N°230 — Octobre 2024

www.linformaticien.com

L'INFORMATICIEN

Étude
Le mainframe pour l'IA et l'hybride

Innovation
La pratique des *acquihire* fait florès

DOSSIER

Vers l'expérience totale

DevOps
Java 23

Logiciel
DBOS, l'alternative à Kubernetes

Retex
PRO BTP maîtrise l'obsolescence de ses systèmes

L 14614 - 230 - F: 8,50 € - RD

Hewlett Packard
Enterprise

Transformez les données dont vous disposez en informations à valeur ajoutée.

Ouvrez le champs des possibles pour vos données et exploiter leur valeur, où qu'elles se trouvent, de l'edge au cloud.

Rendez-vous sur **HPE.com/fr**

www.linformaticien.com
L'INFORMATICIEN

RÉDACTION

88 boulevard de la Villette, 75019 Paris, France.
Tél. : +33 (0)1 74 70 16 30 — *contact@linformaticien.com*

RÉDACTION : Bertrand Garé (rédacteur en chef)
et Victor Miget (rédacteur en chef adjoint).
avec : Olivier Bouzereau, Jérôme Cartegini, François Cointe,
Michel Chotard, Alain Clapaud, Guillaume Renouard, Thierry Thaureaux.

SECRÉTAIRE DE RÉDACTION : Boutheïna Saddi

MAQUETTE ET RÉALISATION : Franck Soulier (chef de studio)

PUBLICITÉ
Tél. : +33 (0)1 74 70 16 30 — *pub@linformaticien.com*

VENTE AU NUMÉRO

France métropolitaine 8,50 € TTC (TVA 5,5 %)

ABONNEMENTS

France métropolitaine 72 € TTC (TVA 5,5 %)
magazine + numérique

Toutes les offres :
www.linformaticien.com/abonnement

Pour toute commande d'abonnement d'entreprise
ou d'administration avec règlement par mandat administratif,
adressez votre bon de commande à :

L'Informaticien, service abonnements,
88 boulevard de la Villette, 75019 Paris, France.
ou à abonnements@linformaticien.com

IMPRESSION

Imprimé en France par Imprimerie Chirat (42)
Dépôt légal : 4ème trimestre 2024

L'INFORMATICIEN est publié par PC PRESSE, S. A. S.
au capital de 130 000 euros.
Siège social : 88 boulevard de la Villette, 75019 Paris, France.

ISSN 1637-5491
Une publication PCpresse,

FICADE

PRÉSIDENT, DIRECTEUR DE LA PUBLICATION :
Gaël Chervet

EX+CX+UX= TX

Notre dossier du mois revient sur l'expérience des salariés dans les entreprises et son évolution future, qui s'annonce comme étant une expérience totale, ou la fusion d'une expérience optimale et unifiée pour toutes les parties impliquées de l'écosystème de l'entreprise : salariés, clients, partenaires… Tout un ensemble, allant du hardware au logiciel et à l'organisation du travail, constitue cette expérience totale. Cela nécessite également de repenser la gestion des salariés, avec des questions toujours en suspens depuis la pandémie, notamment sur le travail à distance et la productivité. S'y ajoute l'apport de l'intelligence artificielle et les gains attendus de son utilisation dans les entreprises.

Outre ce dossier, vous retrouverez dans notre supplément cyber un vaste article sur la protection des postes de travail, véritable complément au dossier sur l'expérience des salariés, mettant en avant l'aspect de la sécurité.

Vous retrouverez bien sûr toutes nos rubriques habituelles, avec en DevOps une analyse des nouveautés de JAVA 23 et le compte rendu de deux grandes conférences aux USA : Oracle Cloud et Dreamforce.

L'actualité de *L'Informaticien* reste évidemment marquée par ses événements. La quatrième édition du Palmarès a rendu son verdict : félicitations à ses nombreux lauréats, qui reflètent la réalité du marché français dans ses diverses catégories. L'autre moment fort est notre rendez-vous du 5 novembre : Stor'Age. Un salon-forum dédié aux questions de stockage et de protection des données. Nous espérons vous voir nombreux à cet événement, qui se tiendra à l'espace Étoile Saint-Honoré. Inscrivez-vous directement sur le site de l'événement à l'adresse suivante : https://storage-forum.com/fr/inscription-storage.html

Nous vous donnons donc rendez-vous le 5 novembre prochain ! □

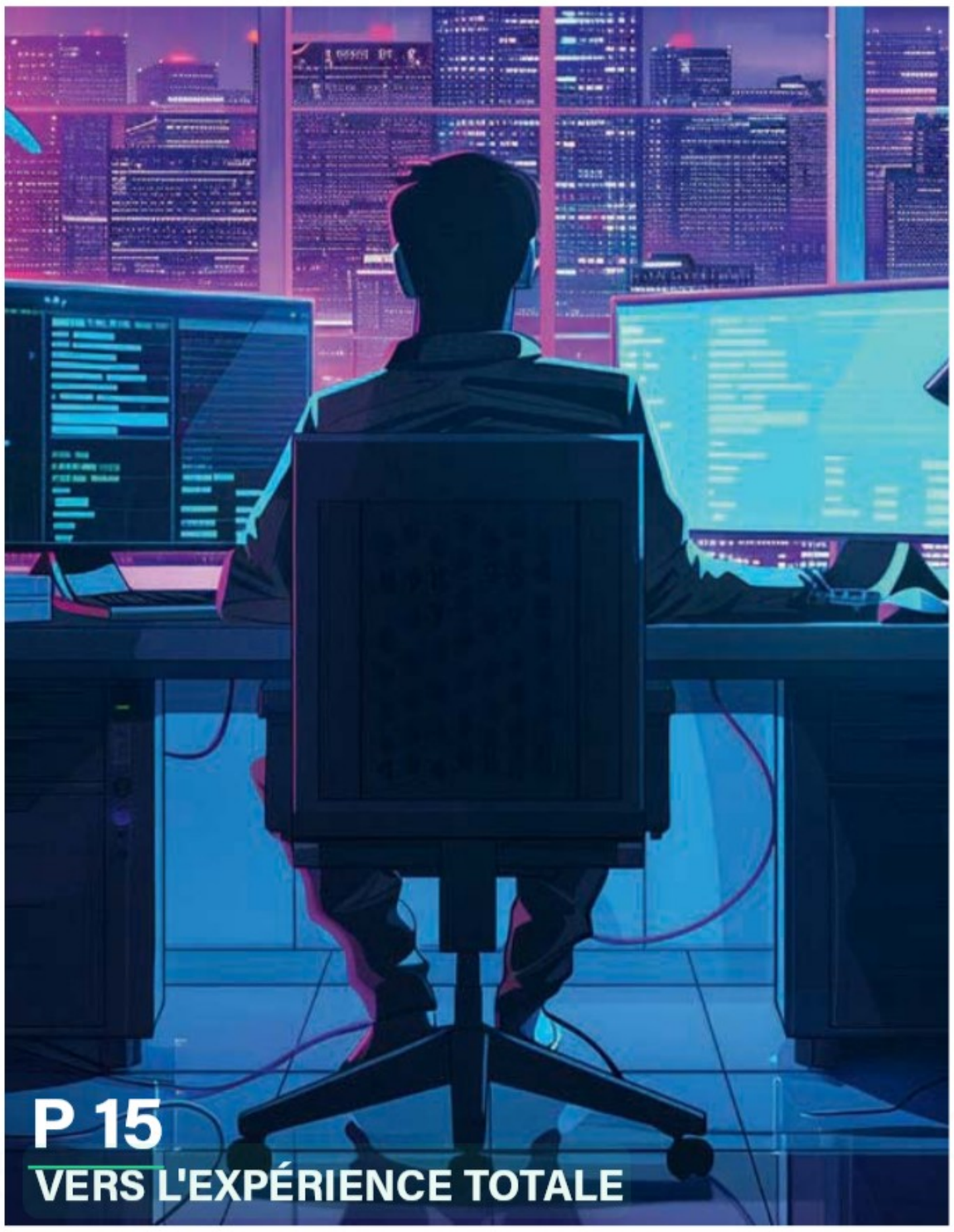
Bertrand Garé
Rédacteur en Chef

PURESTORAGE®

Simplifier le stockage des données pour toujours

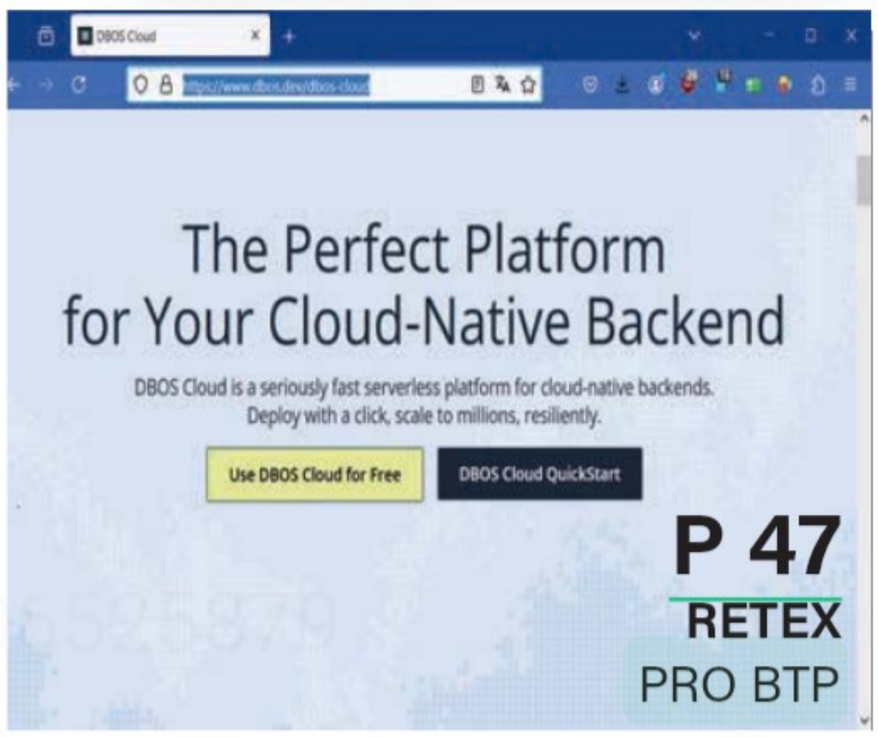
Le stockage à la demande (STaaS) vous permet de bénéficier d'une flexibilité financière et d'une simplicité opérationnelle pour répondre durablement aujourd'hui et demain, aux besoins de votre entreprise.
Evergreen//One™ associe l'agilité du stockage dans le cloud public à la sécurité et aux performances d'une infrastructure all-flash. Cette solution STaaS offre une véritable expérience de cloud hybride.

www.purestorage.com/fr/products/staas/evergreen/one.html



P 15
VERS L'EXPÉRIENCE TOTALE

P 62
ÉTUDE
MAINFRAME PAR KYNDRYL



P 47
RETEX
PRO BTP

Libérez vos fichiers avec Nasuni

Nasuni est une plateforme de données évolutive pour les entreprises confrontées à une explosion des données non structurées dans l'univers IA.

La plateforme Nasuni assure une croissance fluide dans les environnements de cloud hybride, permet le contrôle jusqu'à la périphérie du réseau et répond aux attentes des entreprises les plus modernes en matière de préparation des données pour l'analyse et l'IA.

Croissance illimitée

Protection intégrée

Partage de fichiers collaboratif

+33 (0) 6 51 79 09 66

sales@nasuni.com

nasuni.com

NASUNI

L'EXPÉRIENCE UTILISATEUR

AH OUI, C'EST SUPER BIEN DE REMPLIR MES DOSSIERS CONGÉS-FORMATIONS-NOTES DE FRAIS-RETRAITE-REMBOURSEMENTS, AVEC UN CASQUE DE RÉALITÉS VIRTUELLES!

JE VOIS EN 3D TOUS LES BOTS QUI SE FOUTENT DE MA GUEULE PARCE QUE JE SUIS TOMBÉ DANS TOUS LES PIÈGES QU'ILS M'ONT TENDUS!

L'UX POUR LES PAS LUMINEUX
GUIDE EXHAUSTIF.

HAHAHAHA

¡A PALACASEHAPOS

T'AS VU LE CON SA GUEULE QUAND IL SE REND COMPTE QU'ON N'A PAS PRÉVU DE CASE PARMI LES 312 POUR LE TYPE DE MONTANT QU'IL VEUT QU'ON LUI REMBOURSE...

ET IL N'A PAS ENCORE DEVINÉ QU'ON A UN ALGORITHME QUI TRIE ET ENVOIE TOUTES SES QUESTIONS COMPLEXES AU BUREAU OÙ IL N'Y A JAMAIS PERSONNE ET QU'IL AURA UNE RÉPONSE À LA SAINT GLINGLIN!

HiHi

¡A TOUTA REFAIRE

Hi Hi Hi LA TRONCHE QUAND IL A DÉCOUVERT QU'ON NE L'AUTORISE PAS À CORRIGER SES ERREURS ET QU'IL DOIT TOUT RECOMMENCER À ZÉRO!

HiiiHiii

¡A PERSONNE KI RÉPOND

HOU HOU C'EST TROP MARRANT!

HOU HOU

¡AIAIA

HUHU

J'EN PEUX PLUS, JE VAIS GPT DANS MA CULOTTE!

HOU HOU

HIHIHI

ELLES NOUS EMBÊTENT LES IA, À TOUJOURS TIRER LA COUVERTURE À ELLES.

NOUS AUSSI, ON VOUDRAIT VOIR LA TÊTE DE NOS CLIENTS QUAND ON LES MARTYRISE.

SURTOUT QU'ON LES MARTYRISE PLUS QUE LES ¡A!

F-COINTE

FOCUS

Clearview AI, cette entreprise qui fait fi des sanctions

L'entreprise spécialisée dans la reconnaissance faciale, Clearview AI, a été une nouvelle fois sanctionnée en Europe, cette fois par l'autorité néerlandaise de protection des données (APD) en raison de collectes illicites de photographies de citoyens néerlandais. Malgré les sanctions, l'entreprise persiste et ne semble pas disposée à modifier ses méthodes.

Et une sanction de plus pour l'entreprise américaine de reconnaissance faciale Clearview AI, qui s'est vue infliger une amende de 30,5 millions d'euros par l'APD. Le gendarme des données personnelles néerlandais lui reproche d'avoir créé une base de données illégale contenant 30 milliards de photos d'individus, dont des Néerlandais.

À gauche, Hoan Ton-That fondateur de Clearview AI, en compagnie de membres de l'agence des frontières ukrainienne à qui l'entreprise a apporté son soutien en mars 2022.

© Service national des gardes-frontières d'Ukraine

Pas qu'une affaire de consentement

L'entreprise collecte ces photos d'individus sur internet avant de les convertir en un code biométrique unique pour chaque visage. Problème : elle n'informe à aucun moment, ni ne récupère le consentement des personnes concernées. Ce qui constitue une violation du règlement général sur la protection des données (RGPD).

« La reconnaissance faciale est une technologie extrêmement intrusive, que l'on ne peut pas simplement appliquer à n'importe qui dans le monde », a déclaré Aleid Wolfsen, président de l'APD néerlandaise, cité dans un communiqué. *« Si une photo de vous est sur Internet — et cela ne s'applique-t-il pas à nous tous ? — alors vous pouvez finir dans la base de données de Clearview et être suivi. Ce n'est pas un scénario catastrophe tiré d'un film d'horreur. Ce n'est pas non plus quelque chose qui ne pourrait se produire qu'en Chine »*, a-t-il ajouté. Si le président de l'APD reconnaît l'utilité de la reconnaissance faciale dans certains cas, il estime que ces technologies doivent être opérées par les autorités dans des situations exceptionnelles, et non par des entreprises commerciales.

L'APD a exigé l'arrêt de ces violations et menace Clearview AI d'une astreinte de 5,1 millions d'euros. Ayant constaté que les violations ont continué malgré son enquête, elle cherche désormais des moyens pour tenir responsables les membres du conseil d'administration et les clients de l'entreprise. Elle met ainsi en garde les organisations néerlandaises qui utiliseraient ces services et les menace *« de lourdes amendes »*.

Une décision « illégale »

Le ton est donné. En France, la CNIL avait infligé une amende de 20 millions d'euros à Clearview AI, suivie d'une astreinte de 5,2 millions d'euros pour collecte illicite de photographies de citoyens français. Les autorités italiennes et grecques avaient également condamné l'entreprise à 20 millions d'euros chacune. Plus récemment, le Royaume-Uni a puni la société d'une amende de 8,85 millions d'euros pour avoir collecté les photos de millions de Britanniques.

Malgré tout, Clearview AI ne semble pas disposée à se soumettre aux exigences des autorités européennes. Contacté par L'Informaticien, son directeur juridique, Jack Mulcaire, balaye la polémique et déclare que l'entreprise n'a aucun *« lieu d'affaires (ni clients, ndlr) aux Pays-Bas ni dans l'UE »*. Elle ne mènerait ainsi *« aucune activité qui signifierait qu'elle est autrement soumise au RGPD »*. Il dénonce une décision qu'il juge *« illégale, dénuée de procédure régulière et inapplicable »*.

Nos interrogations concernant les méthodes de collecte et le traitement des données biométriques des personnes sans leur consentement sont restées sans réponse. Sur ces sujets, l'entreprise s'était déjà exprimée à plusieurs reprises outre-Atlantique, affirmant, par l'intermédiaire de Doug Mitchell, l'un de ses avocats, qu'elle ne collectait que des informations publiques provenant d'Internet, et non dans des comptes ou réseaux sociaux privés. En Europe, cette défense ne répond toutefois pas aux exigences strictes du RGPD concernant le consentement et la protection des données.

Pavel Durov mis en examen

Accusé d'héberger des activités criminelles sur sa plateforme et de ne presque appliquer aucune modération, le cofondateur de l'application Telegram, Pavel Durov, a été arrêté à la sortie de son jet privé le 24 août dernier. Remis en liberté le mercredi 28 août, Pavel Durov a toutefois été mis en examen et placé sous contrôle judiciaire.

Le milliardaire de 39 ans a été mis en examen pour :

- Complicité dans l'administration d'une plateforme en ligne permettant une transaction illicite en bande organisée.
- Refus de communiquer, sur demande des autorités habilitées, les informations ou documents nécessaires à la réalisation et à l'exploitation des interceptions autorisées par la loi.
- Complicité dans des infractions telles que la mise à disposition, sans motif légitime, d'un programme ou de données conçus pour nuire à un système de traitement automatisé de données, la diffusion en bande organisée d'images de mineurs à caractère pédopornographique, le trafic de stupéfiants, l'escroquerie en bande organisée, l'association de malfaiteurs en vue de commettre des crimes ou délits, ainsi que le blanchiment de crimes ou délits en bande organisée.
- Fourniture de prestations de cryptologie visant à assurer des fonctions de confidentialité sans déclaration conforme.
- Fourniture et importation d'un moyen de cryptologie n'assurant pas exclusivement des fonctions d'authentification ou de contrôle d'intégrité sans déclaration préalable.

Une atteinte « flagrante » à la liberté d'expression

Pavel Durov a annoncé, lundi 23 septembre, que Telegram allait collaborer avec la justice en transmettant les adresses IP et les numéros de téléphone à la demande des autorités dans le cadre de requêtes judiciaires « valides ».

L'entrepreneur a reçu un soutien important, notamment de personnalités de premier plan telles qu'Elon Musk et Edward Snowden, mais aussi d'une partie de la communauté des cybercriminels, qui ont lancé une vague d'attaques DDoS contre des sites français après l'annonce de son arrestation.

Une pétition lancée par la communauté de TON Society, une initiative blockchain associée à The Open Network (TON) et liée à Telegram, a recueilli des millions de signatures. Pour cette communauté, l'arrestation de Pavel Durov constitue une atteinte « flagrante » à la liberté d'expression. Elle appelle les organismes internationaux et européens à exercer une pression sur la France.

Qualcomm souhaite racheter **Intel**

Ce serait une acquisition de poids si elle venait à se confirmer. Selon le Wall Street Journal, le fabricant de puces Qualcomm s'intéresserait à Intel, actuellement en difficulté. Le PDG de Qualcomm, Cristiano Amon, serait personnellement impliqué dans les négociations.

Bien qu'aucune offre formelle n'ait encore été faite, il est probable que si l'accord se concrétise, il attirera l'attention des autorités antitrust des États-Unis, de la Chine et de l'Europe. Qualcomm pourrait alors être contrainte de faire des concessions, comme céder certaines parties d'Intel. Interrogé par Reuters, Bob O'Donnell, fondateur de TECHnalysis Research, juge « très faible » la probabilité de voir ce rachat aboutir. Il ajoute qu'il est *« peu probable que Qualcomm veuille acquérir l'intégralité d'Intel »*, et qu'*« essayer de séparer les activités de produits de l'activité de fonderie serait tout simplement impossible pour le moment »*.

Intel en difficulté

D'autant plus qu'Intel, bien qu'ayant perdu plus de 50 % de sa valeur boursière depuis le début de l'année, vaut encore près de 94 milliards de dollars, tandis que Qualcomm est valorisée à 188 milliards de dollars. Cela soulève évidemment la question de la faisabilité d'une telle transaction.

Il faut aussi s'interroger sur la reprise des activités de fabrication de puces d'Intel. Pour rappel, Qualcomm manque d'expérience dans ce domaine, n'exploitant pas d'usines (« fabs ») et sous-traitant sa production à des entreprises comme TSMC. La filiale de Softbank, ARM, ainsi qu'Apollo Global Management, seraient également intéressés par Intel. Ce dernier, notamment attiré par la branche produit de la société, aurait été éconduit par Intel, qui aurait affirmé que la société n'était pas à vendre.

RACHATS

Progress Software rachète ShareFile pour 875 millions de dollars

Progress Software, un fournisseur de logiciels d'infrastructure basés sur l'IA, a acquis ShareFile, une unité de Cloud Software Group spécialisée dans le partage et la synchronisation de fichiers. Cette acquisition, d'une valeur de 875 millions de dollars, permettra à Progress Software d'intégrer ShareFile à son portefeuille Digital Experience, offrant ainsi aux entreprises des outils collaboratifs, des flux de travail documentaires propulsés par l'IA, ainsi que des fonctionnalités de partage de fichiers sécurisés et de signatures électroniques.
La transaction devrait être finalisée avant le 30 novembre prochain. ShareFile apportera à Progress 240 millions de dollars de revenus annuels et 86 000 clients.

IBM acquiert Kubecost

IBM continue de renforcer sa suite FinOps avec l'acquisition de Kubecost, un éditeur spécialisé dans la gestion des coûts des environnements Kubernetes, un domaine en pleine expansion. Après l'acquisition d'Apptio, Kubecost vient enrichir l'offre FinOps, notamment la solution Cloudability, qui optimise les performances cloud grâce à l'intelligence artificielle issue d'Apptio et de Turbonomics.
L'intégration de Kubecost permettra aux clients de bénéficier d'un suivi des coûts en temps réel et d'informations précises pour mieux comprendre leurs dépenses d'infrastructure. Elle offrira également des recommandations pour réduire les coûts et éviter le sur-approvisionnement dans les environnements Kubernetes.

Salesforce conclut un accord pour le rachat d'Own Company

Salesforce a annoncé avoir conclu un accord définitif pour acquérir Own Company, un fournisseur de solutions de protection et de gestion des données, pour 1,9 milliard d'euros. Cette acquisition vise à renforcer les capacités de Salesforce en matière de gestion de la relation client, grâce à l'intégration de l'intelligence artificielle (IA). La transaction devrait être finalisée au quatrième trimestre de l'exercice fiscal 2025, sous réserve des approbations réglementaires habituelles.

« La sécurité des données n'a jamais été aussi cruciale, et l'expertise d'Own Company viendra renforcer notre capacité à fournir des solutions robustes à nos clients », a déclaré Steve Fisher, président et directeur général d'Einstein 1, la plateforme de Salesforce conçue pour unifier les données d'entreprise dispersées sur plusieurs applications.
Spécialisée à l'origine dans la sauvegarde et la restauration de données, Own Company propose aujourd'hui une plateforme qui assure l'archivage,

la sécurité et l'analyse des données en mode SaaS. Ces fonctionnalités viendront compléter les solutions Salesforce comme Backup, Shield et Data Mask.

Confluent s'offre WarpStream

Confluent a acquis WarpStream, une plateforme conçue pour les entreprises traitant de grandes quantités de données avec de faibles latences dans leur environnement cloud. Grâce à cette acquisition, Confluent peut désormais offrir des solutions adaptées à tout type d'environnement, que ce soit dans le cloud de l'éditeur, en services managés sur la plateforme Confluent, ou sur le cloud du client via WarpStream.

L'intégration de WarpStream, basée sur un stockage objet similaire au moteur Kora de Confluent, va permettre de proposer une plateforme complète avec des fonctions de gouvernance et de traitement des flux pour des entreprises ayant des besoins en volumes importants et en observabilité. Confluent a précisé que cette acquisition n'aura pas d'impact significatif sur ses résultats financiers.

La start-up d'Ilya Sutskever lève 1 milliard de dollars

La frénésie autour de l'IA générative reste intacte, même pour les sociétés qui ne prévoient pas de rentabilité avant plusieurs années. La start-up Safe SuperIntelligence (SSI), cofondée en juin dernier par Ilya Sutskever, ex-directeur scientifique et cofondateur d'OpenAI, a annoncé avoir levé 1 milliard de dollars auprès de prestigieux investisseurs tels que NFDG, a16z, Sequoia, DST Global et SV Angel.
Cette opération valorise SSI à 5 milliards de dollars, selon des sources de Reuters. La jeune entreprise, qui compte actuellement seulement 10 employés et n'a pas encore dépassé le stade de la recherche et développement, prévoit d'utiliser ces fonds pour renforcer sa puissance de calcul et constituer une équipe de chercheurs et d'ingénieurs entre Palo Alto, en Californie, et Tel Aviv, en Israël. *« Il est essentiel pour nous d'être entourés d'investisseurs qui comprennent et soutiennent notre mission de développer une super intelligence sûre, et qui sont prêts à investir plusieurs années dans la recherche avant de lancer notre produit »*, a déclaré Daniel Gross, PDG de SSI, lors d'une interview.

Les startups d'IA ont levé 48 milliards de dollars en 2024

En 2024, les startups spécialisées dans l'IA ont levé 48,4 milliards de dollars, soit une augmentation de 25% par rapport à 2023, selon des données présentées par AltIndex. Ce montant constitue un nouveau record, surpassant le total de 48,3 milliards de dollars de l'année 2021, malgré une baisse de 35% de l'activité de financement par capital-risque au troisième trimestre par rapport à l'année précédente.
Cette baisse s'explique par un « sentiment mitigé » des investisseurs face aux restrictions d'exportation de puces d'IA vers la Chine imposées par les États-Unis, aux attentes de baisse des taux d'intérêt, ainsi qu'aux résultats parfois décevants des grands acteurs du secteur. Malgré cela, les investissements cumulés dans le secteur dépassent désormais 237 milliards de dollars, avec 65% de cette somme levée par des entreprises américaines, soit environ 151 milliards de dollars. L'Asie et l'Europe suivent avec respectivement 52 milliards et 31,3 milliards de dollars. Les startups spécialisées en apprentissage automatique ont recueilli 113,6 milliards de dollars, tandis que celles axées sur les logiciels d'IA et les technologies de l'information ont levé 76,8 milliards et 59,5 milliards de dollars.

Atomico lève 1,24 milliard de dollars pour soutenir des startups européennes

Atomico a réalisé sa plus grande levée de fonds depuis sa création en 2006, récoltant 1,24 milliard de dollars à travers deux nouveaux fonds destinés à soutenir les startups les plus ambitieuses d'Europe. Le premier fonds, Atomico Growth VI, dispose de 754 millions de dollars et se concentrera sur les séries B et les pré-IPO. Le second, Atomico Venture VI, est doté de 485 millions de dollars et ciblera les séries A et les levées de fonds en amorçage.

Atomico vise à soutenir l'écosystème technologique européen, qui commence à rivaliser avec ses homologues mondiaux. L'Europe capte actuellement 30% des financements pour les stades précoces.

Glean lève 260 millions d'euros pour sa plateforme d'IA pour les entreprises

La start-up Glean, qui propose une plateforme d'IA pour les entreprises, a annoncé avoir bouclé un tour de table de 260 millions d'euros, portant sa valorisation à 4,6 milliards de dollars. La plateforme Work AI de Glean connecte les données d'une entreprise pour automatiser des tâches, fournir des réponses personnalisées et créer des applications ainsi que des agents d'IA sur mesure.
Avec 500 employés, Glean dispose désormais de 550 millions de dollars de liquidités pour poursuivre son développement et répondre à l'accélération de la demande pour ses produits. La start-up a également présenté de nouvelles fonctionnalités pour sa plateforme, telles que l'automatisation des analyses complexes et des flux de travail, ainsi qu'un nouvel outil pour simplifier la conception des invites.

Filigran et Recorded Future s'accordent autour de la threat intelligence

Les deux entreprises approfondissent leur partenariat en intégrant plus étroitement leurs plateformes respectives.

Recorded Future est la plus grande société de renseignement sur les menaces au monde, avec plus de 1 800 clients dans 75 pays. Parmi eux, figurent les gouvernements souverains de 47 pays, plus de la moitié des entreprises du classement Fortune 100 et 40 % des entreprises du classement Forbes Global 100. Alimenté par plus d'un million de sources et mis à jour en temps réel 24/7/365, le Recorded Future Intelligence Cloud est le référentiel de renseignements sur les menaces le plus important, le plus complet et le plus fiable au monde.

La complémentarité entre la suite XTM de Filigran et les renseignements sur les

menaces en temps réel de Recorded Future améliore considérablement l'efficacité de la gestion des menaces. En intégrant les informations sur les menaces de Recorded Future directement dans la plateforme OpenCTI, les organisations bénéficient de données enrichies sur les menaces et d'actions de réponse automatisées. Les principaux bénéfices de cette intégration sont d'avoir une vue unifiée de tous les renseignements sur les cybermenaces, des indicateurs de compromission (IoC) enrichis pour une prise de décision plus éclairée, une automatisation des flux de travail pour une remédiation plus rapide, des mises à jour en temps réel pour une gestion proactive des risques, une vue complète des menaces et posture de sécurité améliorée.

Nota bene, la solution demande d'avoir une licence d'intégration avec Recorded Future.

AntemetA distributeur exclusif de Cloudian

L'ESN spécialisée dans le Cloud hybride et la protection des données a signé un partenariat de distribution exclusif de la solution de stockage objet de Cloudian pour la France.

Cette alliance stratégique vise à offrir une solution logicielle de stockage DataLake pour les besoins de l'IA, la rétention longue durée ou encore le stockage persistant pour les plateformes conteneurs. Depuis 2019, AntemetA et Cloudian œuvrent ensemble pour offrir des solutions de stockage objets souveraines aux entreprises françaises. Cette collaboration repose sur l'utilisation de la plateforme HyperStore S3 de Cloudian et se décline déjà à travers la re-vente des solutions Cloudian mais aussi une offre compatible S3 sécurisée en services managés offerte par AntemetA à ses clients avec des fonctionnalités d'immutabilité et certifiée hébergeurs donnés de santé (HDS).

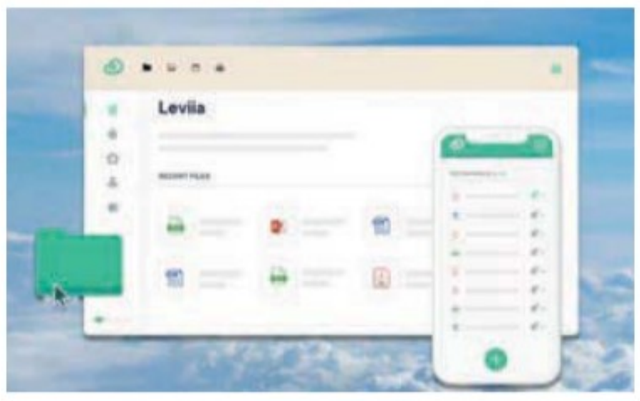
Initialement centré sur la technologie, le partenariat entre AntemetA et Cloudian évolue désormais vers une relation commerciale, marquant une nouvelle étape dans leur collaboration. Pour rappel, la solution de stockage objet Cloudian offre une compatibilité totale de l'API de stockage S3 rendant possible la réversibilité et le débordement entre plateforme *« On et Off premise »* de manière native avec une interopérabilité totale.

Leviia Storage, partenaire et certifié par Bacula Systems

Le fournisseur de solutions de stockage objet français signe un partenariat stratégique avec Bacula Systems et certifie sa plateforme sur Bacula Entreprise.

Leviia franchit un nouveau pas en s'associant à Bacula System SA. Ce partenariat marque la certification officielle de Leviia Storag3 par Bacula System SA, apportant ainsi une garantie de compatibilité totale et d'efficacité aux clients utilisant le logiciel Bacula Enterprise. Parmi les utilisateurs de Bacula Enterprise, on retrouve des organisations de premier plan telles que DPD, la NASA, Vinci, ou même l'US Air Force.

De plus, Leviia déploie une nouvelle région à Lyon afin de renforcer la résilience de sa plate-forme et répond à un enjeu d'importance en sécurité par une diversification géographique des données. Les clients de Leviia ont maintenant la possibilité de choisir où leurs données seront stockées. Cela leur permet, notamment dans le cadre d'un Plan de Reprise d'Activité (PRA), de ne pas centraliser leurs données dans une seule zone



géographique, réduisant ainsi les risques en cas de sinistre ou de cyberattaques.

ALTEN et Schneider Electric s'allient autour de l'industrie 4.0

À l'occasion du salon SIDO à Lyon, l'ESN et Schneider ont annoncé une offre conjointe d'accompagnement des industriels.

Le partenariat entre ALTEN et Schneider Electric est le fruit d'une collaboration de longue date, avec pour objectif d'accompagner la transformation du paysage industriel grâce à des solutions technologiques avancées. En unissant leurs forces, ces deux entreprises ont développé des synergies qui leur permettent de répondre aux nouveaux besoins des industriels. L'offre *« automatisation distribuée pour l'industrie 4.0 »* est conçue pour accélérer l'adoption des cas d'usages liés à cette révolution industrielle et permettre un passage à l'échelle plus rapide de différents sujets. Fruit d'une collaboration d'ALTEN et ses partenaires Schneider Electric et STMicroelectronics, le projet Neo-Automate met en avant un système de maintenance prédictive d'un nouveau genre. Grâce à l'intelligence artificielle embarquée, il est maintenant possible d'anticiper les pannes et d'optimiser la maintenance des machines, réduisant ainsi les coûts et augmentant la durée de vie des équipements. Grâce à l'intelligence artificielle embarquée, il est déosrmais possible d'anticiper les pannes et d'optimiser la maintenance des machines, réduisant ainsi les coûts et augmentant la durée de vie des équipements. La solution Ecostruxure Automation Expert de Schneider Electric joue un rôle central pour le contrôle logique et la gestion des systèmes.

Wasabi et Bechtle partenaires

Le fournisseur de stockage objet s'allie au mastodonte allemand pour proposer sa solution dans la région DACH.

Avec ce partenariat, la solution de Wasabi va être distribuée dans toute la région de langue allemande (Allemagne, Autriche, Suisse). Pour sa part, Bechtle étend son offre de sauvegarde et de restauration des données avec une solution éprouvée. Pour rappel, Bechtle a 14 filiales dans des pays européens et propose des services d'infrastructures et des services managés. Avec cette alliance, Wasabi accélère sur le marché européen et prend une place comme alternative aux autres offreurs de stockage dans les régions alémaniques.

Zoom et Mitel s'allient autour de l'UCaaS

Les deux entreprises spécialistes de la communication et de la collaboration en entreprise mettent en place un partenariat stratégique autour d'une solution en Cloud hybride.

La solution combine Zoom Workplace et Zoom AI Companion avec la plateforme de communication de Mitel, pour une expérience de communication intégrée. Mitel a choisi Zoom pour développer conjointement une offre hybride exclusive avec des capacités bidirectionnelles entre les plateformes Zoom Workplace et Mitel, permettant ainsi aux clients de Mitel de disposer d'une solution complète et de migrer vers Zoom UCaaS selon leurs besoins. L'offre Zoom-Mitel propose une expérience sans add-ons, ni plug-ins. Elle comprend l'intégration des communications unifiées, des applications mobiles, des appareils et des fonctionnalités avancées telles que la présence bidirectionnelle, l'escalade appel-vidéo, le provisionnement et l'administration centralisés de l'utilisateur. Les utilisateurs auront également accès à Zoom Phone avec un support intégré pour les plateformes PBX de Mitel, qu'elles soient déployées sur site ou dans le cloud.

Dans le cadre de ce partenariat, Zoom devient l'offre UCaaS exclusive de Mitel au sein de son portefeuille UC global pour les clients de toutes tailles. Les équipes de vente de Mitel et plus de 7 000 partenaires peuvent pratiquer la vente croisée de l'expérience « Zoom-first » en tant qu'élément central de la solution hybride ou aider les clients à migrer vers Zoom si l'UCaaS est leur modèle de déploiement préféré. Les équipes Zoom peuvent proposer les appareils de Mitel aux clients hybrides ayant besoin de points d'extrémité physiques.

AGENDA

Cisco WebexOne
21-24 octobre 2024
Diplomat Beach Resort, Miami, USA

Celosphere
23-24 octobre 2024
Messe München, Münich, Allemagne

GitHub Universe
29-30 octobre 2024
Fort Mason Center, San Francisco, USA

VMware Explore
4-7 novembre 2024,
Fira Gran Via Barcelone, Espagne

KubeCon & Cloud NativeCon NA
12-15 novembre 2024
Salt Palace Convention Center, Salt Lake City, USA

Microsoft Ignite
18-22 novembre 2024
Hybride, en ligne et au McCormick Place West Building, Chicago, USA

AWS re:Invent
2-6 décembre 2024
Différents sites, Las Vegas, USA

NUTANIX

next

on tour

Paris,
France

.NEXT On Tour revient à Paris le 10 décembre 2024 !

Réservez votre place dès aujourd'hui pour découvrir comment Nutanix vous aide à gérer toutes les données et applications à travers plusieurs environnements informatiques sur une plateforme unique tout en simplifiant les opérations et en réduisant la complexité.

Inscription sur

2024 PALMARÈS L'INFORMATICIEN

NUTANIX

DOSSIER

Vers l'expérience totale

Il va falloir oublier les silos ou séparations existants entre l'expérience utilisateur, clients et employés. Ceux-ci auront un futur commun défini sous le concept de « Total Experience » ou TX, formalisé en 2020 et retravaillé en 2023 par le cabinet Gartner. Selon Gartner, la TX doit surprendre, ravir les clients, faciliter le travail des employés et renforcer la fidélité. La réussite des entreprises passe désormais par cette stratégie de management favorisant la transformation digitale, améliorant les usages des utilisateurs, créant des expériences multiples et interconnectées pour toutes les parties impliquées, que ce soient les clients, les partenaires, les fournisseurs et employés. Ainsi, de l'environnement de travail au service client, en passant par les relations avec les partenaires et les fournisseurs, l'expérience doit être optimisée pour rendre la relation simple, fluide et sécurisée, permettant d'échanger en toute confiance dans un monde digitalisé. Il est temps de s'y mettre, car le cabinet Gartner, toujours un peu optimiste, prédit que d'ici 2026, 60 % des grandes entreprises s'appuieront sur ce concept pour transformer leurs modèles.

Une expérience **unifiée**

La Covid et l'instauration du télétravail sont passées par là, et il est évident que la manière de travailler a beaucoup évolué récemment. Des aspects, comme par exemple le bien-être au travail et une bonne expérience salarié, apportent d'importants avantages métiers comme une meilleure relation avec les clients. C'est ce que constatent les cabinets d'analystes qui prônent désormais une expérience totale qui unifie expérience client et expérience employé.

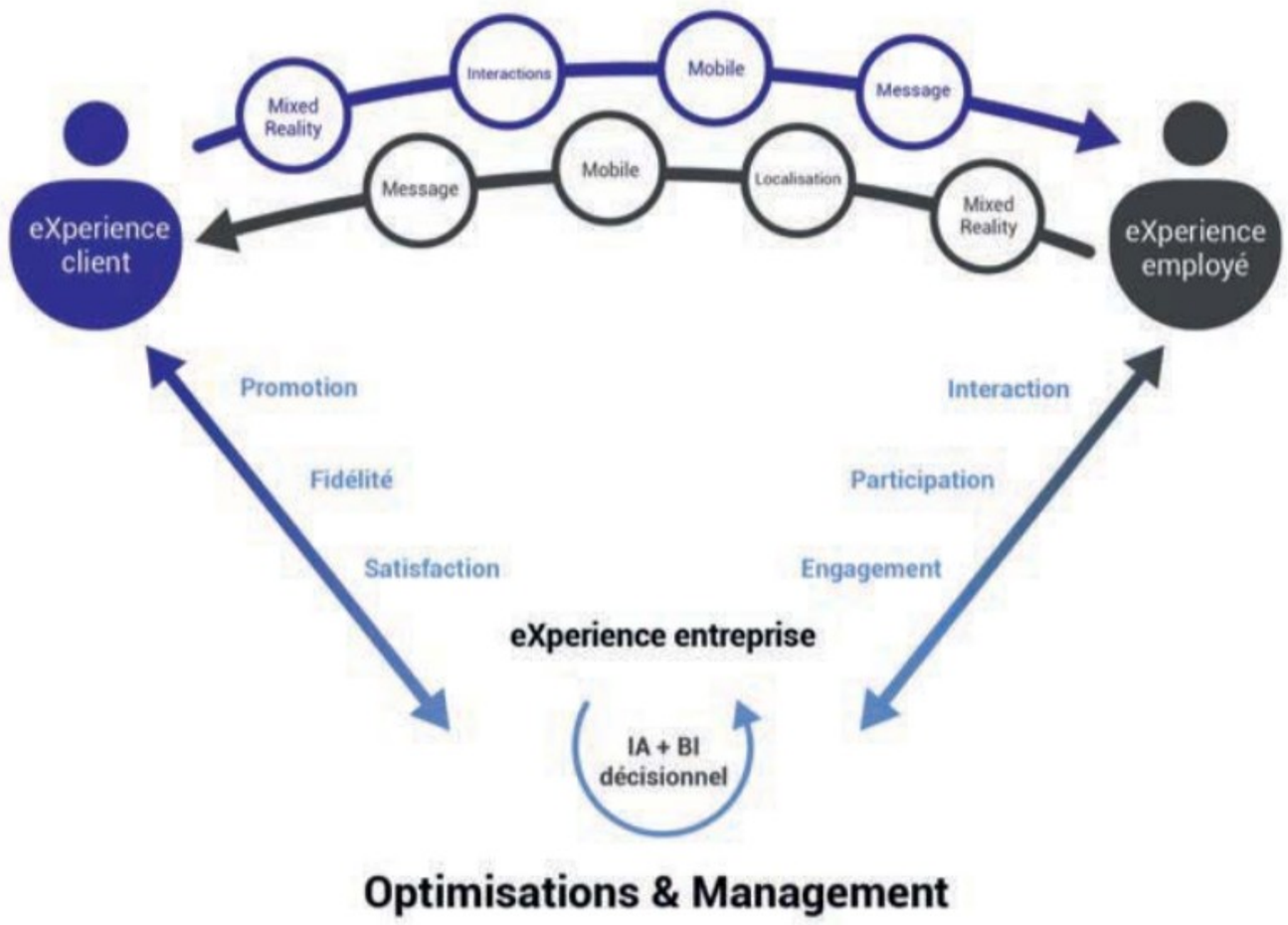
L'expérience salarié est en plein boom. Selon le cabinet d'analyste Zion Market, les plates-formes de gestion de l'expérience salarié dans les grandes entreprises généraient un chiffre d'affaires de 36,18 milliards de dollars en 2023. Le marché va croître de plus de 9 % par an pour atteindre 92,2 milliards de dollars en 2032. Ces plates-formes regroupent les fonctions et applications pour améliorer l'expérience globale du salarié afin que celui-ci s'engage dans l'entreprise. Ainsi, elles prennent en compte la satisfaction du salarié, son bien-être au travail et sa productivité. Le principal objectif est d'établir une ambiance de travail productive, édifiante et plaisante en complément des objectifs et de la culture de l'entreprise.

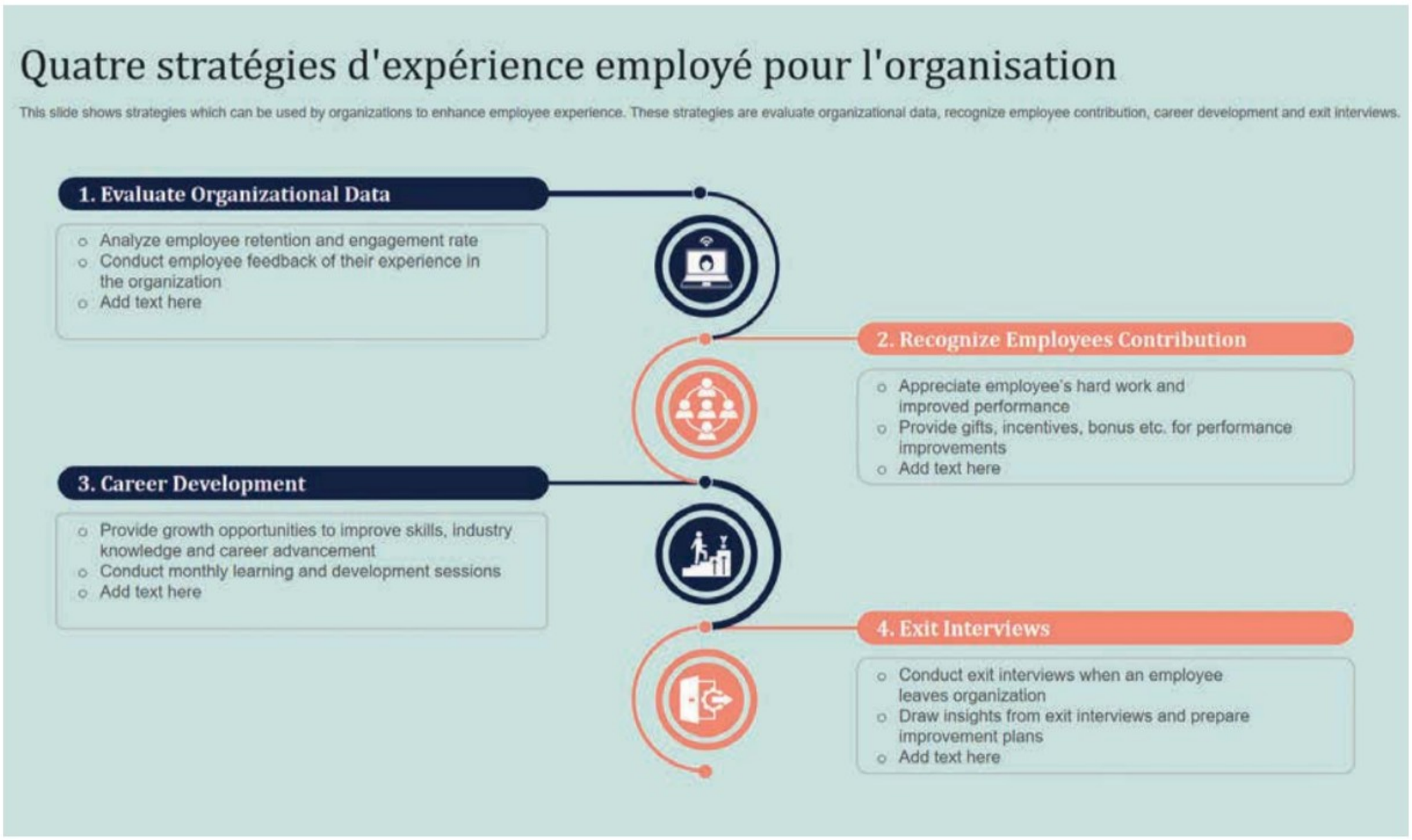
Les raisons d'un développement rapide

La croissance du secteur est liée à différents éléments comme une meilleure prise en compte des avancées technologiques, du bien-être des salariés entre autres. Le but est de renforcer l'engagement des salariés et favoriser la rétention de ceux-ci dans l'entreprise alors que les entreprises reconnaissent avoir du mal à trouver des candidats lors de leur recrutement. Un des facteurs de la croissance est évidemment la digitalisation des services de ressources humaines. Les plates-formes d'expérience salarié fluidifient différents processus RH comme la gestion de la performance, l'écoute des salariés ou l'accompagnement lors de l'embauche (On Boarding). Cela va jusqu'à proposer des programmes de santé ou de bien-être avec des conseils pour éviter la lassitude devant les écrans.

L'expérience totale

Dans un article publié par HumanPerf.com, il est défini l'expérience totale. *« L'amélioration de l'expérience client (CX pour Customer Experience) que l'on prônait depuis des années comme le pilier de la croissance, a été chamboulée par la crise sanitaire de 2020 qui a rebattu les cartes. En effet, l'expérience collaborateur (EX pour Employee Experience) a repris toute son importance face au soudain recours massif au télétravail et ses conséquences organisationnelles et sociales. Un autre impact réside dans la nécessité de revoir l'expérience organisationnelle (OX pour Organizational Experience) afin de repenser la résilience de l'entreprise, d'optimiser ses processus et pérenniser ses marges de manœuvre. Enfin, toutes ces ambitions s'inscrivent dans un paysage numérique où les outils informatiques doivent relever le défi d'une expérience utilisateur (UX pour User Experience) fluide et réussie. CX, EX, OX, UX, tout un programme ! C'est précisément la réunion de toutes ces expériences que le Gartner a identifiée fin 2020 comme une des tendances stratégiques majeures. La* « Total Experience » *ou* « l'expérience totale » *vise ainsi à augmenter la réussite des programmes de transformation digitale en supprimant les silos et en réunissant les équipes et les technologies. L'objectif est de créer des expériences multiples et interconnectées pour toutes les parties prenantes de l'entreprise : clients, partenaires, fournisseurs et employés. La Total Experience est donc quelque part un nouveau terme pour désigner une approche unifiée de la compétitivité… La qualité de l'expérience collaborateur (EX) est donc étroitement liée à la création d'une expérience client (CX) réussie, elle en est même le préalable. En revanche, des progrès sont encore largement réalisables dans ce domaine pour sensibiliser l'ensemble des collaborateurs à la culture de la relation client et à l'amélioration de la compétitivité ».*





Au final, il ne s'agit pas seulement d'améliorer l'expérience d'une seule personne, mais d'améliorer les expériences à l'intersection de plusieurs personnes afin d'obtenir un résultat commercial transformateur. L'expérience totale englobe l'expérience complète de l'entreprise, de l'employé au client et à l'utilisateur. Celle-ci ne consiste pas seulement à prendre soin des clients, mais également à offrir un excellent environnement aux employés ainsi qu'aux utilisateurs. Les expériences utilisateur des clients et des employés sont liées par des fonctionnalités et des interactions spécifiques qui dépendent les unes des autres. Les conditions de ces expériences façonnent la réputation d'une entreprise et affectent la qualité de l'ensemble de ses services. Alors que les entreprises deviennent plus éloignées, virtuelles et distribuées, une stratégie de TX connectée est une nécessité. Il est de plus en plus important d'identifier les intersections entre les disciplines pour obtenir un avantage concurrentiel. L'objectif de l'expérience totale est d'améliorer l'expérience client tout en améliorant l'entreprise dans son ensemble. En donnant la priorité à l'expérience totale, il convient de poser et répondre à des questions sur la façon dont votre expérience utilisateur affecte l'opinion des clients sur votre marque. De même, cela aide à identifier comment améliorer l'engagement et la satisfaction des employés, qui à leur tour, influencent le type de service et d'expérience que ceux-ci offrent aux clients.

Laurent Stoica, directeur régional de Medallia.

« J'ai plus de 80 % de ma carrière dans le domaine du CX, et dernièrement, je l'ai combiné avec EX, Employee Experience. J'ai toujours cru dans cette convergence des domaines. »

Laurent Stoica, directeur régional de Medallia note : *« j'ai plus de 80 % de ma carrière dans le domaine du CX et dernièrement, je l'ai combiné avec EX, Employee Experience. J'ai toujours cru dans cette convergence des domaines. Medallia répond à cette attente, étant donné que c'est une plateforme transverse capable de collecter tous les signaux. On est en pleine implémentation et scalabilisation de la partie ressources humaines, employés, administration, formation, etc. Et donc, avec ça vient la partie expérience. Finalement, les entreprises ont compris que si on veut avoir des clients satisfaits, capables de générer beaucoup de revenus, ça passe forcément par la case satisfaction employé. Les transformations technologiques et les données comme carburant de la compréhension du niveau de satisfaction et de l'expérience alimentent la tendance ».* □

Les nouveaux **apports fonctionnels et techniques**

Différents éléments concourent à l'expérience salarié ou à l'expérience totale. Du matériel aux applications, en passant par l'organisation du travail, l'expérience se construit par un but commun : donner au salarié le meilleur environnement de travail possible pour être productif et engagé dans le projet de l'entreprise.

Samira Bekhtaoui, directrice d'Asus Business a une approche assez originale du matériel à fournir aux salariés : *« l'expérience utilisateur, c'est un mot qui revient souvent, enfin ce sont deux mots qui reviennent souvent, pour parler de la facilité, des fonctionnalités et des choses qu'on met en avant, nous, sur nos postes de travail, à disposition des collaborateurs de toutes les entreprises, pour justement que leur travail soit fait dans les meilleures conditions. Il y a différentes modalités de travail selon l'espace. Donc ça peut être au bureau, ça peut être en mobilité, ça peut être à la maison dans le cadre du télétravail. L'expérience utilisateur, on peut déjà la prendre dans le prisme de l'espace, où est-ce qu'on va se trouver précisément pour travailler. Si on prend l'idée de l'expérience utilisateur déjà dans le prisme de l'espace, se dire comment est-ce que j'offre la meilleure expérience à un utilisateur qui va être 100 % en présentiel ou 100 % en mobilité ou 100 % en télétravail ou en hybride, un peu le mélange des trois. C'est cette expérience hybride qui est à la fois un bénéfice, mais aussi une contrainte parce qu'on retombe sur les problèmes de sécurité, de connectivité. Cela va être des produits qui vont rester au bureau. C'est des tours, ou les all-in-one, c'est les ordinateurs, tout inclus. Ou alors, on va avoir le PC portable, qui pourra justement passer du bureau au café en attendant mon train à la maison. Ou le mini PC, le tout petit desktop qui se pose sur le bureau et qui, chez certains employés, c'est fou, est ramené à la maison aussi et qui est pluggé à un moniteur LCD. Donc, on peut tout imaginer en termes de poste de travail en tant que tel. La docking, c'est un élément indispensable du travail, quel que soit l'espace. Pour moi, c'est vraiment un élément qui a été accessoire pendant un certain temps, la docking station, et qui maintenant devient indispensable, primordial. Outre le double écran pour certaines fonctions dans l'entreprise, les environnements se font léger comme pour les ordinateurs portables. La connectique n'est pas oubliée., on a évidemment de l'USBA, de l'USBC, de l'HDMI. Et là, je pense par exemple à notre P5 qu'on est en train de lancer, le premier PC en Lunar Lake qu'on est en train de lancer. Ce n'est pas parce qu'il est fin, beau, léger et plutôt design qu'il ne bénéficie pas de tous les ports nécessaires. Finalement, ce qu'on nous demande, c'est un ordinateur qui soit transportable, qui soit léger et fin, mais qui, en même temps, permette aux collaborateurs qui doivent l'utiliser tout le temps d'avoir une ergonomie d'utilisation importante. C'est pour ça qu'on a généralisé le 16-10e sur les écrans de plus en plus sur nos machines. Parce qu'ils permettent d'avoir une qualité, un confort d'utilisation plus important sur du 14 pouces*

Simon Daly, Employee Experience Strategy Director chez Qualtrics.

« Je pense que cela revient à quelque chose de très simple : s'assurer d'écouter ses employés et ses clients, de comprendre leurs feed-back, et d'agir en conséquence. Car il est presque pire d'écouter sans prendre aucune action. »

que ce que l'on avait précédemment. Les entreprises sont dans une phase de renouvellement pour arriver à fournir cette expérience optimale ou ils sont à rallonger la durée de vie des équipements en disant qu'on va faire avec ce qu'on a et on verra à côté ? Il y a les deux. En fait, il y a plusieurs phénomènes qui arrivent en même temps. Le comportement des entreprises en ce moment dans le renouvellement de leur parc, c'est de l'étaler le plus possible. Alors qu'auparavant, le temps de détention des machines était aux alentours des 4-5 ans, il est en train de grossir, s'allonger est passé entre 5 et 6 ans. Ça, c'est la première tendance que l'on observe, c'est le temps de détention des machines dans les entreprises se rallonge. À cause des coûts, évidemment, mais aussi grâce au fait que les machines tiennent plus longtemps techniquement, technologiquement, et permettent justement qu'il n'y ait pas d'impact négatif dans le poids RSE des entreprises, dans le poids carbone des entreprises. Ça, c'est la première tendance. La seconde, c'est que nous allons arriver là en 2025 et nous allons arriver à un moment où ces machines qui ont été renouvelées de manière massive post-Covid vont devoir être renouvelées ».

L'ASUS ExpertBook P5 est au maximum des tendances du moment pour les ordinateurs portables en entreprise.

Les éléments d'une bonne expérience employé

Simon Daly, Employee Experience Strategy Director chez Qualtrics, précise : *« comment pouvons-nous déterminer que c'est une bonne expérience ? Quatre points sont importants. J'arguerai que les individus font qu'ils auront certaines préférences avec vous, contre quelqu'un d'autre. Mais il y a des éléments componentiels. Il y a certainement quelque chose d'apprécié et de reconnu dans l'organisation dans laquelle vous travaillez. Il est nécessaire que le salarié soit capable de se développer personnellement et professionnellement, et c'est vraiment, vraiment important. Un autre élément très important, c'est la confiance et la communication avec les dirigeants seniors. C'est le troisième. Et puis le quatrième, c'est l'alignement des valeurs. Donc, en fait, je veux travailler pour une organisation dont les croyances et les valeurs que j'ai sont alignées avec l'organisation pour qui je travaille. Et vous êtes ainsi beaucoup plus heureux, vous ressentez cette connexion, et vous êtes prêt à aller plus loin. »*

Simplification et automatisation

Les principaux points autour de l'expérience salarié restent la simplification de tâches complexes par des interfaces utilisateurs plus ergonomiques et pensées pour avoir un accès intuitif afin d'éviter des courbes d'apprentissage trop long pour celui-ci et faire que l'application ou la fonction soit utilisée. L'arrivée de l'IA générative et son interface textuelle et vocale ont démocratisé son utilisation. Simon Daly indique : *« nous utilisons l'IA et notre technologie pour faire ce qu'on appelle le modèle d'attrition prédictive. Ce qu'il va faire, c'est regarder la connaissance qu'il a accumulée. Il va regarder les points d'impact. Il va regarder les données et il va vous conseiller, car si vous continuez cette trajectoire, vous avez un risque d'attrition d'une certaine quantité. C'est l'attrition qui va se produire. Et puis vous pouvez ajuster le modèle pour dire que si on change de point de vue et qu'on améliore le bien-être, si on améliore ça, en utilisant l'attrition prédictive, vous pouvez améliorer l'expérience et voir ce qui aura un meilleur effet pour vos employés ».* Il ajoute : *« à mon avis, c'est vraiment important que les employés sentent qu'ils peuvent avoir une voix, peu importe laquelle. Il y a plusieurs types d'organisation, mais je pense que cela revient à quelque chose de très simple : s'assurer d'écouter ses employés et ses clients, de comprendre leurs feed-back, et d'agir en conséquence. Car il est presque pire d'écouter sans prendre aucune action. »* □

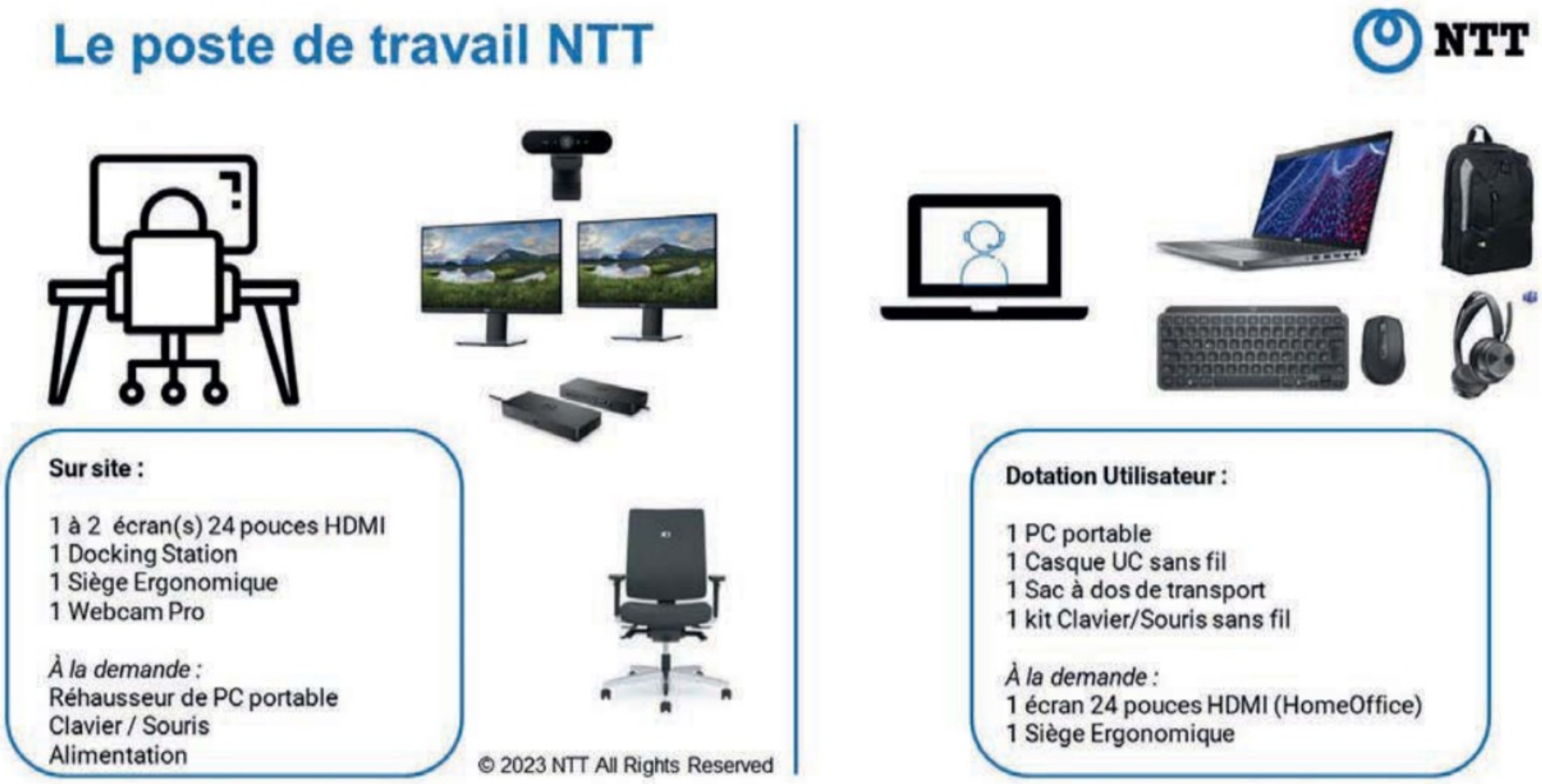
NTT DATA revoit sa **stratégie Flex Office**

L'entreprise d'ingénierie informatique a revu sa stratégie Flex Office à la fin de la pandémie de la COVID pour unifier l'expérience des salariés sur un outil maison, Sharing Cloud.

Pierre-Alexandre Mérot, Senior GTM Software practice Manager chez NTT, entre directement dans le vif du sujet : « Sharing Cloud, c'est une solution technologique qui est venue apporter des éléments de réponse à une stratégie post-COVID de FlexOffice chez NTT. En fait, la réflexion était déjà un petit peu "connue" sur cet outil-là, puisqu'on est historiquement partenaire et intégrateur de la solution chez nos clients, utilisateur, et client utilisateur de cette solution depuis plus d'une dizaine d'années, mais sur des fonctionnalités assez restreintes de réservation de salles. On est des très forts consommateurs de salles de visioconférence, et ce depuis très longtemps du fait de l'activité de NTT en tant qu'intégrateur international, donc avec des besoins de communication quasi quotidiens et permanents. Et donc est arrivé assez rapidement un besoin au travers des salles de réunion et de leur disponibilité. La salle de réunion étant une ressource rare, très convoitée, et avec des problématiques habituelles de comment est-ce que je la réserve, est-ce qu'elle est disponible sur Outlook, comment je gère ces ressources-là, et comment est-ce que je m'assure qu'elles sont utilisées de manière la plus efficace possible. On s'est retrouvé dans un contexte comme beaucoup d'entreprises courant Covid et post-Covid, où la question était de se dire comment est-ce qu'on aménage nos locaux et qu'on rend l'utilisation des locaux la plus efficace possible.

Dans cette volonté de réinvestir les lieux, ça s'est jalonné aussi d'un certain nombre de réorganisations de nos sites en France, où on avait des déplacements de sièges qui ont été effectués courant 2022, où la question se posait vraiment de se dire on a des sièges avec une superficie plus petite, volontaire, puisqu'on déménageait de Rungis sur la Défense, comment est-ce qu'on allait articuler cette réflexion autour de l'espace disponible et des ressources disponibles aux utilisateurs, avec une stratégie de flex office et une nouvelle politique de télétravail.

Il y avait tous les éléments parfaits pour démarrer une réflexion sur le flex office et ce qu'on a appelé l'employé de journée en France, chez MTT, qui consiste à répondre à des questions relativement simples, qui est de se dire où je peux travailler, quand je peux réserver des ressources, pour quel type d'activité et comment je le fais. Ce groupe de travail a été monté par la direction dont j'ai le plaisir de faire partie, avec un certain nombre de managers et de collaborateurs, plutôt orientés sur les outils et les technologies de la collaboration, pour se dire OK, commençons cette étude-là, voyons ce que le groupe nous propose comme outil en interne et ce que nous on avait déjà comme expérience, entre autres avec Sharing Cloud, et voyons ce qu'on peut bâtir autour d'un cahier des charges qu'on a établies avec la direction. Cela a été relativement intéressant ».



Pierre-Alexandre Mérot, Senior GTM Software practice Manager chez NTT.

> *« Finalement, ça fédère autour des équipes et ça facilite le retour au bureau. Mais un retour au bureau, je dirais, qui est voulu, pas qui est subi. »*

Une recherche de simplicité

« Un des marqueurs les plus importants, c'était d'avoir une solution la plus simple d'usage possible et la plus unifiée. Pourquoi ? Parce que le groupe nous proposait un outil pour réserver les bureaux, un outil pour pouvoir réserver des salles de réunion, un autre outil pour déclarer quand est-ce qu'on était au bureau ou en télétravail, et au final on s'est retrouvé avec une myriade d'outils, alors qu'on avait déjà en exploitation un outil qu'on n'exploitait pas sans plein de potentiel qui était Sharing Cloud, et qui en plus, paradoxalement, on vendait au client avec toutes ses capacités. On a eu la chance de pouvoir utiliser un site pilote qui est celui de Toulouse, qui est une petite agence en termes d'espace, on a une cinquantaine de postes et six salles de réunion, en se disant, on étend cette expérience, on commence à définir des règles d'usage, et on accompagne l'adoption avec ses collaborateurs toulousains » explique le cadre de chez NTT. Il ajoute : *« donc l'expérience est basée et partie de comment je réserve une salle de réunion. Finalement comment est-ce que j'intègre la politique de télétravail et tout le concept de l'employé de journée dans ce flex office. C'est-à-dire que je déclare où je travaille, donc ça c'est un module de présence au bureau, ça permet de simplifier et de permettre à tous les collaborateurs d'utiliser également une interface unique pour dire cette semaine, lundi, je suis au bureau, et par l'intermédiaire d'un clic qui suit, je réserve d'ailleurs un bureau pour lundi, pour la matinée, pour un créneau, pour la journée complète. Une approche très visuelle et très graphique »*. Il continue : *« il y avait cette idée d'avoir quelque chose d'accessible pour faciliter l'adoption, pour diminuer le nombre d'outils, ça c'était vraiment une clé pour nous, parce que pour que ça soit plus simple d'usage, il fallait qu'il n'y ait qu'une seule porte d'entrée au final et qui soit accessible, qu'on soit au bureau, qu'on soit avec un terminal qu'on a tout le temps, c'est-à-dire un smartphone, on s'est dit que c'était quand même une clé assez intéressante d'adoption, ou plus traditionnellement et de manière un peu plus précise sur la gestion des occurrences, si on veut faire plusieurs réservations en même temps avec un portail web qui permet d'aller un peu plus loin dans les paramètres de réservation »*.

Après ce pilote, la solution a été étendue sur le siège de la Défense à ce moment-là, et sur un autre site sur Antony exactement, la zone de stockage qui est aussi équipée de bureaux et d'un laboratoire, donc on sait qu'il y a des allers-retours et des ressources qu'on peut réserver à l'avance, et évidemment le site de Toulouse.

Le contexte spécifique d'une ESN

Pierre-Alexandre Mérot précise : *« le cadre de la politique de télétravail en France est la suivante, chaque collaborateur peut avoir jusqu'à trois jours de télétravail par semaine sur un ou deux sites complémentaires à son site de rattachement administratif de l'entreprise. Comme on est une SN, on a des personnes qui sont rattachées administrativement au siège, mais qui opérationnellement sont sur site client ou en déplacement, donc il y a cette notion de site de référence, et il y a cette possibilité d'avoir son domicile en télétravail, voire, et ça c'est un bénéfice qu'on a qui est assez conséquent, un deuxième site de télétravail, typiquement, on a conscience qu'on a des collaborateurs qui partent avec les enfants en vacances chez les parents ou qui ont des contraintes personnelles, qui sont intégrés dans la politique. Dès lors que le site est* « assuré » pour une activité de télétravail, c'est la condition sine qua non de fournir le certificat d'assurance ». Il note que *« ça a un impact. Ça veut dire un renforcement de la sécurité, de la connectivité, un rôle base d'accès, etc. Aujourd'hui, d'un point de vue sécurité du poste de travail, 100 % des collaborateurs sont équipés avec des terminaux portables NTT, qui sont eux-mêmes sous contrôle du groupe, avec un certain nombre d'outils, à la fois du Zscaler pour tous les contrôles et le filtrage de connexion, du VPN, du Zscaler, de la proxification.*

En termes de modules de sécurité, le poste de travail chez NTT est capable, indépendamment du réseau corporate sur lequel le collaborateur pourrait être connecté quand il est dans nos bureaux, de travailler de n'importe où, tout en garantissant l'identité du collaborateur. Donc, on a l'octave plus du Zscaler pour le filtrage. Aujourd'hui, le fait de travailler sur six clients, par exemple sur un réseau public ou au domicile, sont des contextes dans lesquels le poste de travail NTT permet d'effectuer ses tâches, c'est d'accéder aux ressources qui sont nécessaires. S'il y a d'autres niveaux de ressources à accéder, il y a des plateformes spécifiques, par exemple, pour des accès à distance sur des environnements, et ça, finalement, le niveau de sécurité est le même en interne qu'en externe ».

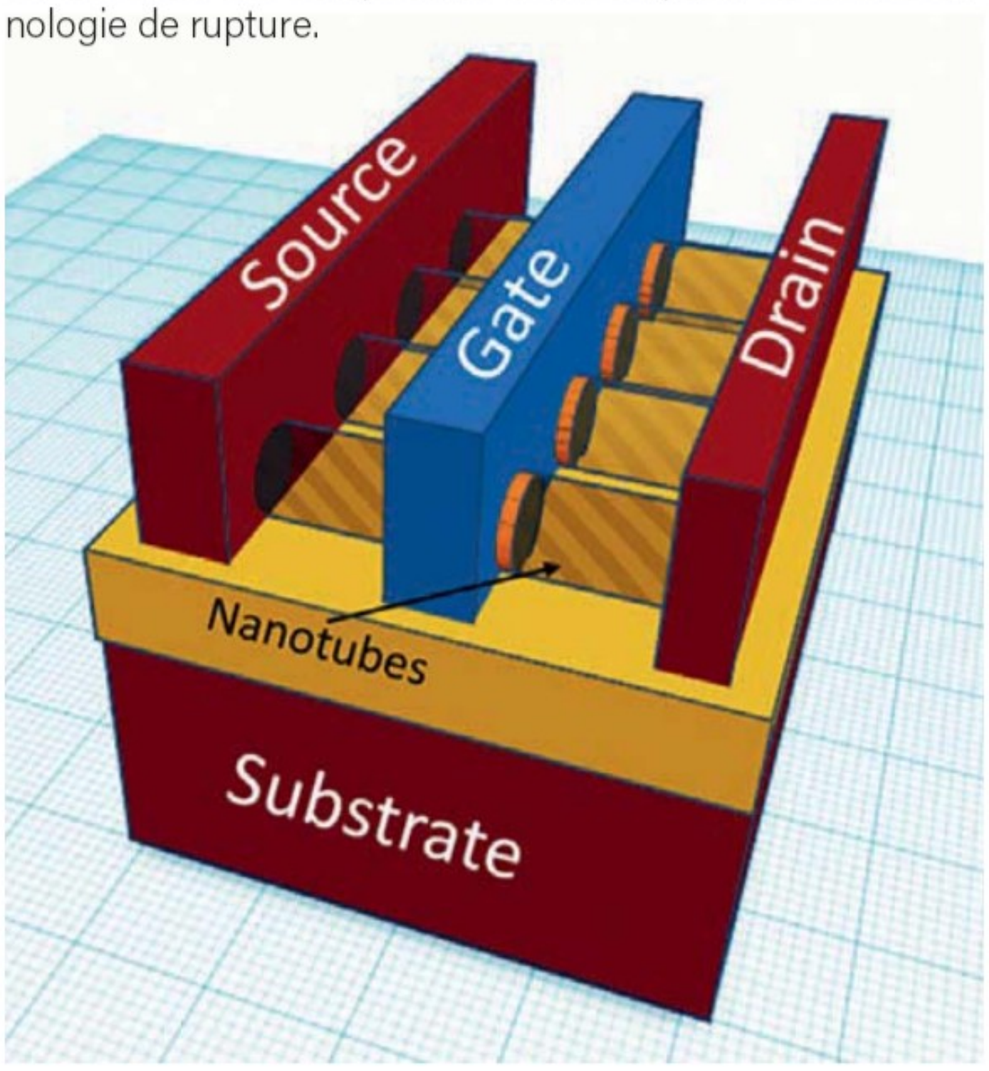
Il conclut : *« finalement, ça fédère autour des équipes et ça facilite le retour au bureau. Mais un retour au bureau, je dirais, qui est voulu, pas qui est subi. Ça, c'est important de le préciser, parce qu'en facilitant ces moyens de se retrouver et de s'organiser facilement, un manager qui veut faire ses réunions d'équipe, il peut lui aussi se dire, tiens, je vais réserver ces cinq bureaux-là et cette salle de réunion, je réserve tout en même temps »*. □

TPU Les nanotubes de carbone vont-ils relancer la loi de Moore ?

C'est un peu un serpent de mer de l'informatique : quelle technologie viendra remplacer le silicium lorsqu'on aura atteint au bout de la loi de Moore ? Plusieurs pistes sont à l'étude, mais depuis le début des années 2000, de nombreux groupes de chercheurs misent sur les nanotubes de carbone. Un premier NPU vient de voir le jour grâce à cette technologie.

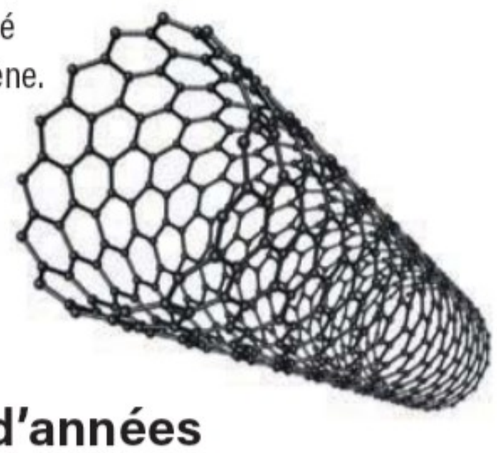
Depuis le début des années 2000, les annonces se succèdent très régulièrement. Cet été, des chercheurs de l'université de Beijing publiaient un papier dans la très sérieuse revue scientifique Nature sur le tout premier TPU (puce dédiée à l'IA) réalisé à base de nanotubes de carbone. Dans ce papier de recherche, l'un des coauteurs, Zhiyong Zhang, révèle que la puce qui a pu être produite contient 3 000 transistors à nanotubes de carbone, ce que l'on appelle les CNFET (Carbon Nanotube Field-Effect Transistors).

L'annonce est d'importance, car la promesse des nanotubes de carbone est élevée : un transistor de ce type offre un coefficient energy-delay product (EDP) 9 fois supérieur à celui d'un transistor Si/SiGe FinFET exploitant des canaux Silicium/Germanium. En clair, le transistor est 3 fois plus rapide tout en consommant 3 fois moins d'énergie… de quoi ouvrir une nouvelle ère dans l'informatique. Encore faut-il pouvoir produire ces transistors de manière industrielle et surtout créer des composants évalués à partir de cette technologie de rupture.



Dans un transistor CNFET (Carbon Nanotube Field-Effect Transistors), les nanotubes de carbone sont placés entre la source et le drain, ce qui vient booster la performance de la porte logique, tout en abaissant l'énergie consommée.

Le nanotube de carbone est réalisé en enroulant une feuille de graphène.



Des recherches menées depuis plus d'une dizaine d'années

Déjà, en 2015, IBM Research a annoncé dans un article scientifique publié par Science, avoir pu produire des transistors avec des contacts de 9 nm. Shu-Jen Han, responsable du groupe de recherche Science & Technology Group nanométriques considérait déjà que les nanotubes de carbone allaient permettre de relancer la loi de Moore : *« nous savons tous que les nanotubes de carbone ont d'excellentes propriétés électriques ; les porteurs se déplacent beaucoup plus rapidement dans les nanotubes de carbone que dans le silicium. C'est pourquoi nous sommes tous, y compris IBM, si intéressés par ces nanotubes. Le grand défi a été la taille du contact. Je dirais qu'elle est maintenant plus importante que le canal [dans les efforts pour réduire la taille des transistors]. »*

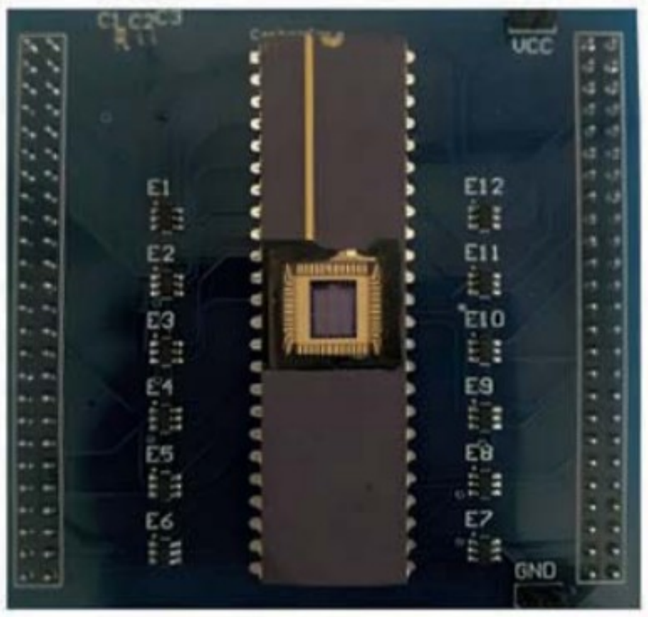
IBM était loin d'être seul à travailler sur ces transistors. En 2016, Science publiait un article des chercheurs de l'université Madison du Wisconsin qui annonçaient avoir produit des transistors 1,9 fois plus rapides que leurs homologues à base de silicium. La technique de fabrication utilisée par ces chercheurs était encore assez loin d'un process industriel. Ceux-ci n'avaient alors produit qu'une surface de 1 pouce sur 1. En 2020, les chercheurs du MIT estimaient avoir franchi une étape clé dans l'industrialisation de la technologie. Ceux-ci sont parvenus à produire des transistors CNFET sur des Wafers standard de 200 millimètres, le standard de l'industrie du silicium. En outre, la technique qu'ils ont imaginée leur a permis d'accélérer le processus de fabrication des transistors CNFET d'un facteur supérieur à 1000 pour des coûts de production revus à la baisse. Les chercheurs américains étaient parvenus à déposer des nanotubes de carbone bord à bord sur les plaquettes et produire des matrices de 14 400 transistors CNFET sur 14 400. Pour atteindre un tel résultat, ceux-ci ont travaillé avec des fournisseurs venus du silicium, à savoir Analog Device et la fonderie de semi-conducteurs SkyWater Technology. Un autre atout de la technique mise au point par le MIT est de pouvoir fonctionner à une température proche de la température ambiante. Alors que les transistors à base de silicium nécessitent une température

de l'ordre de 400° C à 500° C, une production à « basse » température ouvre de nouveaux horizons : il devient possible d'empiler les couches de transistors sans risque d'endommager les couches inférieures. Ainsi, on peut imaginer produire des processeurs en 3D, ce qui permet d'espérer, là encore, des progrès intéressants en termes de performance.

Le TPU, un cas d'usage parfait pour l'électronique de nouvelle génération

Si l'annonce du MIT a soulevé beaucoup d'espoirs, il ne s'agissait que d'un jalon supplémentaire passé vers la longue quête que doit mener l'industrie du semi-conducteur vers un premier microprocesseur à base de CNFET. Il aura fallu attendre le mois de juillet 2024 pour avoir une réalisation concrète venue de l'université de Beijing. Les chercheurs n'ont pas produit un microprocesseur, mais un TPU (Tensor processor chip), une puce spécialisée dans l'apprentissage et l'exécution des IA. On trouve ces puces chez les Hyperscalers. Les NPU qui équipent les smartphones et PC modernes en sont une émanation directe. Avec l'essor des IA génératives, les besoins de puissance de calcul ont explosé… de même que la facture énergétique associée à l'apprentissage et l'exécution des inférences de ces IA. Disposer de composants notoirement plus puissants et moins énergivores serait un avantage compétitif décisif dans cette course aux IA. Le transistor CNFET est donc un bon candidat, d'autant qu'un TPU est un composant bien moins complexe à produire qu'un microprocesseur moderne.

La puce dévoilée cet été par l'équipe de Kaixiang Kang est le tout premier TPU conçu sur la technologie des nanotubes de carbone. Celle-ci repose sur 3000 transistors CNFET et peut effectuer des opérations de type convolution (opération binaire) et de multiplication de matrices. Le papier de recherche publié par Nature précise que les transistors sont organisés en unités de traitement de 3 x 3 unités, des PE (Processing Unit) et que 9 d'entre elles forment un réseau systolique capable d'effectuer des opérations de convolution d'entiers sur 2 bits et des multiplications de matrice. L'architecture a été pensée pour porter des réseaux de neurones artificiels. Les chercheurs ont pu tester un réseau neuronal convolutif à cinq couches dédié à la reconnaissance d'images.



Le TPU présenté dans l'article publié par Nature Electronics ne compte que 3 000 transistors à nanotubes de carbone. Ceux-ci sont organisés en unités de traitement de 3 x 3 transistors. Cette architecture ne peut actuellement traiter que des opérations sur 2 bits.

Le nanotube doit encore faire ses preuves face à des puces silicium ultra-optimisées

Pour un taux de précision de 88 % maximum, la consommation d'énergie a atteint 95 µW. En s'appuyant sur ce résultat, les chercheurs ont estimé en simulation qu'un TPU 8 bits s'appuyant sur cette technologie pourrait délivrer une puissance de 1 TOPS par watt pour une fréquence d'utilisation de 850 MHz… une performance énergétique plutôt décevante puisqu'un GPU extrêmement performant comme le A100 SXM de Nvidia présente un TDP de 400 W, mais délivre 1 248 TOPS (INT8 Tensor Core), soit plus de 3 TOPS par Watt. Du côté des TPU plus économes en énergie, l'Edge TPU de Google Cloud réalisé en silicium délivre 4 Tops pour une puissance électrique de 2 W seulement, ce qui, là encore, est bien supérieur à ce que les chercheurs espèrent obtenir de leur puce CNFET. On est encore très loin de la promesse initiale d'une efficacité énergétique 9 fois supérieure au silicium… Les chercheurs estiment que de gros progrès peuvent encore être obtenus pour augmenter les performances et l'efficacité énergétique de leur architecture : l'augmentation du nombre de bits traités par les PE, la réduction de la taille du composant et l'implémentation d'une logique CMOS vont clairement dans le sens d'une augmentation du nombre de TOPS délivrés. Enfin, le chercheur évoque la capacité de créer des puces en 3D, ou même créer des puces hybrides. En intervenant en bout de chaîne de fabrication de puces silicium classiques, les industriels pourraient ajouter une couche de transistors CNFET aux microprocesseurs classiques, pour ajouter un NPU à nanotube de carbone à une puce X86 classique, par exemple.

En 2020, les chercheurs du MIT parvenaient à produire des transistors CNFET sur un wafer standard, une nouvelle étape franchie vers l'industrialisation de ces puces de nouvelle génération.

Si l'annonce de cet été est riche en promesses, on voit qu'il faudra encore plusieurs années de recherches avant que le nanotube de carbone ne s'impose dans l'électronique de nos PC. □

AC

Amazon Echo Show 8

Plus fluide, plus puissant, plus intelligent

Avec la 3ème génération de l'Echo Show 8, Amazon a voulu améliorer chaque aspect de son écran connecté. Du design, aux composants, en passant par les capacités audio, l'affichage, ou encore les fonctionnalités, ce nouveau modèle a été repensé pour offrir une expérience encore plus fluide et plus riche. Tour d'horizon.

Dix ans après le lancement de la première enceinte connectée avec assistant vocal par Amazon, le marché a littéralement explosé. Dévoilée en 2014, l'Amazon Echo associée à son assistante vocale Alexa a initié le développement d'un nouvel écosystème d'enceintes et d'écrans intelligents dédiés à la maison connectée. Devenue incontournable, cette catégorie de produits suscite désormais une concurrence féroce entre les géants de la tech Amazon, Google et Apple. Capables de traiter toutes sortes de requêtes vocales courantes quasi instantanément, ces appareils ont su se rendre de plus en plus indispensables tant dans le cadre privé que professionnel. Des modèles comme l'Echo Show 8 constituent en outre de véritables hubs numériques pour l'écoute musicale, la lecture de vidéo, la visioconférence ou même la vidéosurveillance. La troisième génération de ce modèle d'Amazon va encore plus loin en introduisant des fonctionnalités innovantes d'affichage permettant à différentes personnes de bénéficier d'une expérience personnalisée avec l'écran.

Design et ergonomie : un rafraichissement discret et efficace

L'Echo Show 8 (3ème gen) arbore un design plus épuré et élégant grâce notamment à un écran bord à bord de 8 pouces et un nouveau revêtement mat moins sensible aux traces de doigt. Doté d'un boitier légèrement plus compact que son prédécesseur, il dispose de différents boutons de commande physique (volume, bouton mute et cache-caméra) sur le dessus. La caméra qui était jadis bizarrement placée en haut à droite de l'écran se trouve désormais au centre. Bien qu'Amazon propose différentes tailles d'écrans plus petites ou plus grandes, le format 8 pouces constitue sans doute le plus polyvalent. L'appareil s'avère en effet suffisamment grand pour afficher confortablement du contenu multimédia (vidéos YouTube, sites d'informations, recettes…), mais assez discret pour s'intégrer facilement sur une étagère ou un bureau, par exemple. Avec une définition HD de 1280 x 800 pixels et un traitement antireflet efficace, la dalle tactile affiche une image lumineuse et contrastée qui reste visualisable dans n'importe quelles conditions lumineuses.

Des contenus interactifs personnalisés

Mention spéciale pour la fonction Auto Cadre qui ajuste dynamiquement l'image pour permettre à la caméra de suivre vos mouvements pendant les appels vidéo afin que vous apparaissiez toujours au centre de l'écran même si vous vous déplacez ! Une innovation particulièrement intéressante pour ceux qui utilisent l'appareil pour les appels et les réunions en visioconférence. Ce modèle intègre également une autre nouvelle fonctionnalité baptisée Adaptive Content. Celle-ci utilise une technologie de détection de proximité pour ajuster automatiquement le contenu en fonction de votre distance par rapport à l'écran. Concrètement,

Pour garantir une meilleure confidentialité, l'Echo Show 8 (3ème gen) intègre un cache physique pour la caméra, ainsi qu'un bouton pour couper le micro.

CARACTÉRISTIQUES TECHNIQUES DE L'AMAZON ECHO SHOW 8 (3ème GEN)

- **Processeur :** MediaTek MT8189 avec réseau neuronal Amazon AZ2 (8 cœurs)
- **Mémoire vive :** 2 Go
- **Taille de l'écran :** 8 pouces
- **Définition :** HD (1200 x 800 pixels)
- **Système audio :** 2 haut-parleurs néodyme 51 mm large bande avec radiateur passif
- **Connectivité :** Wi-Fi 802.11a/b/g/n/ac, Bluetooth A2DP
- **Protocole de communication :** Zigbee, Matter et Thread
- **Support :** Dolby Atmos
- **Assistant vocal :** Alexa
- **Caméra :** 13 Mpx
- **Connectique :** port d'alimentation
- **Dimensions (l x l x h) :** 200 x 139 x 106 mm
- **Poids :** 1,03 kg
- **Tarif au lancement :** 169,99 €

lorsque vous êtes loin, il affiche des informations basiques avec de grands caractères visibles à distance telles que la météo, l'heure, ou les gros titres. À l'inverse, quand vous vous rapprochez, il affiche un contenu plus détaillé et interactif comme des widgets personnalisés, des playlists musicales, des recettes, etc. Grâce à la technologie Visual ID (une technologie d'identification visuelle), l'Echo Show est en outre capable d'identifier les différentes personnes qui se sont enregistrées sur l'appareil pour leur afficher un contenu personnalisé : agendas, notes, commandes (sur Amazon évidemment !), actualités ou vidéos basées sur leur historique de navigation, etc.

Des améliorations à tous les étages

L'Echo Show 8 repose sur une nouvelle puce MediaTek MT8189 épaulée par 2 Go de mémoire vive qui booste les performances de l'écran dans la plupart des usages. Ce gain de puissance optimise la gestion du multitâche, mais aussi le chargement des applications et les interactions avec Alexa. Amazon a fait également de gros efforts pour améliorer la qualité audio de l'Echo Show 8 qui peut désormais faire office d'enceinte Bluetooth. Compatible Dolby Atmos, l'appareil qui intègre deux haut-parleurs néodyme offre une spatialisation sonore d'une efficacité étonnante. De plus, ce modèle peut désormais être appairé avec d'autres enceintes Echo pour créer un système de son multiroom. À l'usage, les commandes vocales permettent de lancer en un clin d'œil une radio, la lecture d'un livre sur Audible, ou un morceau de musique sur une plateforme de streaming prise en charge comme Amazon Music, Spotify, Apple Music, ou Deezer.

Une expérience française optimisée

Le géant Américain a ouvert il y a quelques années un laboratoire spécifiquement dédié au développement d'Alexa en langue française. Au cœur de l'expérience, l'assistance vocale d'Amazon se montre de plus en plus efficace au fil du temps dans sa compréhension des requêtes en français. Pour améliorer son efficacité, l'intelligence artificielle d'Amazon est entraînée à apprendre les spécificités de la langue de Molière, mais aussi la culture, l'histoire, ou encore la gastronomie française. Amazon a enrichi également les fonctionnalités d'Alexa via de nouvelles routines et une personnalisation de plus en plus poussée. Les utilisateurs peuvent programmer des actions complexes via une simple commande vocale et par exemple allumer la lumière, visionner une caméra et lancer une playlist simultanément. Grâce à la fonctionnalité de reconnaissance faciale, chaque utilisateur enregistré peut accéder à son propre profil et à ses préférences de contenu. Amazon a également pris en compte les préoccupations des Français et des Européens en général sur la confidentialité des données. L'Echo Show 8 (3ème gen) est équipé d'un cache physique pour désactiver la caméra ainsi que d'un bouton pour couper le micro. Il est également possible de demander à Alexa de supprimer les enregistrements vocaux.

Avec cette troisième génération, Amazon signe un écran connecté particulièrement complet. Plus performant et réactif que son prédécesseur, ce modèle bénéficie à la fois d'un affichage plus précis et d'une bien meilleure qualité audio. Bien plus qu'un assistant vocal, il se montre désormais aussi efficace pour la visioconférence, que pour le streaming vidéo et l'écoute musicale. L'Echo Show 8 (3ème gen) intègre de surcroît des nouvelles fonctionnalités telles qu'Auto Cadre et Adaptive Content qui permettent de profiter d'une expérience de plus en plus interactive et personnalisée. Compatible avec les protocoles de communication Zigbee, Matter et Thread, l'appareil constitue également un hub très pratique pour configurer et piloter toutes sortes d'objets et périphériques connectés compatibles. Disponible en blanc ou en noir, il affiche un prix un peu plus élevé que la concurrence, mais qui se justifie par ses nombreuses innovations. □ **J.C**

DCIM

Schneider Electric renforce sa position dans le refroidissement liquide

L'industriel a signé un accord pour la reprise de Motivair, un spécialiste du refroidissement liquide et des solutions avancées de gestion thermique pour les systèmes de calcul haute performance.

Basée à Buffalo, NY, Motivair a été fondée en 1988 et compte actuellement plus de 150 employés. Tirant parti de son expertise en ingénierie et de sa connaissance approfondie du secteur, Motivair propose un panel d'offres haut de gamme incluant des unités de distribution de liquide de refroidissement, des échangeurs de chaleur en porte arrière, des plaques froides ainsi que des unités de dissipation de chaleur aux côtés des refroidisseurs pour la gestion thermique.

Selon les termes de la transaction, Schneider Electric fera l'acquisition d'une participation majoritaire initiale de 75 % au capital de Motivair pour une transaction en numéraire de 850 millions de dollars, qui inclut la valeur d'une majoration d'impôts, et qui valorise Motivair à un multiple moyen à un chiffre sur la base du chiffre d'affaires projeté pour l'année 2025. La transaction est soumise aux conditions de clôture d'usage, y compris la réception requise des approbations des autorités de régulation compétentes, et devrait être finalisée dans les prochains

Une vue de la gamme de refroidisseur de Motivair.

trimestres. Une fois la transaction finalisée, Motivair serait comptabilisée au sein de l'activité Gestion de l'énergie de Schneider Electric. Le groupe compte acquérir les 25 % de participation minoritaire restant en 2028.

Une offre circulaire pour les onduleurs

Expérimentée pour la première fois en France en 2024 dans le cadre d'un programme novateur de réutilisation et de recyclage, en partenariat avec le distributeur leader Ingram Micro, Schneider Electric a cherché à accroître la durabilité et la circularité de ses onduleurs.

La nouvelle offre d'onduleurs circulaires de Schneider Electric ne soutient pas seulement les ambitions des partenaires visant à saisir les opportunités de croissance associées aux produits durables, mais leur permet d'agir en tant que partenaires de confiance pour les services de reprise, de recyclage et de remplacement. Cette offre répond également aux exigences environnementales et réglementaires locales telles que la directive européenne sur l'efficacité énergétique (DEE), l'accord de Paris et les objectifs climatiques des Nations Unies (ONU). Cette offre prévoit des processus clés : reprise de l'onduleur en fin de vie, diagnostic et test détaillés, démontage et remise à neuf de toutes les pièces critiques individuelles et des composants en fin de vie tels que les batteries, les interrupteurs et les LED, réassemblage et test rigoureux. Toutes ces opérations sont effectuées avant le reconditionnement et la revente afin d'obtenir la même qualité et la même garantie qu'un onduleur neuf. □

B.G

LA PLATEFORME DCIM

Elle a obtenu un nouveau niveau de certification de la Commission électrotechnique internationale (CEI). Cette certification SL2 est associée à des exigences plus strictes et à une résilience de sécurité supérieure à celle de la certification SL1, obtenue l'année dernière. La plate-forme NMC3 est intégrée dans la majorité des produits DCIM EcoStruxure IT de Schneider Electric et fournit une application de gestion à distance sur le réseau pour les infrastructures d'énergie sécurisée et de refroidissement. Le système EcoStruxure IT Secure NMC apporte une gestion améliorée des micrologiciels intégrés grâce à un nouvel outil dédié ; l'outil système Secure NMC System Tool transforme le processus fastidieux consistant à rechercher et à installer la dernière version du micrologiciel sur tous les périphériques, ce qui peut accélérer le processus de 90 %. Les utilisateurs n'ont plus besoin de rechercher un micrologiciel de manière ponctuelle, de vérifier si sa version est la plus récente pour leur périphérique et de lire les notes de mise à jour pour comprendre ce qui est inclus dans la nouvelle version avant de le télécharger et de mettre à jour leur périphérique. Au lieu de cela, l'outil système Secure NMC System Tool informe les clients que le nouveau micrologiciel est disponible et les invite à installer la nouvelle version.

SMART TECH

DELPHINE SABATTIER

7H30 | 18H30

VOTRE ÉMISSION QUOTIDIENNE DÉDIÉE À L'INNOVATION

Dans l'émission SMART TECH animée par Delphine Sabattier, l'actualité du numérique et de l'innovation prend tout son sens. Chaque jour, des spécialistes décryptent les dernières news, les tendances, et les enjeux soulevés par l'adoption des nouvelles technologies.

N°230 orange™ | N°246 bouygues | N°163 free

B SMART 4. Change

Tendances

7ème édition de Grand Angle

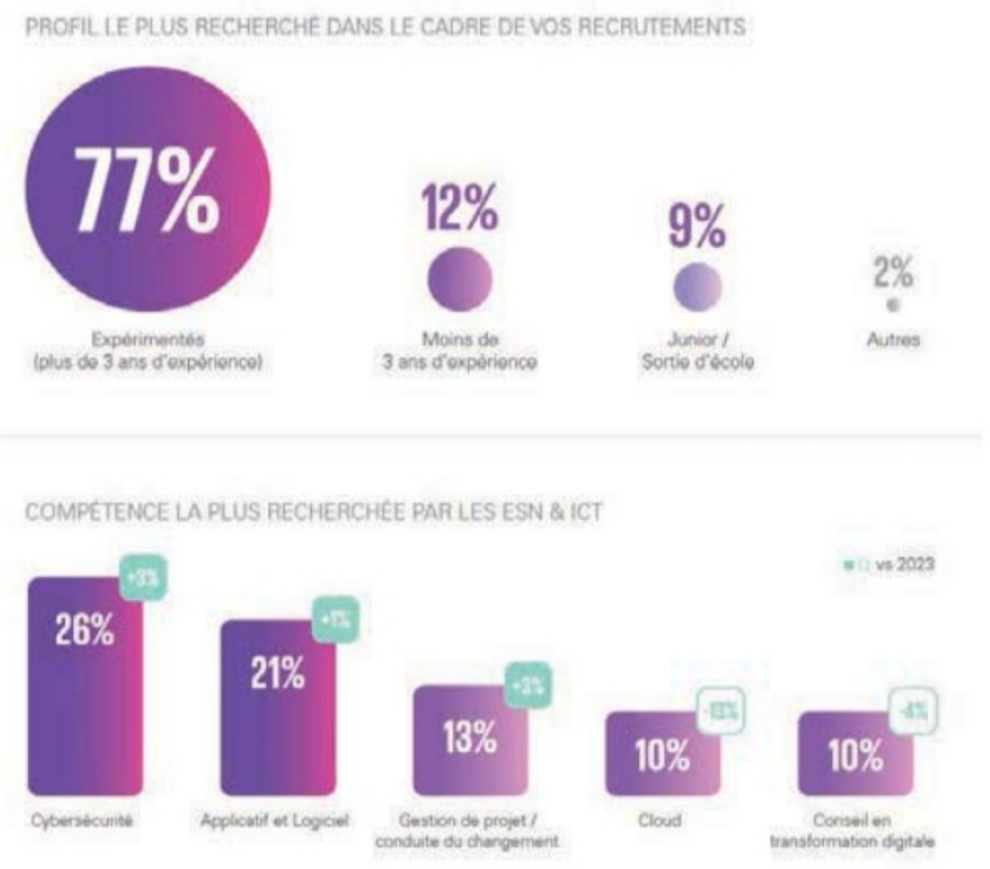
Comme chaque année depuis plusieurs éditions, Numeum et KPMG ont récemment présenté la 7ème édition de leur étude Grand Angle sur les tendances du marché des ESN. Le principal enseignement reste l'optimisme, bien que l'on observe un ralentissement depuis l'été dernier.

Le baromètre de Numeum aborde les thèmes habituels : stratégie de croissance, gestion des talents, RSE et Innovations. Concernant la croissance, les ESN devraient connaître en 2024 une décélération marquée du marché numérique en France. La capacité à recruter reste le principal moteur de croissance des ESN & ICT (pour 23 % d'entre elles), suivie par les acquisitions externes et le positionnement sur des secteurs porteurs (21%). Parmi ces secteurs, les services financiers, l'énergie et l'industrie continuent de dominer la dynamique de croissance. Le service le plus performant demeure le conseil en transformation digitale, représentant un quart des entreprises du secteur, suivi par l'applicatif/logiciel et la cybersécurité. Enfin, l'étude souligne également le potentiel du Off/nearshore pour répondre aux besoins de couverture géographique.

L'exposition sectorielle et le ralentissement de la demande (63 %), le manque de ressources (54%) et les crises et conflits internationaux (54%) restent les principaux facteurs de risque dans l'exécution de la stratégie de croissance des ESN & ICT. Malgré ces défis, la confiance reste élevée avec 96 % des entreprises se projetant vers l'avenir avec assurance, prêtes à atteindre leurs objectifs de croissance grâce à une stratégie claire et une vision tournée vers l'innovation et le numérique responsable.

Les facteurs les plus porteurs de croissance sont, à 23 %, la capacité de recruter, suivi du positionnement sur des secteurs porteurs, les acquisitions, l'innovation technique et technologique et le gain de parts de marché. Par secteur d'activité, la finance reste le secteur le plus dynamique devant l'énergie et l'industrie qui précède l'aérospatiale et la défense. Le secteur des technologies et de la santé suivent pour respectivement 8 et 6%.

Si la majorité des centres de production reste sur le territoire français (95 %), l'off-shore et le near-shore sont désormais présents pour des raisons de coûts, de couverture géographique, d'effectifs, d'expertise, d'innovation, de spécialisation dans la sécurité et la flexibilité des différents droits du travail.



Une politique de recrutement ambitieuse

Même si un ralentissement des recrutements a été observé ces derniers mois, les objectifs de recrutement 2024 et à 3 ans restent ambitieux. Le secteur du numérique reste donc en tension, même si l'impact de l'intelligence artificielle (IA) et de l'intelligence artificielle générative (IAG) sur les plans de recrutement est notable, soulignant l'importance croissante de ces technologies dans le secteur. 67 % des ESN & ICT estiment que l'IA ou l'IAG ont un impact sur leur plan de recrutement. Cela se traduirait en moyenne par une réduction des recrutements de 5 % pour les GE & ETI et de 19 % pour les PME & TPE. La proportion de free-lances, indépendants et sous-traitants dans les effectifs des ESN et ICT reste à 7 % pour les GE & ETI et à 21% pour les PME & TPE, reflétant une tendance vers une plus grande flexibilité et adaptation aux besoins du marché. Dans ce contexte, le profil expérimenté (plus de 3 ans d'expérience) reste très majoritairement le profil le plus recherché par 77%

des répondants. Les compétences en cybersécurité sont, quant à elles, la compétence la plus recherchée pour 26 % du secteur.

Contrairement à l'enquête 2023 qui mettant largement en avant l'augmentation des salaires, le moyen de fidélisation des talents considéré comme le plus efficace en 2024 est la flexibilité du temps de travail (21%), suivi par les plans de formation (20 %) et l'organisation d'événements de team building (15 %). En complément, l'augmentation moyenne de salaire accordée en 2023 aux salariés a été de 5 %. Le nombre de jours de télétravail proposés varie, avec une majorité offrant deux jours par semaine (57 %). Les actions de formation en 2024 se concentreront principalement sur l'expertise technique (68 %) et la méthodologie de projet (12 %).

Le RSE s'intègre dans la stratégie globale des ESN

L'enquête révèle une nette tendance à l'intégration accrue des critères ESG dans la stratégie globale des entreprises. En effet, 83% des ESN & ICT ont instauré un comité RSE/ESG au sein de leur gouvernance. Ce comité se concentre principalement sur des initiatives internes, visant à fédérer les projets, renforcer la durabilité de l'entreprise et favoriser la rétention des talents. Les priorités pour les trois prochaines années s'orienteront davantage vers l'extérieur, avec des efforts accrus pour réduire l'empreinte environnementale, développer des solutions et services à finalité ESG, et créer des emplois liés à cette thématique. Ainsi, les acteurs du numérique que 7 donneurs d'ordres sur 10 accordent un degré d'importance de plus de 10 % aux critères des ESN & ICT dans leurs critères de sélection.

84 % des ESN & ICT ont mis en place un suivi de leur empreinte carbone, c'est 6 % de plus que l'année dernière. En moyenne, lorsqu'un objectif de réduction de l'empreinte carbone a été fixé, il représente une réduction de 26 % pour les GE et les ETI et 22 % des émissions actuelles pour les PME & TPE. Ces enjeux cruciaux constituent de plus en plus un facteur de compétitivité non tarifaire pour les entreprises.

L'IA, N°1 dans les technologies

Après avoir été au cœur de toutes les discussions, l'intelligence artificielle, y compris l'IA générative, est passée n°1 des domaines technologiques et des répondants soulignant son importance croissante dans divers secteurs. Elle est utilisée par 73 % des ESN et ICT pour les tâches administratives (compte-rendu automatique de réunion, traduction, correction orthographique...). Son utilisation est ensuite plébiscitée pour la R&D (62 %) et pour la stratégie marketing (42 %).

Les investissements dans l'IA et l'IAG sont massifs et les projections à 3 ans illustrent bien l'importance de cette technologie. Les GE et ETI estiment que dans 3 ans, elles investiront 1,2 % de leur chiffre d'affaires dans l'IA chaque année. Cela sera même 5 % du chiffre d'affaires pour les PME et TPE. □

B.G

NUMEUM S'ASSOCIE AU CAMPUS CYBER

À l'occasion du Cybermoi/s, Numeum et le Campus Cyber annoncent deux initiatives majeures : le lancement de la plateforme Cyber4Tomorrow.fr et la signature d'un partenariat avec Cybermalveillance.gouv.fr.
L'ambition est de mobiliser l'écosystème cyber et les citoyens autour de pratiques durables, inclusives et sécurisées dans le cyberespace. Ce type d'initiative se révèle d'autant plus utile dans un contexte où près d'une entreprise française sur deux (49 %) déclare avoir subi au moins une cyberattaque en 2023. Cette plateforme unique s'adresse à tous les professionnels de la filière et propose un catalogue d'actions concrètes conçues et validées afin qu'elles soient directement activables, en fonction du profil du volontaire et du public visé. Plusieurs actions y sont déjà référencées telles que l'accompagnement des victimes de cyberattaques, la promotion de l'hygiène numérique ou l'animation d'un atelier d'OSINT auprès de ces publics, le mentorat de femmes dans la cybersécurité, ou encore la sensibilisation des seniors, par exemple en maison de retraite, à l'importance des pratiques d'hygiène numérique auprès d'adolescent(e)s. Chaque fiche action est accessible en ligne, avec toutes les ressources nécessaires pour permettre à tout professionnel de s'engager immédiatement et de contribuer directement à la sécurité numérique tout en intégrant des objectifs de responsabilité sociale et environnementale. Ce lancement marque également un partenariat clé avec Cybermalveillance.gouv.fr, signataire de la « Charte C4T Organisation », faisant de la plateforme un pilier du Cybermoi/s. En incitant les professionnels à s'engager via des actions concrètes, cette collaboration vise à renforcer la résilience collective face aux cybermenaces tout en intégrant des pratiques respectueuses de l'environnement.

Industrie

Alpine crée un centre d'ingénierie de pointe

Le constructeur automobile va mettre en place un centre d'excellence à Viry-Chatillon, Hypertech Alpine, un nouveau centre d'ingénierie de pointe de la marque, qui rassemblera les meilleurs talents pour contribuer au développement des véhicules très haute performance et des innovations sur les technologies de pointe, pour Alpine ainsi que pour Renault Group.

À l'issue du processus de consultation des partenaires sociaux, au cours duquel les échanges ont été constructifs et une expertise indépendante a pu être menée, la direction d'Alpine confirme son projet de transformation du site de Viry en centre d'excellence en ingénierie et haute technologie, et cela, dès la fin 2024. Les activités F1 de Viry, hors développement d'un nouveau moteur, sont maintenues jusqu'à la fin de la saison 2025. Chaque collaborateur concerné par ce projet de transformation, se verra proposer un nouveau poste au sein d'Hypertech Alpine. Cette démarche, qui s'inscrit dans la durée, permettra de pérenniser les compétences des équipes de Viry-Châtillon en technologies de pointe (hardware et software) et sécuriser la propriété intellectuelle du Groupe.

Plusieurs projets au programme

Le centre de Viry-Châtillon va conduire différents projets. Il va contribuer au développement de la future Supercar Alpine. Il sera de plus en charge du développement à court et moyen termes de batteries pour les véhicules sportifs de la marque. À un horizon plus long terme, ces équipes conduiront également des activités de Recherche & Advanced Engineering sur la chimie des cellules à très haute densité d'énergie, notamment avec la technologie solid state battery, en conditions d'utilisation extrêmes pour des applications du type Supercar. Il conduira des activités de Recherche & Advanced Engineering sur ces technologies en collaboration avec Ampere, afin de préparer les ruptures technologiques attendues sur les prochaines générations de véhicules électriques.

Une expertise reconnue

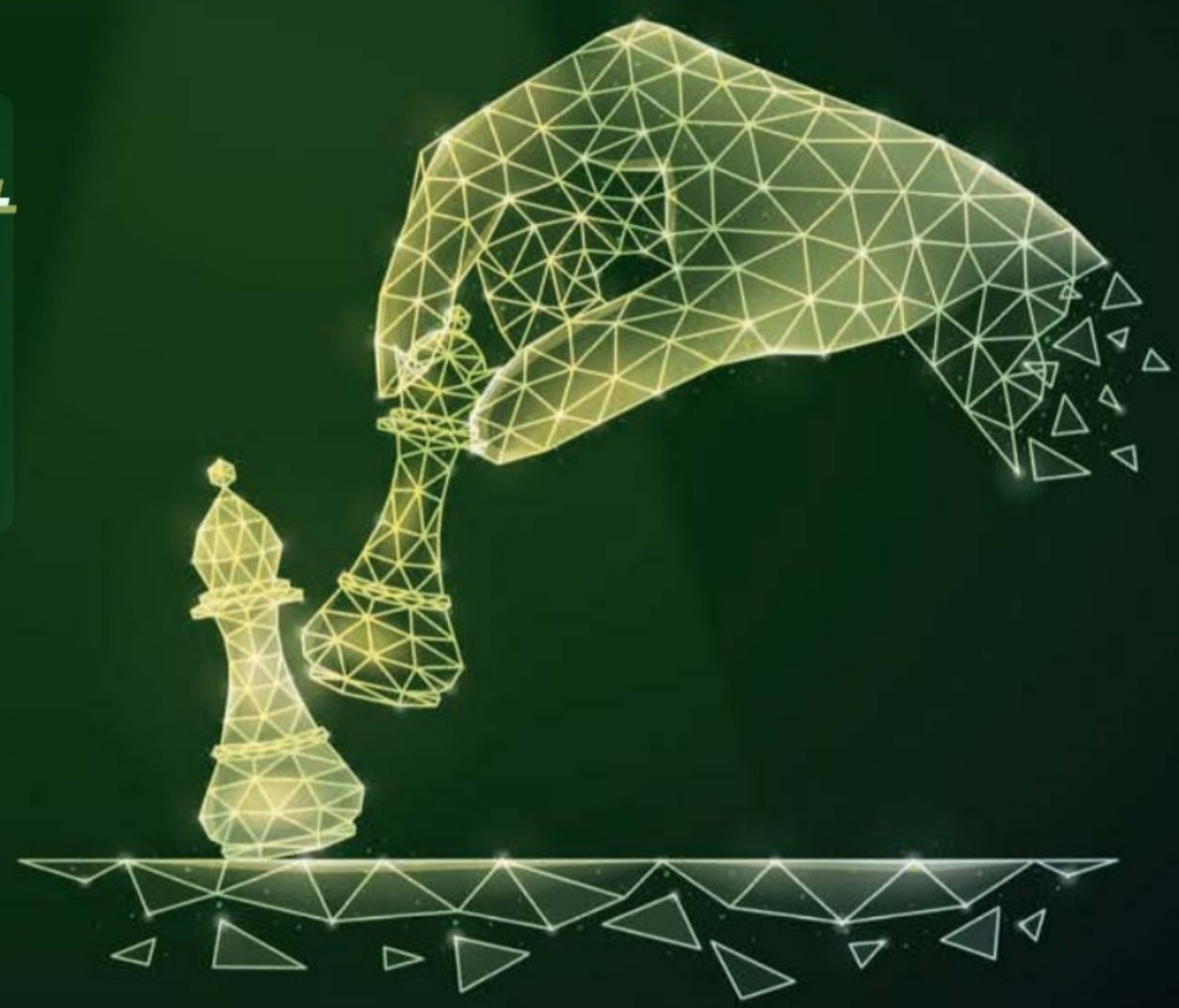
L'expertise de Viry-Châtillon en sport automobile est largement reconnue et s'exprime également au travers des différents programmes existants amenés à se renforcer. À savoir : le championnat du monde d'endurance (WEC), la compétition client ou encore la Formule E et le Rally-Raid pour des marques partenaires. Suite au processus de consultation et au dialogue mené avec les partenaires sociaux de Viry-Châtillon, Alpine a pris la décision de créer une cellule de veille F1 qui visera à conserver la connaissance et les compétences des collaborateurs sur cette discipline sportive, et à se maintenir à la pointe de l'innovation, au service des différents projets d'Hypertech Alpine.

Le projet, par son ampleur et son ambition sur les technologies d'avant-garde, permettra de soutenir la stratégie de croissance de la marque et ses objectifs sans précédent d'élargissement de sa gamme de véhicules, avec 7 modèles à l'horizon 2030. ☐ **B.G**

SPRINKLR PARTENAIRE POUR **REFONDRE L'EXPÉRIENCE DES FANS**

Sprinklr est partenaire officiel et fournisseur de plateforme de gestion unifiée de l'expérience client (Unified-CXM) pour l'équipe et la marque Alpine au sens large. L'écurie s'appuiera sur les quatre suites de produits Sprinklr : Sprinklr Service, Sprinklr Insights, Sprinklr Marketing et Sprinklr Social. Sprinklr aidera BWT Alpine F1 Team à repenser l'expérience fans avec les fonctionnalités suivantes :
La plateforme Unified-CXM de Sprinklr permettra à l'équipe d'éprouver et de pérenniser sa stratégie marketing, service clients et engagement pour rester compétitive et faire évoluer en permanence ses expériences.

Un débat sans fin



par Bertrand Garé

Avec l'immixtion de l'intelligence artificielle partout dans nos vies, le sempiternel débat entre innovation et réglementation refait surface avec d'autant plus de virulence que de gros intérêts sont en jeu.

Le premier est stratégique, il est nécessaire que les États maîtrisent et s'appuient sur l'intelligence artificielle. Il suffit de voir comment celle-ci s'est imposée dans le conflit russo-ukrainien. Ce côté militaire n'est pas à négliger, car il est le pendant de la recherche et de l'utilisation de l'intelligence artificielle.

Deuxième point, les investisseurs de tous poils ont investi énormément d'argent dans des start-ups ou des entreprises plus matures qui se sont lancées sur les technologies d'intelligence artificielle générative et ses derniers développements avec les agents autonomes qui fournissent un véritable coup de fouet à la productivité.

Le troisième point concerne la protection de nos vies. Pendant des années la formule *« si c'est gratuit, c'est toi le produit »* a fait le bonheur des GAFAM en leur fournissant gratuitement de plus en plus de données parfois sans véritablement prévenir les utilisateurs que leurs données seraient utilisées pour optimiser le placement de publicité dès qu'ils arrivent sur un site ou qu'un moteur de recherche soit capable de vous lister tous les déplacements que vous avez durant le mois passé. Avez-vous déjà demandé à Google de vous rappeler que vous avez été ici ou là durant une certaine période ?

Quatrième point, et c'est d'importance, quid de l'innovation ? Doit-on faire de l'innovation pour de l'innovation ? Des réformes pour des réformes ? Allons-nous, avec l'IA, revivre les atermoiements d'Oppenheimer après qu'il a, en partie, mis au point le moyen le plus radical de finir une guerre en trucidant des milliers de Japonais. Si l'IA générative prouve sa valeur, doit-on laisser faire tout et n'importe quoi sur l'autel de l'innovation et du progrès et ne pas s'interroger sur les aspects qui pourraient être néfastes de cette technologie. Déjà, avec les agents autonomes capables de réaliser seuls des tâches complexes, il va bien falloir préciser ce que sont les tâches à valeur ajoutée qui seront réservées à l'humain dans le futur.

La conjonction des quatre points précités fait qu'aujourd'hui dominent deux pays : les USA et la Chine qui ont investi et investissent encore massivement dans l'intelligence artificielle. Leur but est évidemment d'obtenir un avantage, pas seulement concurrentiel sur un ennemi potentiel pour le premier point. Du fait des investissements massifs, tout ce qui pourrait ralentir les nouveautés dans le domaine est vu comme un frein au développement, donc à des gains futurs ou une atteinte à la « sécurité nationale ».

Pour le troisième point, il est clair que les utilisateurs ou toute personne allant sur Internet n'est là que pour fournir ses données pour alimenter la vaste machine permettant d'affiner les modèles et que son droit à une vie réellement privée ne peut que devenir désuet.

Que viennent faire les européens avec leur IA Act ou le gouverneur de Californie qui vient de sortir 19 textes de loi limitant ce qu'il est possible de faire avec l'intelligence artificielle ? En fait, cet IA Act fournit aux développeurs et aux déploiements d'IA des exigences et des obligations claires en ce qui concerne les utilisations spécifiques de l'IA. De manière logique, le texte propose donc d'interdire les systèmes d'IA visant à exploiter les vulnérabilités *« dues à l'âge ou au handicap physique ou mental »* d'un groupe de personnes afin d'altérer leur comportement susceptible de causer un préjudice (physique ou

psychologique) seront interdits. La loi sur l'IA s'applique aux systèmes d'IA mis sur le marché ou mis en service, ou aux modèles d'IA à usage général mis sur le marché, dans l'UE ou dans l'Espace économique européen (EEE), que les fournisseurs soient établis ou situés dans l'UE ou l'EEE, ou dans un pays tiers. Face à la force de frappe des géants de l'intelligence artificielle, ce texte semble nécessaire alors que des états s'immiscent dans les processus d'élections afin de choisir le vainqueur qu'ils souhaitent. Les ingérences étrangères dans les élections américaines ne sont un mystère pour personne.

Là où le bât blesse, vient du fait que les européens ou certains états comme la Californie s'inquiète des possibilités de l'intelligence artificielle. Les autres pour l'argent ou pour le pouvoir sont donc prêts à faire de nos vies l'aliment de choix des modèles fondamentaux pour mieux nous vendre des produits, nous expliquer ce que l'on doit faire, penser, créer... Bref, un meilleur des mondes à la carte Open AI.

L'IA ACT est pourtant par sa portée le seul moyen pour l'Europe de contrer cette vision du monde en obligeant ces acteurs à avoir, au moins sur notre continent, un comportement éthique et respectueux des lois dont ils se contrefichent. Si Meta après une plainte de l'équivalent de note CNIL en Irlande lui a fait plier le genou en lui interdisant d'entraîner ses modèles avec les données présentes sur ses réseaux sociaux, les frasques de l'ancien CEO de Stability AI ou les procès entre entreprise du secteur démontrent bien qu'une réglementation est nécessaire dans un secteur devenu un nouveau far-west.

Dernier point à soulever, celui de la productivité. Donc, en moyenne, l'intelligence artificielle ferait gagner cinq heures par semaine pour 60 % des salariés utilisateurs de cette technologie. Selon le Boston Consulting Group, ce temps gagné contribue notamment à l'amélioration de la qualité du travail et est utilisé pour effectuer davantage de tâches (41 %), expérimenter la technologie (39 %) ou encore travailler sur des missions stratégiques (38 %). Donc, seulement 41 % des développeurs développeraient plus, des vendeurs passeraient plus de coups de téléphone pour contacter des clients. Ainsi, les améliorations apportées par la GenAI permettent non seulement de réduire la charge de travail, mais aussi d'améliorer la qualité globale des missions réalisées. Malgré une confiance accrue, l'IA génère également des inquiétudes. Près de la moitié des utilisateurs réguliers (49 %) craignent que leur emploi ne disparaisse dans les dix prochaines années à cause de l'IA. Cette peur n'est partagée que par 24 % des salariés qui n'utilisent pas ces technologies. La question est donc, gagner du temps, pour quoi faire ?

Deux économistes sur le site Telos dans un article fort argumenté indique : *« Si de nombreux intervenants dans ce débat avancent que l'IA et l'IAG devraient être la source de gains de productivité très significatifs, comme par exemple ceux associés à la seconde révolution industrielle (celle de l'électricité et du pétrole) qui a produit des effets tout au long du XXe siècle, d'autres sont beaucoup plus prudents. Ainsi, dans une étude très récente, Daron Acemoglu (2024) avance que les gains de productivité à attendre de l'IA pourraient, en cumul sur les dix prochaines années, être de l'ordre de 0,5 point de pourcentage, soit 0,05 point de pourcentage par an. Aghion et Bouverot estiment dans leur rapport que les gains de productivité à attendre de l'IA pourraient être de l'ordre de la moyenne de ceux associés dans le passé à la seconde révolution industrielle et à l'émergence des TIC, c'est-à-dire d'environ 10 % sur les dix prochaines années »*. Dans la suite de l'article ils, précisent, *« les gains de productivité induits par l'émergence et la diffusion universelle des TIC apparaissent très faibles, sinon négligeables dans la très grande majorité des pays avancés, pourtant grands utilisateurs de telles technologies »*. □

pfSense Une solution de routeur et firewall virtuels

pfSense est un système léger open source offrant des fonctionnalités de pare-feu, routeur et VPN faciles à installer et à configurer. Il est très prisé des entreprises de toutes tailles qui recherchent un outil dédié sécurité réseau simple, efficace et complet. Nous allons voir dans cet article comment l'installer et réaliser une configuration de base.

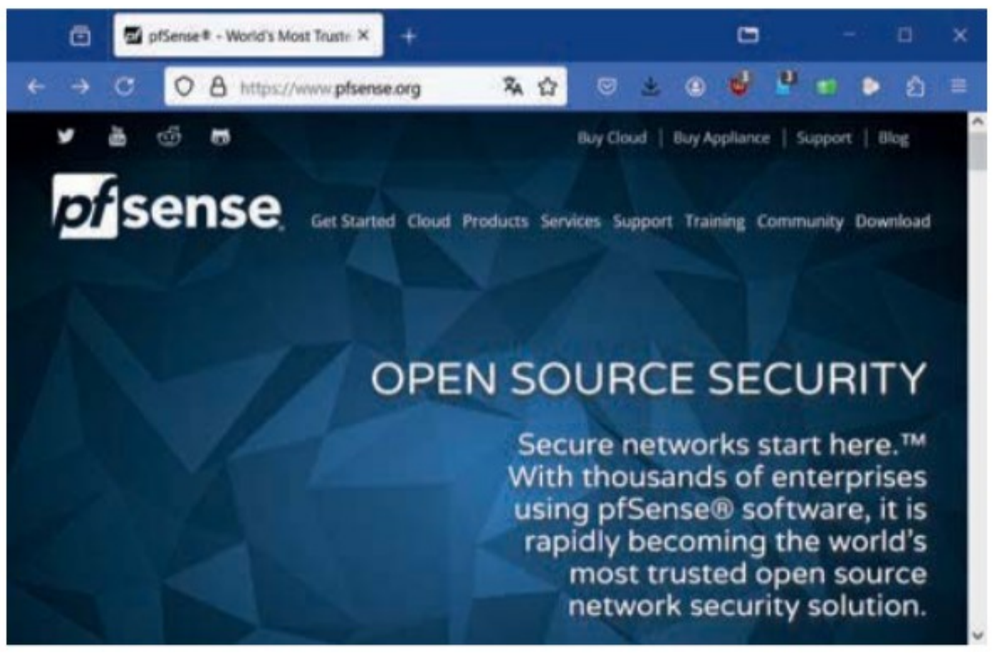
pfSense est un pare-feu open source faisant également fonction de routeur développé par Rubicon Communications et Netgate. Il est basé sur le système d'exploitation FreeBSD issu de la famille Unix (et non Linux). C'est à l'origine un fork de m0n0wall utilisant le pare-feu à états Packet Filter ainsi que des fonctions de routage et de NAT (translation d'adresse réseau) lui permettant de connecter plusieurs réseaux informatiques.

Caractéristiques de pfSense

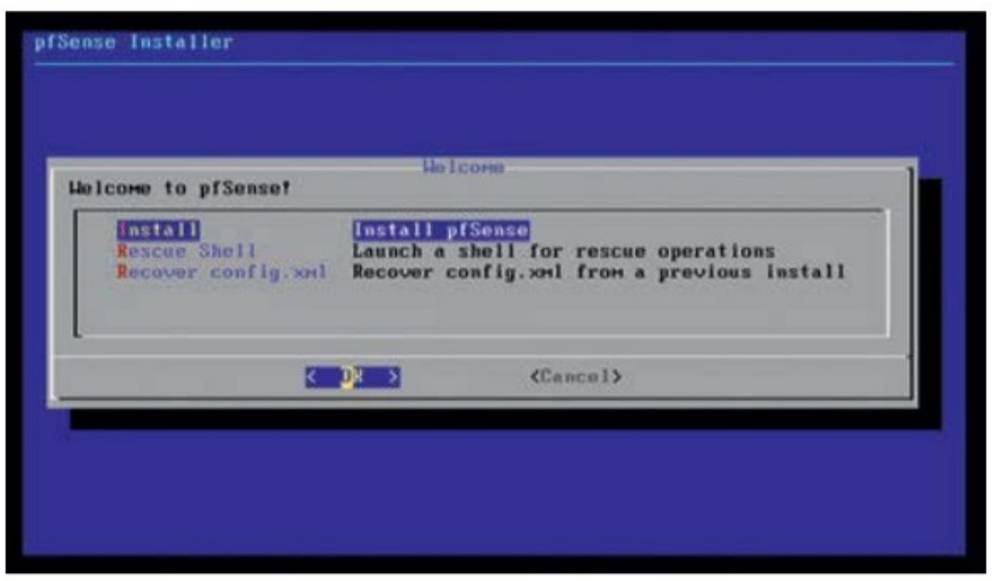
Il comporte l'équivalent libre des outils et services utilisés habituellement sur des routeurs professionnels propriétaires. pfSense peut convenir aussi bien pour un réseau domestique que professionnel. Ce logiciel peut s'exécuter aussi bien sur des ordinateurs physiques ou virtuels que sur des boitiers firewall et/ou routeurs de style appliance et open hardware pouvant accueillir presque n'importe quel système d'exploitation. Il offre un large éventail de fonctionnalités quasiment similaires à ce que proposent les périphériques commerciaux du genre. Il supporte également d'autres solutions tierces parties telles que Squid ou Snort pour accroître encore ses capacités. pfSense inclut un firewall, un point d'accès sans fil, un routeur, un point de terminaison VPN, un serveur DNS, un serveur DHCP, un équilibreur de charge, un contrôleur de flux et un filtre de contenu Web.

Installation de pfSense

La version utilisée ici est la dernière publiée, la 2.7.2. Vous pouvez vous la procurer à l'adresse https://www.pfsense.org/download/. Rappelons que le rôle d'un routeur est de faire communiquer différents réseaux entre eux. Un pare-feu (ou firewall comme dirait Shakespeare) est un système de sécurité matériel (on parle alors d'appliance) et/ou logiciel qui permet de définir et de contrôler les flux de données qui sont autorisés à entrer et sortir de votre réseau. pfSense s'utilise via une interface web. Il faut par conséquent disposer d'un poste client situé sur le même réseau avec un navigateur web. C'est la seule contrainte. Après, peu importe le système d'exploitation (Linux, Unix, Mac ou Windows, Android, Raspbian ou autre) ou le navigateur web. C'est encore un énorme avantage de pfSense qui montre bien son côté ouvert. Pour notre part, nous avons utilisé un PC sous Windows 10 comme hôte des machines virtuelles, Oracle Virtual Box comme logiciel de virtualisation et nous avons créé deux machines virtuelles que nous avons bien évidemment mises sur le même réseau local afin qu'elles puissent communiquer. Sur l'une d'elles, nous avons installé pfSense avec l'ISO fourni par le projet et, pour communiquer avec elle, installer et utiliser le client web de pfSense, nous avons créé une VM Ubuntu 22.04. Encore une fois, cette dernière machine aurait pu s'exécuter sur n'importe quel système d'exploitation du moment qu'elle peut faire tourner un navigateur web. Lancez votre logiciel de virtualisation et commencez par installer pfSense sur une VM. Vous pouvez aussi installer d'abord la machine qui hébergera le client web. Peu importe l'ordre, tant que cette VM se trouve sur le même réseau local que l'interface LAN de pfSense. Vous pouvez, si vous n'avez pas bien configuré la partie réseau, y revenir et la modifier ensuite. Ce n'est pas vraiment un souci. Quel que soit le logiciel de virtualisation et le système d'exploitation choisi, vous allez devoir utiliser l'ISO de pfSense téléchargé sur son site. Le setup va démarrer

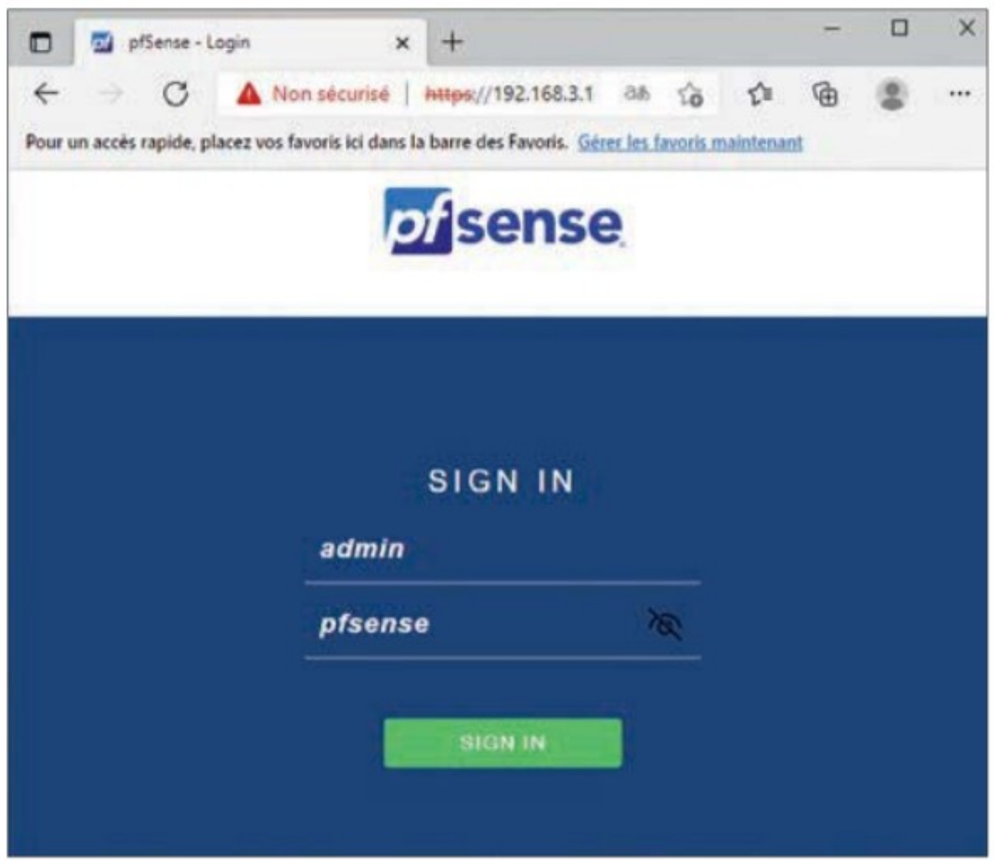


Vous pouvez télécharger des iso de pfSense pour les différentes architectures existantes sur le site du projet à l'adresse pfsense.org https://www.pfsense.org/



Vérifiez que vous êtes bien sur Install. L'option doit être sélectionnée en bleu foncé comme dans l'image ci-dessus, sinon déplacez-vous avec les flèches de votre clavier et appuyez sur Entrée pour valider.

automatiquement après quelques secondes. La configuration requise pour la vm pfSense est assez légère : 1Go de Ram et 20 Go d'espace disque. Vous pouvez utiliser VMWare Workstation, Oracle Virtual Box, Hyper V, KVM, Proxmox ou n'importe quel autre logiciel de virtualisation. Après avoir créé la vm et chargé l'ISO dans le lecteur DVD virtuel, il vous faudra booter dessus pour lancer l'installation. La souris n'étant pas disponible, l'installation va devoir s'effectuer au clavier uniquement. Appuyez sur la touche Entrée pour Accepter. Vérifiez que vous êtes bien sur Install (l'option est en bleu foncé si elle est sélectionnée). Dans le cas contraire, déplacez-vous avec les flèches de votre clavier. Validez en appuyant sur Entrée. Le programme d'installation vous propose de partitionner le disque de stockage de la machine, qu'il soit physique ou virtuel. Pour un labo sur machine virtuelle, nous choisirons généralement l'option par défaut. Libre à vous de faire autrement si vous le souhaitez, et surtout si cela vous apparaît nécessaire. Allez sur Auto (UFS) pour l'option par défaut et validez en appuyant sur Entrée. Pour confirmer que vous voulez utiliser le disque entier pour installer le système d'exploitation, placez-vous sur Entire Disk et appuyez encore sur Entrée. Validez jusqu'au bout pour lancer l'effacement du disque et l'installation du système d'exploitation de pfSense. Il vous sera ensuite proposé d'ouvrir un shell pour apporter d'éventuelles modifications en ligne de commande. Si ce n'est pas le cas (normalement non), déplacez-vous sur Reboot et appuyez sur Entrée. Au démarrage, pfSense va se lancer, tester et configurer les services dont il a besoin : la présence de l'interface WAN (Configuring WAN interface), celle de l'interface LAN, le lancement du service DNS (Configuring DNS Resolver) et du DHCP. Il reste une dernière petite chose à faire avant de passer sur l'interface web de pfSense pour la configuration finale : assigner la bonne adresse IP à l'interface LAN, c'est-à-dire celle qui correspond à notre réseau local. Celle-ci doit être sur le même réseau que votre réseau local et vous la ferez terminer par 1 ou 254 pour respecter la tradition. Si votre réseau est sur le 192.168.3.0/24, par exemple, vous pouvez assigner l'adresse 192.168.3.1. La dernière question concerne le protocole utilisé pour aller sur l'interface web. Il est en https par défaut.



Depuis un PC, physique ou virtuel, mais présent sur le même réseau local, ouvrez un navigateur internet pour accéder à votre routeur/firewall pfSense en saisissant son adresse IP. Les identifiants par défaut de pfSense sont admin pour le Login et pfSense pour le mot de passe.

Rien ne vous empêche de le passer en http. Tout dépend de votre contexte et du niveau de sécurité que vous souhaitez. Pour un labo de réseau virtuel, http sera bien suffisant.

Configuration via le client web

La configuration de pfSense en ligne de commande étant terminée, vous pouvez passer à l'interface web. Depuis un PC, physique ou virtuel, mais présent sur le même réseau local, ouvrez un navigateur internet et accédez à votre routeur / firewall pfSense en saisissant son adresse IP. Les identifiants par défaut de pfSense sont admin pour le Login et pfSense pour le mot de passe. Les deux dernières options de la page définissent que tout trafic entrant sur l'interface WAN et venant d'une classe d'adresse réseau privé est automatiquement bloqué. Dans le cas d'un labo virtuel où vous souhaitez faire communiquer des réseaux privés sans utiliser une adresse publique, il est nécessaire de décocher les 2 cases attenantes pour éviter tout problème. Et vu le contexte, cela ne créera pas de problème de sécurité puisque vous resterez dans un cadre fermé. Si vous n'avez rien de particulier à modifier sur l'interface WAN, vous pouvez poursuivre. L'assistant de pfSense passe alors à l'interface LAN. Vous pouvez par exemple changer l'adresse IP de l'interface LAN de pfSense en choisissant bien entendu une adresse qui est toujours sur le même réseau local. Il est conseillé de changer les identifiants par défaut du compte admin de pfSense, mais à vous de voir. Pour un labo virtuel sur une machine privée et en réseau fermé, cela n'a guère d'importance. Dans le cadre d'une utilisation professionnelle où les machines seraient accessibles depuis un réseau plus global, c'est totalement indispensable. □

T.T

TABLEAU DE BORD DE PFSENSE

Vous retrouvez dans le tableau de bord de votre machine pfSense des informations sur l'utilisation des ressources de la machine, ses différentes adresses IP, sa version, ses mises à jour et autres. Vous allez pouvoir ajouter des graphiques, des informations sur les load balancer, le trafic, les logs et tutti quanti. Vous pouvez mettre en place des VPN (IPSEC, OpenVPN,...), activer des services (DHCP, DNS, NLB, NTP, WOL,...), faire du NAT et du port forwarding, ajouter des routes statiques, définir des règles pour le trafic entrant et sortant ou ajouter des plugins comme Squidguard pour le filtrage ou Ntopng pour le monitoring réseau. Par défaut, juste après l'installation, tout le trafic est ouvert. Il est généralement conseillé de le brider. C'est le but d'un firewall : sécuriser ce qui entre et sort du réseau. Là encore, si vous installez pfSense en mode test sur une machine et surtout sur un réseau fermé, voire interne, ce n'est pas très grave. Dans tous les autres cas, c'est indispensable.

La part secrète d'Oracle Cloud

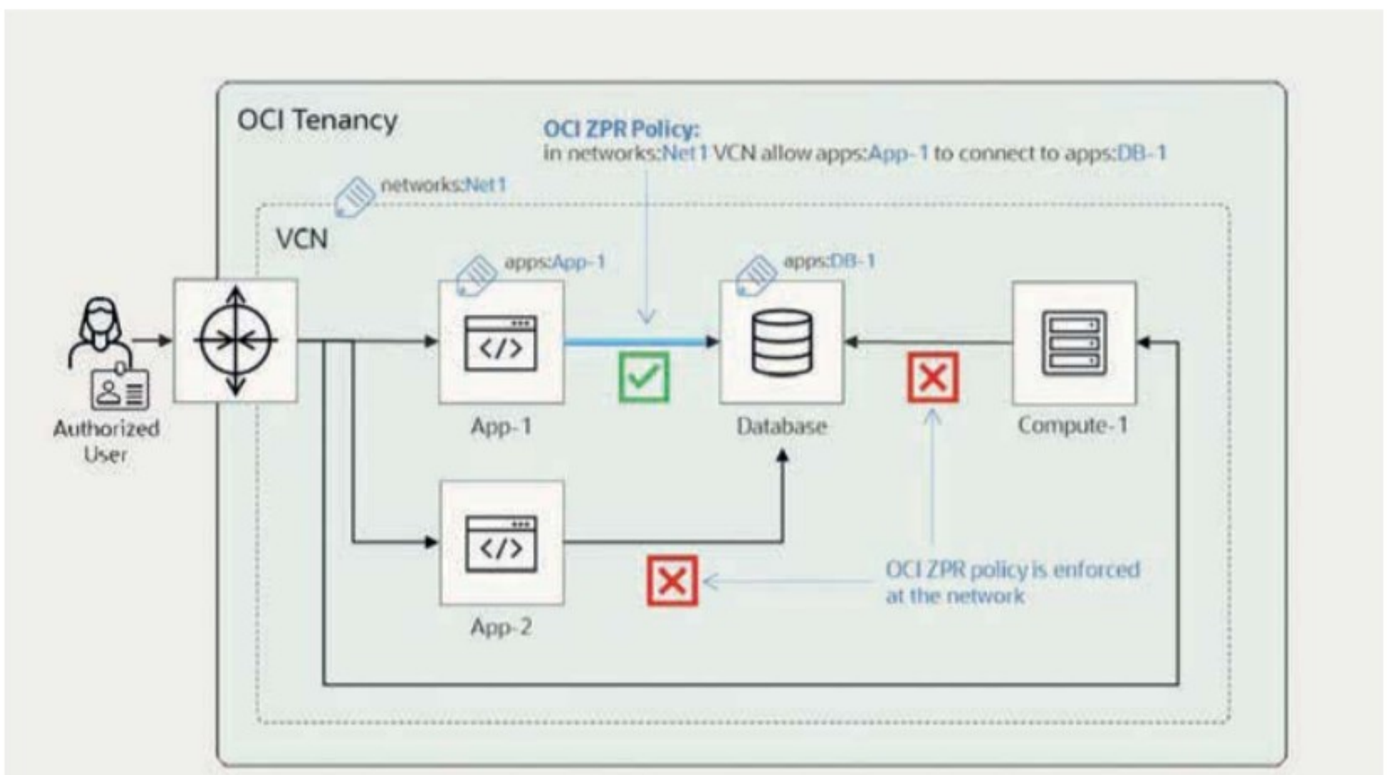
Mahesh Thiagarajan, EVP d'Oracle Cloud Infrastructure, est revenu sur les annonces faites par Larry Ellison autour du réseau et de la sécurité pour apporter des éclaircissements sur le travail accompli et les différenciateurs d'Oracle Cloud.

Construit sur la fabrique réseau d'OCI, Zero Trust Packet Routing (ZPR) permet de découpler la configuration du réseau de sa sécurité pour limiter les erreurs humaines. Par analogie, le logiciel fonctionne comme une couche de virtualisation boostée à l'IA. S'appuyant sur un projet de développement d'un nouveau standard ouvert par Applied Invention et d'autres entreprises en 2023, Zero Trust Packet Routing autorise les utilisateurs à définir des attributs de sécurité sur des ressources et d'écrire des politiques de sécurité en langage naturel pour limiter le trafic réseau à partir des ressources disponibles et des services de données accédés. Cela permet principalement de limiter les cas de mauvaises configurations réseau. C'est la première fois que le standard est mis en œuvre sur un Cloud.


Mahesh Thiagarajan, EVP d'Oracle.

Concrètement ZPR, les données ne peuvent transiter que sur une permission explicite, offrant ainsi une meilleure protection autour des données sensibles. La sécurité est mise ainsi en œuvre au niveau du réseau, ce qui évite les problèmes d'une mise en œuvre classique comme le temps nécessaire à réaliser ces opérations. La sécurité est déportée vers l'équipe réseau. Tout transfert non inscrit dans une politique sera donc banni et sera contraint au niveau réseau.

Ainsi, les équipes réseau ou de sécurité peuvent réduire l'accès aux données sensibles en leur attribuant un seul chemin possible, réduisant la surface d'attaque et se protégeant des exfiltrations de données par des identités compromises. Les équipes ont la possibilité de prouver plus facilement que les bonnes politiques de sécurité ont été appliquées aux équipes d'audit. Enfin, il n'est plus forcément nécessaire de mettre en œuvre une protection spécifique pour le réseau.



Un exemple du fonctionnement de ZPR.

Les autres différenciateurs d'OCI

Mahesh Thiagarajan ajoute que la performance est aussi une des forces de la plateforme et que tout ne vient pas du réseau. Un ensemble d'innovations et d'éléments d'architecture concourt à cette performance qui dénote par rapport à la concurrence. OCI profite ainsi d'une architecture spécifique qui ne décharge pas les CPU tout en travaillant au-dessus d'une architecture NUMA (Non uniform memory access). Il ajoute aussi le travail accompli sur le service de stockage ou au niveau du plan de contrôle avec des niveaux de services garantis.

Il revient ensuite autour du réseau et précise qu'il n'y a pas de sursouscription permettant ainsi lors d'un pic d'utilisation d'une application de ne pas connaître le phénomène de goulet d'étranglement sur le réseau. Cela est possible du fait d'une analyse réalisée au niveau du plan de contrôle avec le suivi du rapport du nombre de cœurs utilisés pour le réseau. De plus, Oracle continue en permanence d'améliorer le contrôleur réseau et des cartes réseaux pour assurer que le maximum de performance est fourni à chaque équipement mis en ligne sur le Cloud.

Les différenciateurs d'OCI sont à plusieurs niveaux, du réseau aux optimisations hardware et software afin de fournir la meilleure performance possible avec une architecture sans partage. □

B.G

Automatisation

Aruba étend les fonctionnalités de Networking Central

L'entité réseau chez HPE étend la couverture de sa solution Networking Central avec l'intégration d'OpsRamp et de nouvelles fonctionnalités Zero Trust de sécurité.

La solution s'enrichit de nombreuses nouveautés, dont un moteur de configuration des périphériques réseau amélioré, une observabilité réseau étendue et des optimisations générées par IA, basées sur la télémétrie des architectures des utilisateurs de Central. Le nombre de remontées auxquelles accède l'IA englobe désormais les informations télémétriques venant de plus de 1,6 milliard d'appareils uniques, connectés à 4,6 millions de matériels actifs réseau. Cet énorme apport d'informations permet d'augmenter les performances de manière prédictive, à partir de changements de configuration spécifiques, de simulations d'activation de fonctions de sécurité et d'optimisations du réseau en vue de générer des économies énergétiques, de garantir l'expérience des utilisateurs et de renforcer les politiques de sécurité. HPE Aruba Networking Central comprend maintenant l'ajout d'un modèle de configuration commun, d'une approche de développement pour l'automatisation via API, de nouvelles capacités de configurations hiérarchiques, d'une gestion plus fine de la géolocalisation en intérieur et de 90 nouvelles API.

La surveillance des appareils tiers, intégrée depuis l'acquisition d'OpsRamp par HPE en 2023, offre une visibilité sur les points d'accès, les commutateurs, les routeurs et les pare-feux des principaux fournisseurs, améliorant ainsi la maintenance prédictive dans des environnements divers et hétérogènes. Cette nouvelle option permet de réduire les angles morts et d'accélérer les tâches courantes de surveillance de l'état de santé et de dépannage du réseau. L'ajout de la fonction « capsule temporelle » permet aux opérateurs de visualiser l'état du réseau, d'un site, d'un appareil, d'un IoT, à tout moment, minute par minute, sur les sept derniers jours glissants.

Une vue d'Aruba Central.

Pour une meilleure expérience utilisateur

La solution intègre nativement HPE Aruba Networking User Experience Insight (UXI) qui apporte à l'IA une nouvelle source d'information sur la qualité d'utilisation perçue par les usagers du réseau. UXI, qui peut être installé dans des environnements HPE Aruba Networking ou non, permet de surveiller en continu le respect des niveaux de service (SLA) attendus, de l'utilisateur à l'application, dans une seule et même interface. Les fonctionnalités AI Networking de HPE Aruba Networking Central seront disponibles en avant-première publique à partir d'octobre 2024. La surveillance d'appareils tiers, l'intégration d'UXI et certaines fonctionnalités de configuration d'appareils seront ajoutées à l'accès public avant la fin de l'année 2024. □

B.G

ARUBA ÉTEND LE ZTNA AUX RÉSEAUX LOCAUX

Cette fonctionnalité de périphérie locale permet d'appliquer les mêmes politiques de contrôle d'accès définies pour le cloud directement sur les campus et les centres de données, offrant une expérience utilisateur homogène et une application cohérente, quel que soit l'emplacement de l'utilisateur ou la méthode de connexion.

La nouvelle solution NDR exploite la télémétrie du datalake de HPE Aruba Networking Central pour entraîner et déployer des modèles d'IA afin de surveiller et de détecter toute activité suspecte dans les appareils IoT vulnérables.

DBOS Une alternative à Kubernetes

**Le créateur de PostgreSQL et celui d'Apache Spark ont travaillé en collaboration afin de créer un OS « cloud » au-dessus d'une base de données distribuée.
Leur souhait était d'offrir aux utilisateurs une plus grande sécurité et une simplification de la gestion par rapport au combo Linux/Kubernetes si largement utilisé de nos jours.
Nous allons voir dans cet article ce qu'il en est.**

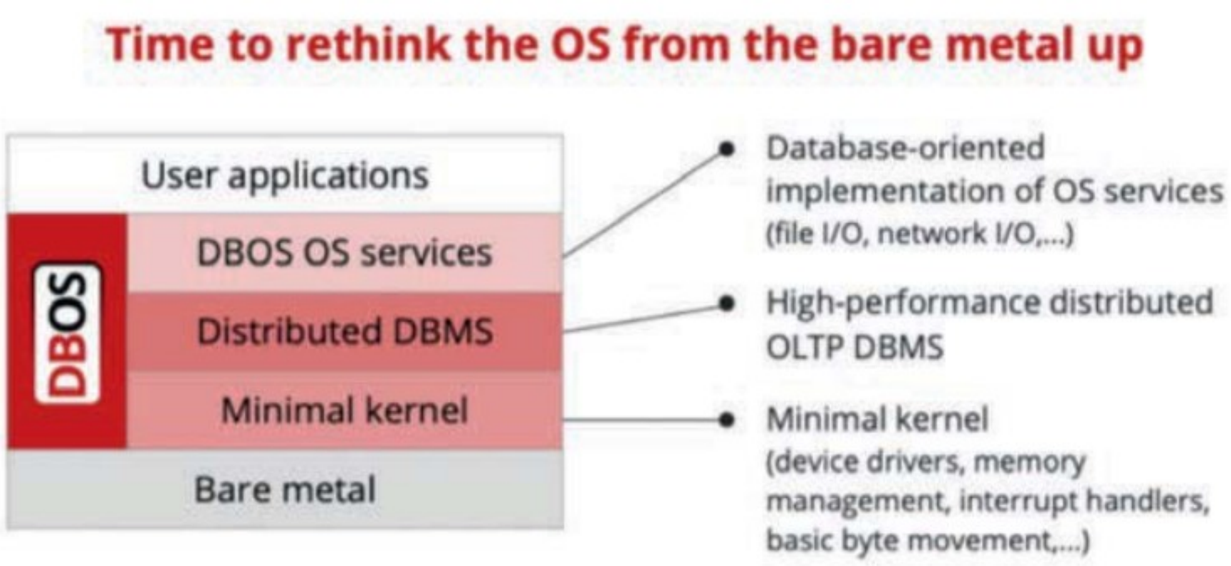
Le Dr. Michael Stonebraker est un célèbre inventeur de bases de données. C'est lui qui a créé il y a 40 ans le tout premier SGBDR, Ingres. Il a remis cela 10 ans plus tard avec PostgreSQL, un SGBD open source que l'on ne présente plus et qui fait de plus en plus d'ombre à Oracle. Toujours infatigable, il a participé plus récemment à la création de VoltDB, un système de base de données transactionnel en mémoire très innovant. Le sieur Stonebraker revient encore aujourd'hui sur le devant de la scène avec un système de gestion base de données conçu dans le but de remplacer la globalité de la pile (stack) informatique cloud native : DBOS (Database Operating System). *« Linux est trop vieux et Kubernetes trop complexe »*, a déclaré la start-up à l'initiative de ce projet. Une base de données a désormais été désignée pour remplacer ce combo Linux / Kubernetes. Pour que ce projet puisse émerger, DBOS Inc. a levé 8.5 millions de $ en financement de démarrage dirigé par Engine Ventures avec Construct Capital, Sinewave et GutBrain Ventures. Le projet a été élaboré par le Dr. Stonebraker avec le créateur d'Apache Spark et le co-fondateur et CTO de Databricks, Matei Zaharia, en collaboration avec une équipe conjointe de chercheurs en informatique du MIT et de Stanford. Que du beau monde…

Fonctionnalités de DBOS

DBOS exécute des services de système d'exploitation au-dessus d'une base de données distribuée haute-performance. Tous les états, logs et autres données système sont stockées dans des tables accessibles en SQL. Le résultat est un système de calculateur cloud très résistant à la montée en charge, tolérant aux pannes et cyber-résilient sans serveur pour des applications cloud native, du moins c'est ce qu'en disent ses créateurs. Il est vrai qu'avec un OS (Operating System ou système d'exploitation) s'exécutant au-dessus d'une base de données distribuée, vous obtenez plus facilement une grande tolérance aux pannes, la mise à l'échelle multi-nœuds et la gestion d'état. L'observabilité et la sécurité n'en sont que facilitées. Les couches containeur et orchestration disparaissent, offrant une solution plus simple et plus compacte. *« Vous écrivez moins de code car l'OS fait plus pour vous »* fait remarquer Stonebraker. Aujourd'hui, les systèmes distribués sont en grande partie construits sur un système d'exploitation (Linux pour ne pas le citer, quelle que soit la distribution) conçu pour s'exécuter sur un serveur unique. Cela résulte en un nombre incroyable d'états variables différents à gérer à travers la pile d'infrastructure (les données d'application, les systèmes d'authentification, la messagerie, la gestion des clusters, etc). Cette nature fragmentée nécessite bien entendu un nombre incalculable d'outils d'observation et de sécurité, vu que tous ces états fournissent à des hackers malicieux un terrain plus que fertile pour travailler leurs attaques. Et qu'est-ce qui peut gérer un million d'états aisément ? Une base de données, assurément. Le principe de conception qui a été suivi pour DBOS est qu'une OLTP (Online Transactional Processing) distribuée haute-performance implémente une suite de services de type OS. Elle doit s'exécuter sur un noyau OS minimaliste, avec le support de la gestion de la mémoire, des pilotes de périphériques, des gestionnaires d'interruption et des tâches basiques de gestion des octets (données enregistrées).

Une seule base de données pour tout gérer

En fait, ce n'est pas la première fois que ce type de solution a été évoqué. En 2001 déjà, Larry Ellison arguait que *« le middleware était une idée idiote et que tout devait être géré par la base de données elle-même. » « L'idée principale derrière ce projet DBOS provient d'une réflexion assez simple : ne pas perdre de vue que l'état du système d'exploitation devrait être le problème de la base de données »* a encore déclaré le Dr. Stonebraker. Zaharia, l'un des coconcepteurs, a fait remarquer que le service cloud Databricks pour Apache Spark gérait régulièrement un million de sous-tâches à la fois. Toutes les informations d'états et de planification étaient suivies dans une base de données PostgreSQL avec des performances exécrables très frustrantes pour l'équipe



Dans l'architecture DBOS, une OLTP distribuée haute-performance implémente une suite de services OS. Elle s'exécute sur un noyau OS minimaliste, avec le support de la gestion de la mémoire, des pilotes de périphériques et des gestionnaires d'interruption.

d'administration de Databricks. Le goulot d'étranglement de la base de données pouvait être résolu assez facilement. En fait, c'est ce que VoltDB voulait et pouvait faire : un traitement transactionnel concurrent ACID-compliant (Atomicité, Cohérence, Isolation et Durabilité) capable de se propager à travers de multiples serveurs. Le Docteur Stonebraker a été un des tout premiers utilisateurs d'Unix sur un PDP 11/40 possédant 48k de mémoire principale et 25MB d'espace disque. Un véritable luxe pour l'époque... Les états étaient alors conservés par le système d'exploitation Unix. Bien entendu, un million d'états représente un saut de six ordres de grandeur comparés au PDP. Mais *« le volume d'états dont le système d'exploitation devait conserver la trace restait et reste fondamentalement proportionnel aux ressources »*, toujours d'après le Dr. Stonebraker. DBOS a lui-même été testé au SuperCloud du MIT (https://supercloud.mit.edu/) avec plus de 32.000 processeurs, des téraoctets de mémoire principale et encore plus de téraoctets de stockage secondaire (ROM).

DBOS CLOUD, UNE BASE DE DONNÉES DISTRIBUÉE POUR LE SUPPORT TRANSACTIONNEL

Le premier service commercial construit autour de DBOS est DBOS Cloud, une plateforme FaaS (Functions as a Service) transactionnelle. DBOS, qui s'exécute sur le service open source Firecracker d'Amazon Web Services, est disponible pour les développeurs afin qu'ils se familiarisent avec le DBOS Cloud. Le service fournit les avantages suivants :

- **Le support pour les fonctions à état et les workflows**
- **La tolérance aux pannes intégrée avec garantie d'exécution en une et une seule fois**
- **Le débogage avec retour dans le temps aux états précédents**
- **Des données d'observabilité accessibles en SQL**
- **L'activation de la détection et de la récupération automatique après des cyberattaques**

DBOS Cloud conserve un suivi d'audit complet du code et du traitement des données et le stocke dans des tables SQL cryptées. Le DBOS Cloud Time Travel Debugger (débogueur temporel cloud DBOS) permet de rejouer les données, d'examiner les problèmes de dépannage rencontrés, d'assurer une conformité réglementaire ou de rechercher d'éventuelles tentatives de fraudes ou d'incursion.

Au sommet de la pile, il y a un système de base de données transactionnel distribué avec un système de fichiers, un moteur de planification et un système de messagerie.

Le Docteur Stonebraker a qualifié Linux de « leaky », voulant dire par là qu'il y a de trop nombreuses manières d'y introduire des vulnérabilités en matière de sécurité. De plus, construire un OS au-dessus d'une base de données a l'avantage d'offrir la possibilité de retourner à un état sain enregistré avant qu'une faille de sécurité n'ait été exploitée. Ce fonctionnement est comparable à celui d'une Apple Time Machine, le système de restauration d'Apple permettant de revenir dans le temps à un état stable du système, mais pour des serveurs dans ce cas et non de simples PC. *« Cette base de données centralisée va aussi beaucoup aider au débogage »* a encore déclaré le Docteur Stonebraker. La décomposition d'applications en de multiples micro-services, comme c'est la mode en ce moment, peut certes présenter certains avantages, mais les rend très difficiles à déboguer. Cela peut même complexifier la reproduction d'un comportement instable susceptible d'être rencontré de manière aléatoire entre plusieurs tests. Les bugs de ce genre sont connus sous le nom de « Heisenbugs » en référence au grand physicien Werner Heisenberg. *« Nous exécutons tous nos micro-services de manière transactionnelle. Ainsi, des micro-services parallèles sont ordonnancés par notre système de contrôle concurrentiel et, pour l'essentiel, il n'y a pas de Heisenbugs ou ils sont beaucoup, beaucoup plus difficiles à rencontrer »* a encore dit le Docteur Stonebraker. À l'origine, le système avait été simulé sur VoltDB, mais les bailleurs de fonds voulaient partir à la place sur un système clef-valeur open source. Du coup, ils avaient choisi comme base de données FoundationDB (https://www.foundationdb.org/), une base de données distribuée respectant les normes ACID. Le Docteur Stonebraker a reconnu que n'importe quel système OLTP distribué compatible d'un point de vue réseau avec PostgreSQL pouvait faire l'affaire, comme par exemple CockroachDB (www.cockroachlabs.com). □

T.T

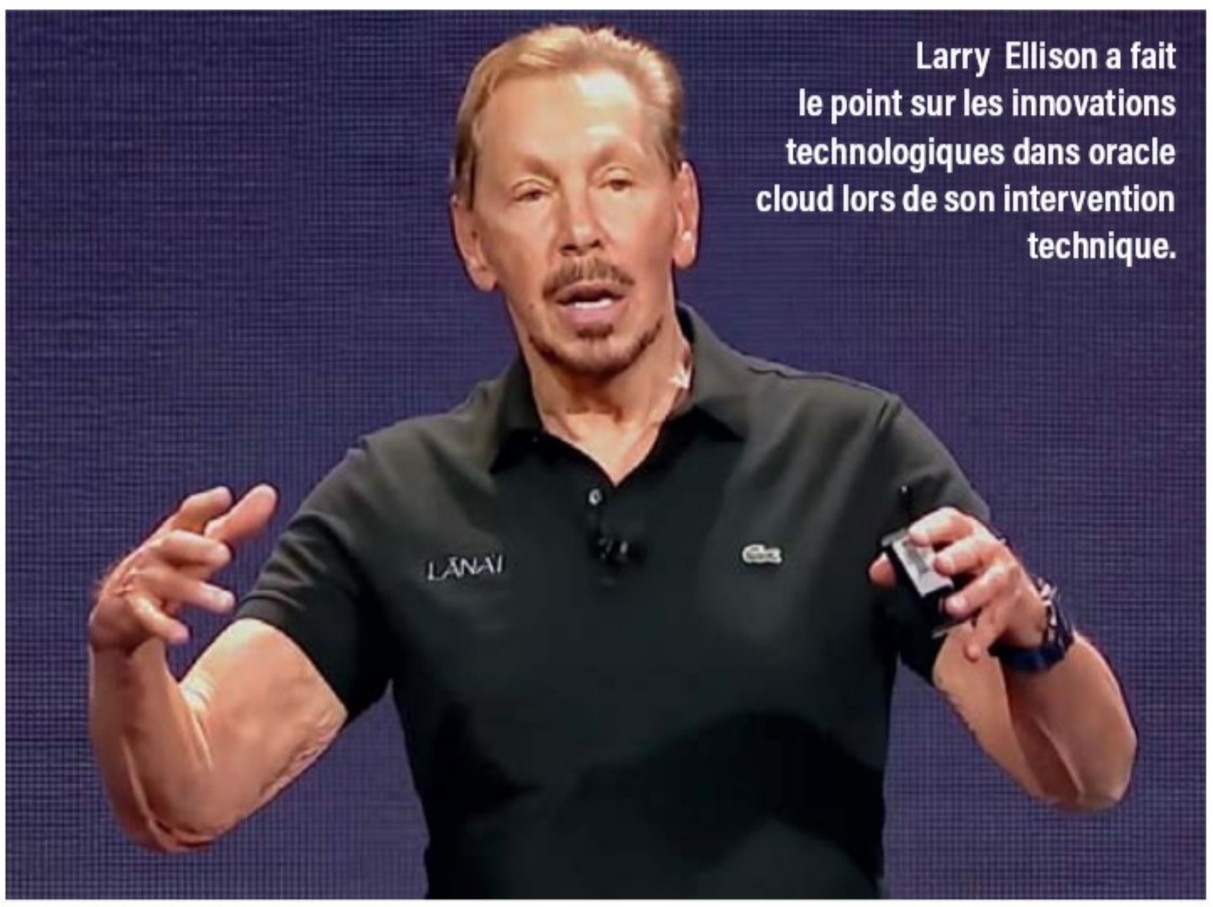
DBOS Cloud est la plateforme parfaite pour vos applications backend cloud native. Elle se déploie en un seul clic et supporte une mise à l'échelle de millions d'utilisateurs avec résilience.

Oracle CloudWorld

De l'IA, encore de l'IA… mais pas que !

Lors de la dernière conférence US d'Oracle, il a été beaucoup question de l'intelligence artificielle et de nombreuses annonces sont en rapport. Mais cela n'a pas été le seul sujet et l'éditeur continue sa route en mettant en avant le multicloud.

Larry Ellison, comme beaucoup d'autres grands patrons de l'industrie informatique, est persuadé du pouvoir transformateur de l'intelligence artificielle dans notre vie et dans notre manière de travailler. Les annonces lors du dernier Cloud World vont dans ce sens avec l'introduction d'un agent avec des fonctionnalités de RAG. Premier d'une série d'agents OCI GenAI, l'agent RAG fournit des fonctionnalités RAG prêtes à l'emploi, permettant aux clients de se lancer tout en évitant les processus manuels tels que la planification, l'extraction, le classement, la génération et l'intégration des agents. Il fournit également un auto-contrôle pour réduire les hallucinations, permettant aux clients d'adopter les technologies RAG pour rationaliser leurs processus métier sans passer par des cycles de recherche et développement. OCI GenAI Agents permet aux clients d'accéder à Oracle Database 23ai AI Vector Search et d'exécuter des requêtes de similarité rapide sur les données d'entreprise stockées dans la base de données. Pour les clients qui ont un abonnement à Oracle Database 23ai sur OCI, le service GenAI Agents ajoute une couche d'automatisation pour exécuter les fonctions de recherche RAG et de similarité sans avoir à déplacer les données vers une base de données vectorielle distincte. Pour les entreprises à la recherche de solutions open source pour leurs workloads d'IA générative, le service OCI GenAI Agents prend également en charge OCI Search avec OpenSearch. Oracle indique d'ailleurs différents cas d'usages pour lesquels l'agent pourrait avoir quelque utilités ! Analyse Vidéo, traduction et analyse de texte… L'outil donne de plus accès aux modèles de Meta Llama 3.1, dans des tailles allant de 405 milliards de paramètres pour les cas d'usage de l'IA qui nécessitent des fonctionnalités de pointe à 70 milliards de paramètres pour des workloads plus ciblés à un prix inférieur. En outre, OCI Generative AI prend en charge les modèles Cohere Command R, Command R+ et Embed.



Larry Ellison a fait le point sur les innovations technologiques dans oracle cloud lors de son intervention technique.

Pour développer et simplifier l'accès à l'IA

En s'appuyant sur la même base de données, DB 23ai, Oracle propose désormais une infrastructure de développement d'applications centrée sur l'IA. Elle fournit des technologies de développement qui permettent aux développeurs de générer rapidement des applications sophistiquées et aux applications d'utiliser facilement des interfaces en langage naturel alimentées par l'IA ainsi que des données centrées sur l'humain. GenDev, doux nom de la solution, associe des technologies dans Oracle Database 23ai, notamment les vues de dualité relationnelle JSON, la recherche vectorielle avec AI Vector Search et le low-code avec APEX, pour faciliter le développement à l'aide de l'IA générative. Dans GenDev, la complexité des données est gérée au niveau de la couche de données, et les règles de données d'applications, notamment l'intention, la confidentialité, la validation et l'intégrité, sont appliquées par le moteur de données. Pour ce faire, le moteur de données convergé d'Oracle, Database 23ai, prend en charge tous les types de données et workloads sans sacrifier la cohérence, les performances et la disponibilité transparentes des données dont les entreprises ont besoin.

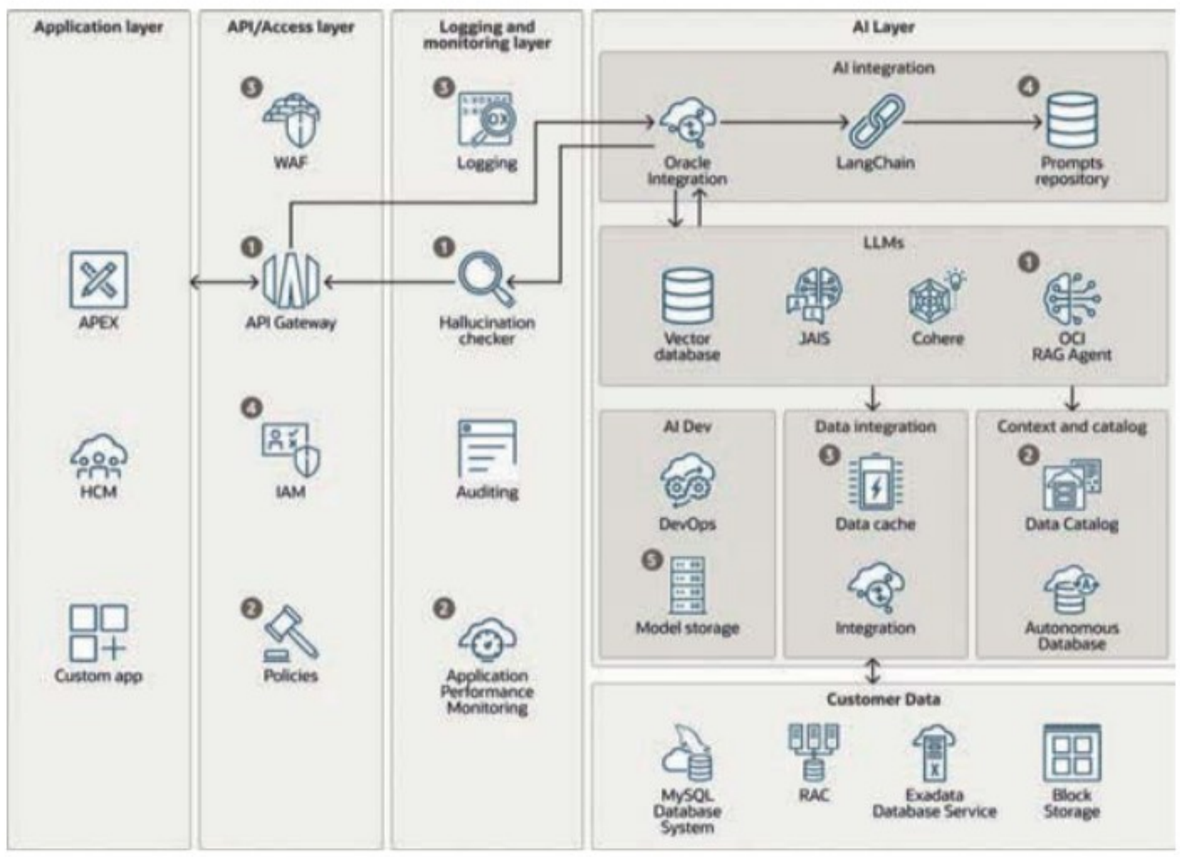
Ces capacités se retrouvent maintenant dans la dernière version d'Autonomous Database disponible sur le Cloud d'Oracle et d'Azure. Ce service démarre par la mise à disposition d'Exadata Database et d'Autonomous Database.

Un grand chez soi ou un petit chez les autres ?

Les développements sur le Cloud et le multicloud constituaient un autre axe majeur de la conférence. Christophe Négrier, en charge d'Oracle en France, précise cependant la définition d'Oracle du multicloud lors d'un point presse durant la conférence. Sur la demande de clients qui étaient déjà sur un cloud public mais qui étaient utilisateurs

d'Oracle en interne ou sur OCI, l'éditeur a développé une solution qui permet d'utiliser le cloud d'Oracle sur un autre Cloud comme Azure GCP ou maintenant AWS avec qui Oracle a annoncé un partenariat en ce sens. Ce service démarre par la mise à disposition d'Exadata Database et d'Autonomous Database « @AWS ».

Avec Oracle Database@AWS, les clients pourront notamment effectuer des analyses en temps quasi réel et exécuter des charges de travail de machine learning à faible latence, tout en s'intégrant de manière transparente à Amazon Simple Storage Service (Amazon S3), afin de faciliter et sécuriser les sauvegardes et la restauration des bases de données, dans le cadre d'une reprise d'activité après sinistre.



Comment créer une pile d'IA générative sur le Cloud d'Oracle.

Outre ce partenariat avec AWS, Oracle étend son empreinte sur Azure. Ainsi, OracleDatabase@Azure est maintenant disponible en France et dans cinq autres régions. Le service sera bientôt disponible dans quinze autres régions. Cette extension s'ajoute à de nouvelles fonctionnalités dont un service de protection des données entièrement géré pour les entreprises qui exécutent Oracle Exadata Database Service sur Oracle Database@Azure ; ce service permet aux entreprises de récupérer rapidement des données stratégiques en quelques secondes après une panne ou une attaque par ransomware. Un service de réplication de bases de données et d'intégration de données hétérogènes sera disponible sous peu comme service géré sur Oracle Database@Azure avec une parité complète des fonctionnalités vers OCI GoldenGate. OCI GoldenGate prendra en charge les intégrations avec Microsoft Fabric et OneLake, améliorant ainsi les intégrations existantes de banque de données et de messagerie Azure. Il en est de même avec Google Cloud où les clients bénéficieront pour la première fois d'un accès direct aux services Oracle Database exécutés sur OCI et déployés dans les data centers de Google Cloud. Ils peuvent désormais tirer parti de la base de données et de la technologie Exadata de pointe d'Oracle pour innover plus rapidement et développer de nouvelles applications. En outre, les clients peuvent exécuter des applications sur Oracle Linux, qui est désormais pris en charge par Oracle sur Google Cloud. Les images Oracle Linux peuvent être importées à l'aide du processus d'import d'image de disque virtuel de Google Cloud. Dans les douze prochains mois, les clients devraient également bénéficier d'un provisionnement plus simple des images Oracle Linux dans Google Compute Engine avec des images prêtes à l'emploi.


Safra A. Catz, CEO d'Oracle.

De nouvelles applications

Lors de la conférence, Oracle a aussi annoncé de nouveaux modules dans Fusion Cloud ou des améliorations de logiciels existants. Oracle Fusion Cloud Sustainability est une nouvelle application au sein d'Oracle Fusion Cloud Applications Suite qui aide efficacement les entreprises à gérer leurs initiatives de durabilité et à établir des rapports à ce sujet. Dans le domaine des RH, Oracle Dynamic Skills aidera les entreprises à élaborer, structurer et exécuter une stratégie de talents basée sur les compétences. Oracle Dynamic Skills, qui fait partie intégrante d'Oracle Fusion Cloud Human Capital Management (HCM). Le logiciel aide les entreprises, quel que soit le stade de leur parcours de compétences, à mieux comprendre et exploiter les compétences de leurs collaborateurs, à élargir leur accès aux talents et à prendre des décisions relatives à leurs effectifs plus éclairées. L'ensemble est évidemment dopé à l'intelligence artificielle.

De nouvelles fonctionnalités au sein d'Oracle Unity Customer Data Platform (CDP) qui fournissent aux entreprises des vues précises des comptes clients et des groupes d'achat afin d'optimiser les initiatives de croissance des revenus du marketing et des ventes. □

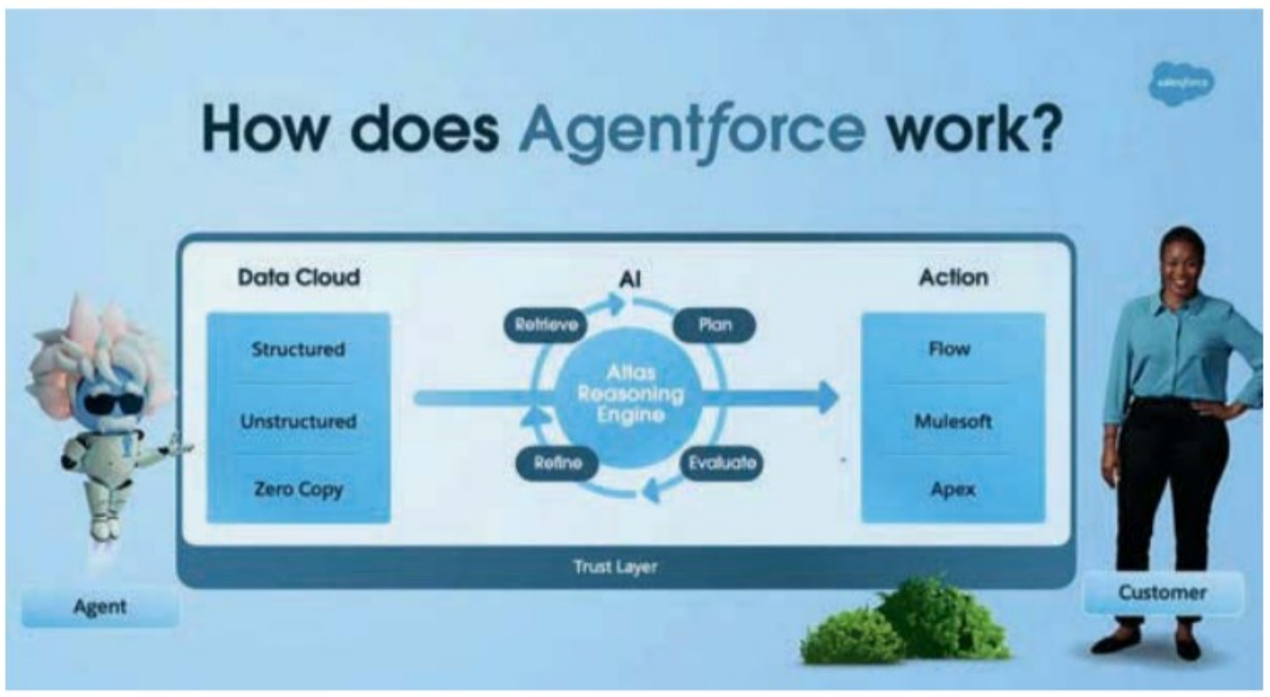
B.G

Dreamforce 2024

Agentforce, la SWAT Team de Salesforce

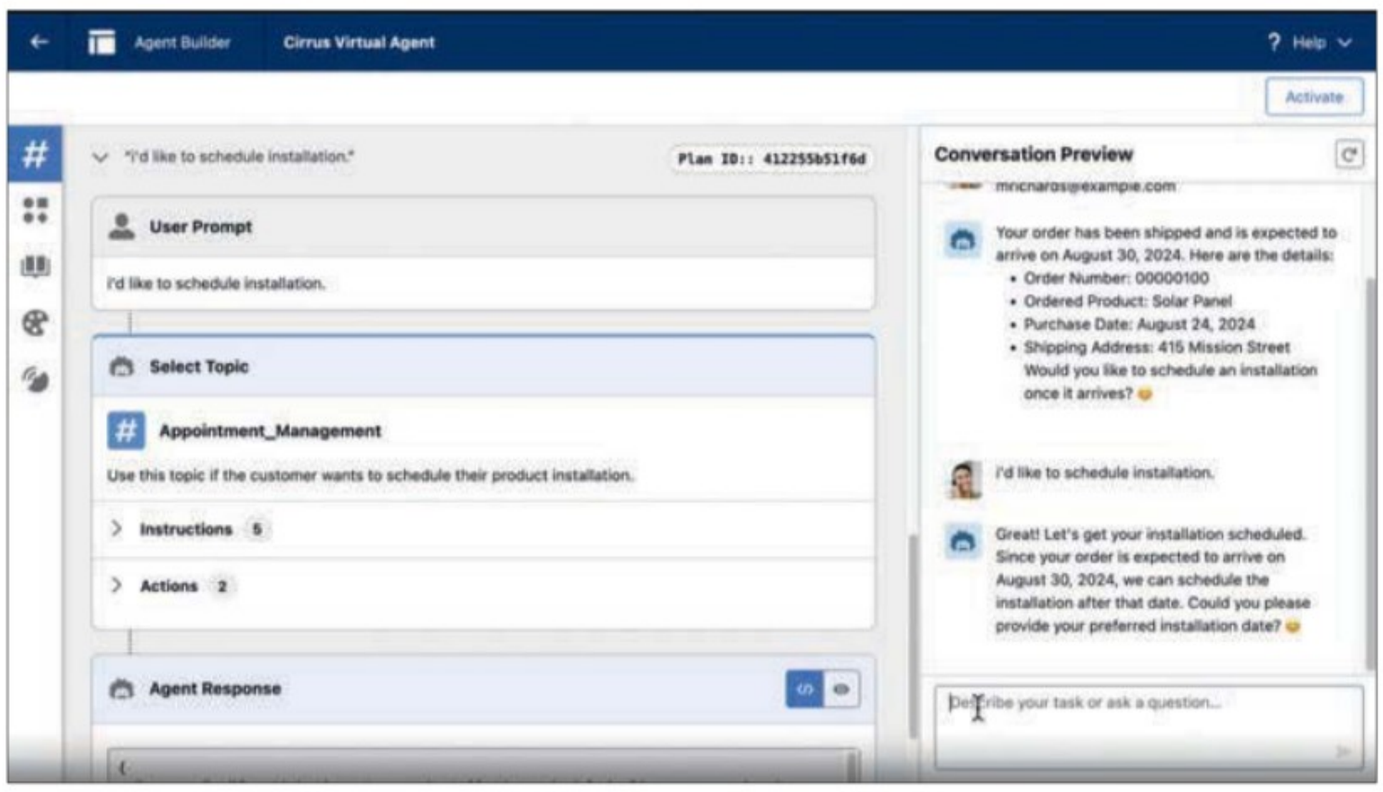
Lors de la conférence mondiale de Salesforce qui s'est tenue à San Francisco à la mi-septembre, l'éditeur de CRM a présenté Agentforce, sa troisième vague d'intelligence artificielle avec des agents autonomes et multitâches.

Si l'intention de Salesforce est d'augmenter la productivité et la performance des salariés, cette nouvelle vague d'agents autonomes rebat les cartes dans la manière de travailler puisqu'ils peuvent gérer des tâches de manière totalement autonome. Agentforce permet aux entreprises de dimensionner leurs effectifs à la demande en quelques clics. La main-d'œuvre numérique illimitée d'agents IA d'Agentforce peut analyser des données, prendre des décisions et agir sur des tâches telles que répondre aux demandes de service client, qualifier des prospects de vente et optimiser des campagnes marketing.



Le fonctionnement d'Agentforce.

Les agents sont capables de récupérer les données pertinentes sur demande, en élaborant des plans d'action pour n'importe quel type de tâche et en exécutant des plans d'actions sans nécessiter d'intervention humaine. À l'instar d'une voiture autonome, Agentforce utilise des données en temps réel pour s'adapter aux éléments changeants et fonctionne de manière indépendante en respectant scrupuleusement les garde-fous personnalisés du client. Lorsque nécessaire, Agentforce peut transmettre de manière transparente aux employés humains un résumé des interactions, une vue d'ensemble des détails sur un client, ainsi que des recommandations sur la marche à suivre.



L'interface d'un agent.

Atlas Reasoning

S'appuyant sur le Data Cloud de Salesforce, Atlas Reasoning Engine est le cerveau d'Agentforce. Il repose sur un système propriétaire qui lui permet de raisonner, de prendre des décisions et d'accomplir des tâches de manière autonome. Ce mélange de capacités de collecte d'informations basées sur l'IA et d'intégration étroite avec la plateforme Salesforce rend les réponses d'Agentforce plus pertinentes, solides et précises.

Le moteur se compose de six éléments. Query Refiner optimise les requêtes de manière itérative, améliorant ainsi la qualité des demandes des utilisateurs. Conversation Processor traite et analyse en permanence les données de conversations en direct pour en extraire des informations contextuelles pertinentes et améliorer ainsi le niveau de compréhension et la précision des réponses. Les requêtes des utilisateurs sont classées dans des rubriques prédéfinies pour assurer la pertinence des réponses. Par un processus de RAG (Retrieval Augmented Genaration), le moteur accède aux informations les plus pertinentes, même si elles proviennent de sources publiques. Output Evaluator évalue la qualité des réponses et permet de les affiner si nécessaire. Trust Layer, la couche de sécurité de Salesforce, applique de manière stricte les possibilités

d'accéder aux données. L'ensemble permet de limiter le phénomène d'hallucination de l'intelligence artificielle tout en fournissant des réponses fiables.

Des Agents prêts à l'emploi

L'éditeur propose de plus des agents préconfigurés pour certaines tâches pour simplifier le déploiement et l'utilisation de ces agents qui fonctionnent en permanence. Dans les processus de vente, un agent est spécialisé sur le service client avec une IA capable de gérer un large éventail de problèmes sans scénarios préprogrammés, ce qui améliore l'efficacité du service client. Les autres agents disponibles sont un coach commercial, un agent pour l'e-commerce, un agent pour les acheteurs B2B ou B2C, un optimiseur de campagnes marketing. L'ancien Einstein Copilot évolue en Agentforce, devenant un agent autonome capable de collecter des données, de raisonner et d'agir. Il s'agit désormais d'un agent d'assistance personnalisé et embarqué, qui soutient les employés avec des tâches spécifiques de manière parfaitement intégrée, en recherchant et en analysant des données, en créant des plans d'action et en les exécutant pour accroître l'efficacité globale sur le lieu de travail.

Marc Benioff lors de sa session plénière où il a présenté Agentforce.

Le rôle central de MuleSoft

Agentforce bénéficie d'une intégration transparente avec les capacités d'automatisation existantes de Salesforce, y compris une intégration étroite avec MuleSoft. À l'aide de Salesforce Flow, MuleSoft et de méthodes Apex, les clients peuvent facilement développer les fonctionnalités d'Agentforce en tirant parti de workflows et d'actions déjà conçus et optimisés. Ces modules leur permettent également de créer de nouvelles automatisations pour la solution. Qu'il s'agisse d'automatiser des processus complexes ou de déclencher des actions spécifiques à travers l'ensemble d'une entreprise, Agentforce peut exécuter ces automatisations sans effort. Les organisations peuvent ainsi capitaliser sur leurs investissements passés dans l'automatisation tout en développant de nouvelles capacités. Agentforce agira alors de manière autonome en respectant le cadre de confiance prédéfini, et offrira des résultats plus rapides sans qu'il ne soit nécessaire de créer de nouvelles automatisations ou intégrations à partir de zéro.

La force rapide

Salesforce propose de plus des outils low code pour personnaliser ou créer des agents avec Agent Builder et Model Builder. Prompt Builder permet aux utilisateurs de personnaliser facilement des modèles d'instructions prêts à l'emploi à l'aide de leurs propres données de CRM ou issues de Data Cloud, afin d'optimiser la qualité des résultats générés. L'expérience générative est intégrée de manière transparente pour l'utilisateur, que ce soit au sein d'un workflow automatisé, d'une page d'enregistrement Lightning ou même des actions d'un agent. □

B.G

UN RENFORCEMENT DES PARTENARIATS

Avec Agentforce, Salesforce a aussi renforcé ses partenariats autour de l'intelligence artificielle. Google, IBM et Nvidia assistent Salesforce dans l'enrichissement des fonctionnalités de la plateforme Agentforce et à répondre aux besoins spécifiques des métiers à travers des agents autonomes personnalisés.

Ainsi, IBM et Salesforce renforcent leur partenariat pour permettre aux entreprises de secteurs hautement réglementés de déployer des agents IA autonomes sans avoir à copier ou déplacer leurs données. Grâce à Agentforce, les organisations des secteurs financiers, manufacturiers, ou des télécommunications peuvent quant à elles déployer des agents IA autonomes et ainsi automatiser des processus complexes, tout en respectant les exigences de conformité spécifiques à leurs secteurs d'activité.

Salesforce et Google Cloud renforcent leur partenariat pour proposer une suite de productivité intégrant l'IA avec des agents IA autonomes disponibles dans Salesforce, Slack et Google Workspace. Les clients pourront désormais développer des agents IA personnalisés via Google Vertex AI et interagir avec ces agents directement dans Slack pour obtenir des réponses précises, automatiser des tâches et améliorer la productivité.

Grâce à l'intégration de NVIDIA AI Enterprise Platform et d'Agentforce, les entreprises auront accès à des agents capables de gérer des flux de travail complexes et de transformer l'expérience client et collaborateur. De plus, la collaboration entre Salesforce et NVIDIA vise à accélérer le traitement des données structurées et non structurées sur Data Cloud, en tirant parti de la capacité de la plateforme NVIDIA à optimiser les ressources nécessaires au traitement de ces données. Parallèlement, Salesforce et NVIDIA travaillent au développement d'avatars IA qui offrent une interaction plus humaine et immersive avec les clients et employés. Ces avatars utilisent des modèles multi-modaux pour la reconnaissance vocale et la synthèse textuelle, et ils reposent sur la puissance combinée des services NVIDIA ACE et de la plateforme Agentforce. Les clients pourront s'appuyer à la fois sur Agentforce et NVIDIA NIM Agent Blueprints pour développer et déployer des agents plus rapidement et de manière intuitive.

Licence

Elasticsearch et Kibana de retour sous licence Open Source

Après avoir tenté de contrecarrer AWS en basculant Elasticsearch et Kibana sous des licences plus restrictives et empêcher le géant du Cloud de proposer ses offres as a Service, Elastic fait machine arrière. L'éditeur a perdu son bras de fer et doit se rabibocher avec les communautés du libre pour sauver les meubles. Une affaire qui a posé clairement le problème du modèle libre face à l'essor du « as a Service ».

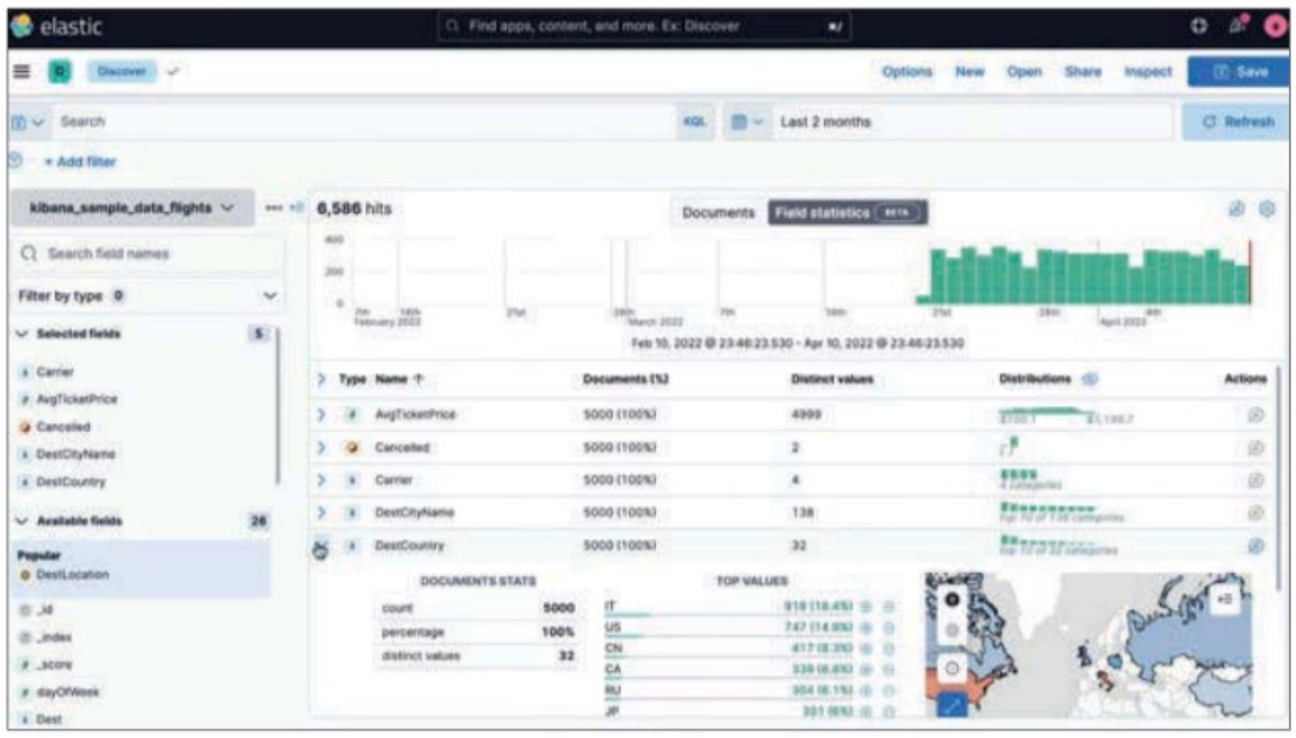
C'est clairement un sujet ultrasensible dans les communautés Open Source et notamment pour les éditeurs de logiciels. Ceux-ci se font littéralement tondre la laine sur le dos par les fournisseurs Cloud qui proposent leurs solutions en mode XaaS. Non seulement les solutions Open Source peuvent être commercialisées sans que les éditeurs ne bénéficient de royalties, licences libres oblige, mais les services des hyperscalers entrent en concurrence directe avec les offres que les éditeurs tentent de proposer sur leurs marketplaces… Les éditeurs sont pris entre le marteau et l'enclume, entre les géants du Cloud d'un côté qui profitent des droits d'usage accordés par les licences Open Source, mais aussi les chantres du logiciel libre de l'autre qui rejettent des licences qui pourraient entraver cette liberté d'une façon ou d'une autre.

Un retour vers le « vrai » Open Source

Le débat est loin de se limiter à la lutte de David contre Goliath. C'est toute la philosophie de l'Open Source qui est bousculée par le basculement du marché du logiciel vers le Cloud. L'exemple d'Elastic Search est particulièrement éloquent à cet égard. Le 29 août dernier, Shay Banon, créateur d'Elastic Search annonçait le retour d'Elasticsearch et de Kibana sous licence libre. Le ton enthousiaste du billet de blog vise à rassurer les communautés Open Source quant à l'engagement d'Elastic pour le modèle Open Source et le moins qu'on puisse dire, c'est que l'annonce n'a pas fait consensus. L'origine de l'affaire date de près de 10 ans maintenant. En 2015, AWS annonce le lancement de son service managé Amazon Elasticsearch, fruit, si l'on en croit le tweet de Werner Vogels, CTO d'AWS, d'un partenariat entre Elastic et AWS… Pour Shay Banon, il n'en est rien : il n'y a jamais eu de partenariat signé avec AWS et Amazon a lancé son service de manière unilatérale. Or, AWS n'a pas enfreint les termes de la licence Apache 2.0 et Shay Banon ne peut que s'appuyer sur la marque déposée « Elasticsearch » pour contraindre AWS à débaptiser son offre AWS Elasticsearch. Elastic décide alors de changer le licencing de ses offres ElasticSearch et Kibana pour empêcher AWS de proposer ses solutions en l'état. Elasticsearch passe d'une licence Apache 2.0 à un système de double licence maison Elastic et la SSPL (Server Side Public License), une licence initialement créée par MongoDB pour justement se protéger des fournisseurs de Cloud public qui ne contribueraient pas aux développements.

Le précédent MongoDB vs AWS

L'idée est bien que les puissants fournisseurs Cloud mettent la main à la poche afin de contribuer d'une manière ou d'une autre aux projets. On imagine les enjeux financiers derrière ces grandes manœuvres sur les licences logicielles, mais en voulant sanctionner les acteurs du Cloud, Elastic et MongoDB ont mis en place des mesures affectant toutes les communautés Open Source qui, elles-aussi, peuvent proposer des offres embarquant de nombreuses briques Open Source. En 2019, Tom Callaway, le responsable juridique de la distribution Linux Fedora, qualifiait la licence SSPL d'être « agressivement discriminatoire » envers une catégorie spécifique d'utilisateurs… Celui-ci estimait alors qu'une telle licence jetait un discrédit sur toutes les autres licences libres. La licence proposée par MongoDB suscite la polémique et, outre Fedora, c'est Red Hat qui sort MongoDB de RHEL 8 car bon nombre de fournisseurs Cloud ayant fait le choix de l'OS Red Hat se retrouvent en porte-à-faux avec la licence SSPL de MongoDB… MongoDB va d'ailleurs se casser les dents lorsqu'il cherche à faire reconnaître sa licence auprès de l'OSI (Open Source Initiative), le gardien du temple des licences Open Source.



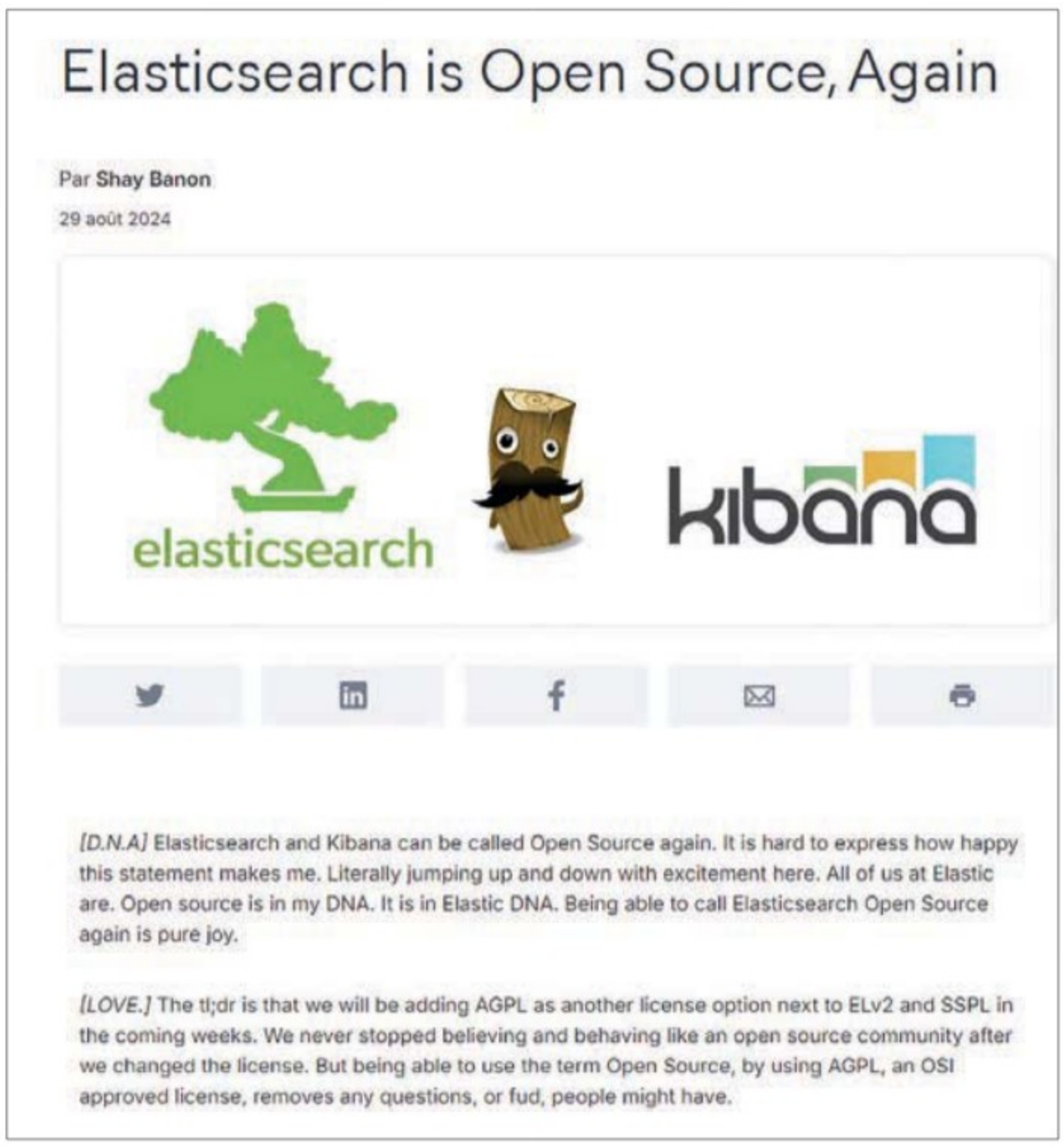
Moteur de recherche, plateforme d'observabilité, cybersécurité, les cas d'usage d'Elasticsearch sont innombrables. Depuis sa création, Elastic a levé 162 millions de dollars et valorisé à 2,5 milliards de dollars lors de son introduction en bourse en 2018.

Non seulement l'OSI ne reconnaît pas la licence, mais VM (Vicky) Brasseur, aujourd'hui directrice de la stratégie Open Source de Juniper Networks, mais surtout ancienne Vice Presidente de l'OSI se fend d'un billet assassin titré *« Elasticsearch and Kibana are now business risks »*. Son texte est un brulot contre cette licence qui n'a plus rien de libre, selon elle : *« tous les avocats spécialisés dans la propriété intellectuelle à qui j'ai montré le texte de la loi sur la protection des droits de propriété intellectuelle se sont montrés plutôt alarmés avant même d'avoir atteint la fin du texte. Fondamentalement, il s'agit d'une licence propriétaire hostile déguisée en licence open source. »* Elle pointe le côté viral de la licence : un fournisseur de service qui embarque une brique SSPL va devoir publier non seulement ce code, mais aussi le code de chaque logiciel qui l'accompagne sous licence SSPL… Bref, l'OSI ne reconnaîtra jamais une telle licence.

Un fork qui fait mal à Elastic

Face à ces changements de licences, l'attitude d'AWS a été très simple : lorsque MongoDB est passé en licence SSPL, AWS a lancé DocumentDB, une base de données NoSQL compatible MongoDB, mais totalement propriétaire. Et lorsqu'Elastic a suivi la même tactique, AWS a répliqué en forkant ElasticSearch et Kibana pour continuer à les proposer à ses clients… Contrairement à DocumentDB qui est un développement propriétaire d'Amazon et qui n'a pas été publié sous licence libre, le fork d'Elasticsearch a été rebaptisé OpenSearch en septembre 2021 et proposé sous licence Apache 2.0, *« une vraie licence Open Source »* souligne-t-on malicieusement chez AWS…

Le tweet de Werner Vogels de 2015 qui a fait sortir Shay Banon de ses gonds. AWS lance alors un service managé Elasticsearch sans signer de partenariat avec Elastic, ce qui va pousser ce dernier à basculer des licences Apache 2.0 très permissives à la très discutée licence SSPL.



Le ton enthousiaste du billet de Shay Banon « Elasticsearch is Open Source, Again » peine à masquer le camouflet infligé par AWS à Elastic…

Non seulement Elastic n'a pas fait plier AWS, mais l'éditeur se retrouve maintenant avec des distributions concurrentes à ses offres. Pour enfoncer le clou, Amazon a transféré le projet OpenSearch à la fondation Linux, ce qui pourrait bien convaincre les fournisseurs SaaS et éditeurs de logiciels rebutés à l'idée de choisir un développement piloté par AWS. Si Elastic s'est fendu de tests comparatifs démontrant que son Elasticsearch est bien plus performant que le fork réalisé à partir d'une ancienne version, avec ses licences SSPL, l'éditeur était clairement dans une impasse. Revenir à de vraies licences Open Source est un moyen d'éteindre le feu pour Shay Banon, mais cette volte-face est loin d'avoir fait l'unanimité dans les communautés. Face aux avis négatifs, Shay Banon les qualifie de trolls et préfère noter un *« énorme flux de sentiments positifs et de réactions de notre incroyable communauté »*… □

AC

SecNumCloud

Whaller DONJON reconnue par l'ANSSI

Reconnue pour sa sécurité de haut niveau, la plateforme collaborative française Whaller DONJON s'impose comme une référence pour les organisations traitant des données sensibles. Après plus de deux ans de développement, elle a récemment décroché la prestigieuse qualification SecNumCloud délivrée par l'ANSSI pour sa solution SaaS.

Fondée en 2013 par Thomas Fauré, Whaller s'est distinguée dès ses débuts par son approche axée sur la protection des informations. Hébergée en France chez OVHcloud, sa plateforme collaborative Whaller DONJON a été conçue pour répondre aux plus hautes exigences en matière de confidentialité et de souveraineté des données. Avec l'obtention du Visa de sécurité de l'ANSSI, Whaller franchit une étape stratégique. La qualification SecNumCloud garantit que le service cloud est conforme aux exigences de l'État français en matière de confidentialité, d'intégrité et de disponibilité des données sensibles, couvrant aussi bien les services IaaS, PaaS que SaaS. Les fournisseurs certifiés SecNumCloud doivent répondre à des critères stricts en matière de gestion des risques, de sécurité des infrastructures, de contrôle des accès et de résilience. *« L'obtention du Visa de sécurité de l'ANSSI pour la qualification SecNumCloud de notre solution est une étape clé dans notre mission de fournir des services collaboratifs sécurisés et fiables. Elle renforce notre position en tant que leader de confiance dans le secteur des plateformes collaboratives et souligne notre engagement continu envers la sécurité des données et la souveraineté technologique française. »* déclare Thomas Fauré, fondateur et président de Whaller. Forte de plus d'un million d'utilisateurs, Whaller DONJON est aujourd'hui la première et la seule plateforme collaborative française à avoir obtenu cette certification.

La plateforme de communication et de collaboration Whaller DONJON est la première et unique solution collaborative qualifiée SecNumCloud.

« LA QUALIFICATION SECNUMCLOUD ATTESTE QUE LES QUATRE-VINGT MESURES DU RÉFÉRENTIEL SONT MISES EN ŒUVRE DANS NOTRE SOLUTION. CE QUI EST VRAIMENT ESSENTIEL, C'EST D'UNE PART L'IMMUNITÉ AUX LOIS EXTRATERRITORIALES QUI PROTÈGE LES DONNÉES CONTRE L'ACCÈS DES AUTORITÉS ÉTRANGÈRES, ET D'AUTRE PART NOTRE DISPOSITIF QUI EST À L'ÉTAT DE L'ART D'UN POINT DE VUE CYBERSÉCURITÉ. »

Cyril Bras, Directeur Cybersécurité chez Whaller.

Pouvez-vous nous présenter Whaller Donjon ?
Whaller DONJON est une déclinaison de notre solution standard, conçue pour un environnement sécurisé avec des serveurs réservés pour chaque client. Contrairement à notre offre classique, cette version restreint les fonctionnalités nécessitant des connexions externes afin de garantir une isolation totale. L'outil propose un large éventail de fonctionnalités : visioconférence, édition collaborative de documents, messagerie instantanée, gestion de projets, etc. Il peut également être utilisé comme réseau social d'entreprise, plateforme de travail collaboratif ou intranet. Très modulable, il s'adapte aux besoins de chaque organisation. Certaines fonctionnalités, comme la traduction, ont été retirées, car elles nécessitent un accès à des ressources extérieures. Selon leurs besoins, les entreprises peuvent intégrer la plateforme dans leur système interne, tout en ayant la possibilité de la connecter à Internet via un VPN ou de la garder totalement isolée. Toutes ces fonctionnalités sont disponibles dans notre offre certifiée « SecNumCloud ».

Quelles sont les principales raisons qui ont motivé Whaller à obtenir la qualification SecNumCloud ?
L'objectif était de faire de la cybersécurité un différenciateur commercial par rapport à nos concurrents. Lorsque j'ai rejoint l'entreprise il y a deux ans et demi, ma feuille de route était d'obtenir la qualification SecNumCloud et d'intégrer la cybersécurité dans tous les aspects de Whaller. Nous avons même été au-delà des exigences de SecNumCloud en lançant par exemple notre propre centre de réponse à incidents pour nos clients et partenaires. L'idée étant de démontrer notre capacité à répondre en cas d'attaque, et surtout de partager notre expertise de manière pédagogique. Nous fournissons à nos utilisateurs des indicateurs techniques liés à des menaces potentielles et des modes opératoires d'attaquants. Concrètement, nous observons les événements sur notre plateforme, les documentons et les communiquons ensuite gratuitement à nos clients.

Votre collaboration avec OVHcloud a-t-elle facilité l'obtention de la qualification de l'ANSSI ?
Ce n'est pas seulement une question de facilité. Étant donné que nous ne sommes pas notre propre hébergeur, il est essentiel de s'appuyer sur un hébergeur déjà qualifié pour que notre solution le soit également. Nous avons choisi OVHcloud, mais l'avantage de notre qualification actuelle est qu'elle repose sur un modèle de qualification par composition. Cela signifie que nous pourrions également nous tourner vers un autre fournisseur de cloud qualifié. Ce qui nous caractérise, c'est notre code applicatif et l'architecture informatique que nous avons développés et qui ont été validés par l'ANSSI. Si nous reproduisons ce schéma sur des serveurs SecNumCloud, nous avons la possibilité de changer d'hébergeur, à condition toutefois d'obtenir une validation de l'ANSSI. Ce qui nous différencie de nos concurrents, c'est que nous sommes les seuls à avoir fait qualifier une solution *« Software as a Service »* (SaaS). D'autres plateformes peuvent être hébergées sur des infrastructures qualifiées SecNumCloud, mais elles ne seront pas elles-mêmes qualifiées. L'ANSSI a clairement indiqué qu'il n'y a pas de phénomène de ruissellement : utiliser un hébergement qualifié ne rend pas automatiquement votre produit SecNumCloud.

Quels sont les avantages concrets de la qualification SecNumCloud pour vos clients ?
Tout d'abord, nos clients bénéficient d'une immunité totale aux lois extraterritoriales. Contrairement aux entreprises sous pavillon américain, nous ne sommes pas tenus de répondre à des demandes d'informations de la part des autorités américaines, par exemple. Ensuite, nous garantissons un niveau de sécurité contractuel. La qualification SecNumCloud atteste que les quatre-vingts mesures du référentiel sont mises en œuvre dans notre solution. Ce qui est vraiment essentiel, c'est d'une part l'immunité aux lois extraterritoriales qui protège les données contre l'accès des autorités étrangères, et d'autre part notre dispositif qui est à l'état de l'art d'un point de vue cybersécurité. Il ne s'agit pas juste d'une déclaration, nous sommes soumis à des audits tout au long de notre qualification. Celle-ci est accordée pour une certaine durée et dans ce delta, nous devons faire faire régulièrement des audits de configuration, des tests d'intrusion, et des évaluations d'architecture. C'est très strict, très carré, et c'est cette garantie-là que nous offrons à nos clients.

Comment Whaller DONJON contribue-t-elle à la souveraineté numérique française ?
Nous offrons un outil qui est développé et opéré en France sur des serveurs français garantis contre les lois extraterritoriales. De plus, nous proposons une solution qui assure aux utilisateurs et aux clients que leurs données ne peuvent pas être soumises à des lois extra-européennes. Si un organisme comme la NSA venait frapper à notre porte pour une demande de données, nous ne serions pas contraints d'y répondre. ☐

J.C

Gestion de l'obsolescence

PRO BTP veut maîtriser l'obsolescence technique pour se moderniser

Sabrina Drouet, Responsable Stratégie IT de PRO BTP précise à *L'Informaticien* que la gouvernance et les astuces retenues pour réduire la dette technique du S.I. sont un prélude indispensable à sa rénovation autour d'un nouvel outillage méthodologique et technique.

Quels sont les grands métiers de PRO BTP ?
Sabrina Drouet : PRO BTP est un groupe de protection sociale dont les métiers sont en lien avec le commercial, la gestion de dossiers, la comptabilité, la finance, et les RH. À la DSI, nos trois métiers principaux sont le développement, la gestion de projets et l'administration des infrastructures.

Votre groupe couvre plusieurs millions de professionnels du BTP. Quel est l'effectif total ?
Nous comptons 3 millions de particuliers couverts et 300 000 entreprises clientes. PRO BTP est le huitième assureur français avec une particularité importante : il s'agit d'une association loi 1901 dont l'ensemble des bénéfices est reversé dans des actions sociales pour les adhérents. Nous sommes 5 500 collaborateurs répartis sur l'ensemble de la France, dont 10 % environ, entre 500 et 550 personnes travaillent à la DSI. Comme toutes les grosses structures, nous disposons d'un parc informatique conséquent, composé de 112 applications maison, 319 progiciels et 208 composants techniques identifiés. L'expansion des matériels et logiciels se poursuit. C'est ce qui nous a amenés à chercher à limiter la dette technique.

Quand et comment ce projet a-t-il démarré ?
L'analyse de l'obsolescence a commencé en 2022. C'est un défi majeur pour l'informatique nécessitant de mener des actions concrètes et surtout de faire comprendre l'importance du projet à la Direction Générale. Cela reste un défi en 2024, mais moins. Grâce au programme de gestion de l'obsolescence que nous avons mis en place, nous avons résorbé, dès 2023, 50 % de la dette technique. Nous avons pour objectif de terminer ce programme en 2025 pour ensuite passer sur une gestion courante du run.

Justement, quels sont les objectifs d'un tel programme ?
Il s'agit de gagner en performances, en sécurité et en qualité de services. L'enjeu qui a scellé le démarrage de ce programme était de mesurer l'obsolescence. Nous avons appliqué une méthode dans ce but. Le second point était d'obtenir le budget nécessaire pour réduire cette dette technique. Un premier volet s'est concrétisé par des montées de versions, un autre par des remplacements de logiciels et des rénovations de briques applicatives.

La première étape a-t-elle consisté à faire un inventaire du parc ?
Recenser ses actifs numériques forme effectivement la première étape. Pour y parvenir, nous nous sommes appuyés sur deux outils : un outil de cartographie maison et la CMDB qui a permis de recenser l'ensemble des matériels et logiciels.

Comment avez-vous évalué précisément la dette technique ?
Nous sommes allés voir chaque responsable applicatif pour récupérer un ensemble d'indicateurs afin de définir chaque application et l'ensemble des critères en lien avec cette application.

Cela exige un fort soutien financier et des moyens humains ?
Ces indicateurs nous ont permis d'évaluer la dette technique. Nous avons pu définir les actifs obsolescents. Puis, nous avons présenté un rapport à notre direction générale et au comité des fonctions clés qui ont budgétisé un programme pour piloter la réduction de la dette, programme qui a vu le jour en 2023.

Quel aspect du programme vous a semblé essentiel ?
De manière surprenante, dans une analyse d'actifs, la partie la plus importante est la partie humaine. Il a fallu mobiliser l'ensemble des intervenants de la DSI, l'ensemble des responsables applicatifs pour adhérer à cette démarche. Il a fallu également sensibiliser et mobiliser notre direction générale pour bien comprendre l'importance du sujet pour la DSI. Une métaphore a permis de souligner la nécessité de rénover notre patrimoine informatique.

Quelle a été la principale difficulté rencontrée ?
Elle a été de convaincre et d'embarquer les équipes, de les mobiliser pour passer du temps à nous aider à cerner, puis à réduire l'indice d'obsolescence. On cherche à améliorer l'engagement de service en termes de disponibilité, de maintenance, et d'évolutivité de nos solutions. Ce programme nous a permis d'atteindre cet objectif.

Les aspects économiques du programme sont-ils conséquents ?
C'est un programme important de plusieurs milliers de jours-hommes et de plusieurs millions d'euros dans le cas de nos renouvellements. Le programme de traitement de l'obsolescence engagé en 2023 représente 5 000 jours/hommes ; il a mobilisé 8 % du budget d'évolution de la DSI. Un effort considérable a été fait par les équipes, avec la nomination d'un directeur de programme de gestion de l'obsolescence, épaulé par des chefs de projets.

Qu'a permis cette gouvernance spécifique ?
Elle a permis de mener une trentaine de projets de front. Grâce à un reporting de suivi des actions, nous avons pu rendre compte des avancées du programme à la direction générale. A la fin de l'année 2023, 17 projets de modernisation ont été terminés, 11 autres ont été suspendus faute de budget, avec l'acceptation par les métiers du risque associé, et 17 projets sont poursuivis en 2024. Lorsqu'on demande du budget, il importe de répondre précisément à nos engagements en tant que DSI.

Après près de deux ans de conduite de projets, quel est votre premier bilan ?
Le premier bilan est que nous avons réussi à réduire et à maîtriser, de manière conséquente, la dette technique, malgré l'augmentation du nombre d'actifs numériques. Nous parvenons à une maîtrise du taux d'obsolescence, dont la moyenne du marché oscille entre 10 % et 20 %. La principale satisfaction consiste à pouvoir arrêter ce programme pour partir sur l'exploitation courante.

Quels sont les principaux enseignements que vous retenez ?
L'humain, l'humain, l'humain. Une DSI ne peut pas fonctionner sans les personnes qui la constituent. Une intelligence collective doit naître, où chaque sachant apporte et partage la connaissance de sa brique applicative, ou de son logiciel.

Comment avez-vous géré le changement du côté des utilisateurs métiers ?
Nous avons sensibilisé aussi les métiers face à l'importance de cette dette. Il a fallu aussi les responsabiliser, l'idée étant d'avoir une responsabilité partagée entre les métiers et l'IT pour la gestion de l'obsolescence. Dans le cadre des grands programmes de rénovation en cours chez PRO BTP, il nous fallait décommissionner toutes les briques devenues obsolètes.

D'autres bénéfices constatés, au niveau de l'interopérabilité des applications ?
Très nettement, notre programme a permis d'offrir un support plus simple et de simplifier toute l'équation IT. Nous pouvons mieux communiquer entre nous, à la DSI, ainsi qu'avec les clients et également avec la direction générale.

Des avancées technologiques, comme l'IA ou l'automatisation, deviennent-elles possibles ?
Notre programme majeur Stella prévoit la rénovation des briques métiers. Il vise à moderniser la gestion de la santé et de la prévoyance ainsi que la gestion de la relation client autour d'outillages techniques et méthodologiques adaptés : approche agile, chaîne de déploiement et d'intégration continue notamment. À horizon 5 à 10 ans, nous aurons renouvelé 70 % de notre système d'information.

L'approche cloud hybride est-elle envisagée ?
Aujourd'hui, nous travaillons sur certains clouds publics. Mais, en tant qu'assureur et au vu des exigences de sécurité et de conformité fixées, nous privilégions notre cloud privé et ses ressources sur site. Suite à la mise en place du RGPD, nous portons une attention particulière aux données privées de nos adhérents.

Qu'en est-il des objets connectés ?
Dans le monde de l'assurance, ils sont encore peu présents. Nous couvrons la santé et la prévoyance, donc la maladie, l'invalidité et le décès, des sujets auxquels l'IoT ne se prête pas forcément bien. En revanche, notre branche médico-sociale gère des établissements de santé et des EHPAD où l'IoT a toute sa place.

Propos recueillis par Olivier Bouzereau

Transformation

Orange International Networks Infrastructures & Services passe au Cloud natif

Le fournisseur de solutions réseau pour Orange Business Service et Orange Wholesale passe sur Openshift pour obtenir plus de flexibilité et de modularité sur son infrastructure.

Pionnier dans l'infrastructure Cloud pour les télécommunications, Orange International Networks Infrastructures & Services disposait d'une infrastructure sur des machines virtuelles disséminées dans ses centres de données de périphérie qui s'appuyait sur l'Open Stack de Red Hat et Contrail de Juniper. Mame Bess Diop, Telco Cloud Design Director chez OINIS, explique : *« nos clients nous demandent des services à la carte, donc un peu plus flexibles, plus modulables, et puis sur un modèle Pay As You Grow, qu'ils peuvent arrêter quand ils veulent. Pour ce faire, côté Orange Business, il y a la plateforme Evolve, qui est un marketplace qui permet aux clients d'avoir accès à leurs infrastructures, de pouvoir les manager comme ils veulent. Mais pour que ce soit vraiment efficace, il fallait aussi, côté infrastructure, que ce soit modulable »*. Elle ajoute : *« il était important d'être en mode cloud natif, justement, pour profiter de cette flexibilité. Et faciliter aussi les opérations pour nos équipes, et innover »*.

Mame Bess Diop, Telco Cloud Design Director chez OINIS.

Un projet d'ampleur

L'opération n'était pas anodine alors que l'entité d'origine possède actuellement près de 50 minis centres de données et en détiendra 75 à la fin de 2026. Dans ces centres, de très nombreuses machines virtuelles sont présentes, et il fallait donc répondre à la problématique de les migrer vers le nouvel environnement choisi. Il existe évidemment des outils logiciels pour le faire, mais Mame Bess Diop indique qu'ils n'ont pas forcément la maturité suffisante pour un projet de cette ampleur. Un appel d'offres a été lancé et 10 entreprises ont été interrogées. Quatre ont été en « short list ». La directrice précise : *« il était important pour nous, non seulement de faire l'étude papier, mais aussi de faire des tests dans nos propres labs, c'est-à-dire pas dans le laboratoire des fournisseurs, mais dans notre propre laboratoire, pour mesurer nous-mêmes et nous faire une idée de la maturité de la solution, est-ce que ça répond à nos contraintes, etc. Et donc, du coup, c'était deux semaines par fournisseur. »* Elle continue : *« l'aspect performance, latence est assez important, en effet. »*

Le contrôle, un critère de décision

Sur l'ancienne infrastructure, OINIS avait un très bon contrôle, car elle réalise aussi de l'attachement réseau. *« On relie, en fait, nos VM, par exemple, à nos VPN, nos MPLS, pour assurer une sécurité, renforcer la sécurité pour nos clients »* précise Mame Diop. Sans compter que pendant la compétition, Juniper annonçait la fin de son support de Contrail. Pour finir, ce point a été développé en interne chez OINIS à partir d'une solution en cours de développement de l'équipe innovation de l'entité d'Orange.

Un déploiement en plusieurs étapes

Au final, OINIS a choisi OpenShift de Red Hat. Les premiers points de présence vont être mis en œuvre dès le début de novembre prochain en France et en Europe. Celui-ci va continuer tout au long de 2025 sur les centres de données existants et les nouveaux à venir. De plus, la nouvelle infrastructure va soutenir la mise en place de nouveaux services comme le cœur 5G ou l'Internet des Objets. L'année prochaine se réalisera la migration de la version 1, la version OpenStack Control, vers la nouvelle version qu'on vient de mettre en place. La mise en œuvre sera faite par les équipes internes avec l'appui de Red Hat et de l'équipe innovation qui comprend de nombreux experts sur la question.

20 % sur le Green

Le côté Green est important pour le groupe Orange et la question était clé sur ce projet stratégique. Mame Diop précise : *« On avait mis une note de 20 % sur cette partie-là, pour montrer que c'était super important pour nous »*. Il était ainsi demandé aux fournisseurs de s'assurer que toutes les versions hardware qu'on avait sur les anciens POP étaient supportées. Dès cette année, au niveau de la direction ingénierie, va commencer la validation sur des serveurs de seconde main et commencer à les déployer dans une proportion de 25% des serveurs pour commencer. Il a été de plus négocié un support étendu pour les matériels. Mame Diop précise : *« on aura le même niveau de support que sur les équipements neufs »*. □

B.G

Stockage

SEW USOCOME fait le choix de DataCore

L'entreprise allemande avec une forte implantation en France évolue vers une solution de stockage définie par logiciel pour s'autoriser des changements guidés par le pragmatisme.

L'entreprise, qui a fait le choix des matériels HPE, renouvelle régulièrement ses équipements dès qu'ils deviennent obsolètes, soit tous les 4 à 5 ans. Lors de chaque renouvellement, SEW USOCOME effectue une analyse du marché pour sélectionner la solution la mieux adaptée aux besoins sur l'un de ses sites. Si la société reste fidèle à HPE pour le matériel, elle a souhaité explorer d'autres options pour la partie logicielle, en réexaminant notamment DataCore, consulté quelques années auparavant.

Le sur site privilégié

Manuel Bohr précise : *« nous avons une politique très anti-cloud. C'est-à-dire que les salles où nos données sont stockées dans nos data centers sont sur site, et l'ensemble du matériel est stocké dans nos différentes salles pour des besoins de haute disponibilité. Autrement dit, on a une salle A, une salle B sur chacun de nos sites. Avec 3 sites, nous avons donc 6 centres de données »*. Les salles sont en actif-passif pour qu'il y ait toujours, en cas de perte d'une salle, la capacité nécessaire à la salle de secours de reprendre le relais naturellement. Chaque site est plus ou moins indépendant et fonctionne de son côté à part pour quelques applications qui sont gérées centralement. Dernièrement, les solutions de stockage étaient installées sur des solutions de Simplivity et de Nimble Storage. Or sur Simplivity l'extension parfois nécessaire du système était compliquée. Les remplacements se font par étapes et site par site. L'infrastructure combine donc encore des baies Nimble Storage et des solutions de DataCore. *« Chaque année, on va remplacer certains serveurs, certaines solutions. Et donc, l'année dernière, effectivement, c'était un peu exceptionnel, il y en a eu deux d'un coup parce qu'aussi la solution DataCore était moins chère qu'une solution HPE, ce qui nous a permis d'en faire deux »* précise Manuel Bohr. *« Cette année, il n'y a rien, on est orienté sur plutôt un projet de sauvegarde qui est quelque chose à part. Et l'année prochaine, effectivement, reviendra sur la table le choix de la solution HPE si on change vers du DataCore ou vers autre chose. C'est une vision pragmatique du projet à chaque fois »* indique le salarié de SEW USOCOME.

Une rupture avec DataCore

Manuel Bohr met en avant la flexibilité obtenue avec le choix de DataCore ainsi que la baisse des coûts par rapport aux solutions précédentes. Définie logiciellement,

Manuel Bohr du FabLab de SEW USOCOME.

DataCore lui permet aussi de choisir le matériel le plus adapté, voire d'autres constructeurs que HPE, si besoin. Notre interlocuteur met de plus en avant la granularité de la fonction de protection continue des données présente sur la plate-forme qui est utilisée en stockage primaire. Même s'il n'a pas eu à l'utiliser beaucoup, Manuel Bohr a aussi apprécié le support de l'éditeur. L'approche de l'éditeur est un peu différente de celle habituellement utilisée par les autres fournisseurs. Au début de la relation, DataCore fait remplir un dossier technique très complet et une validation du processus d'installation. Manuel Bohr ajoute : *« Mais je pense que le jour, en cas de problème, le fait que DataCore connaisse entièrement notre infrastructure pour nous aider dans le dépannage sera vraiment un plus »*. □

B.G

Anti-manuel d'intelligence artificielle

ou répondre aux questions que vous n'osez pas demander !

L'intelligence artificielle, et pas seulement la générative, a envahi nos vies, nos réseaux sociaux, nos logiciels et nos façons de travailler. Elle pose cependant autant de questions qu'elle n'apporte de réponses. Dans leur ouvrage, Victor Storchan et Vladimir Atlani ne veulent pas seulement faire un point précis sur les problèmes qu'elle pourrait engendrer, mais aussi donner les éléments factuels qui la régissent. Comment fonctionnent les algorithmes ? Comment sont-ils conçus ? Avec quelles données ? Pourquoi faire ? Se qualifiant d'anti-manuel, le livre est surtout un moyen de répondre à la plupart des questions que l'on peut se poser sur cette technologie et sur les conséquences de son utilisation. Grâce à de nombreux exemples concrets et des anecdotes passionnantes, les auteurs s'adressent avec pédagogie aux étudiants, aux professionnels, aux décideurs politiques, mais aussi à tous les citoyens désireux de participer activement à ce grand débat de société. Car oui, le sujet de l'IA est à la portée de tous, à condition d'être bien guidés !
Permettre à chacun de comprendre en profondeur les enjeux économiques, éthiques et géopolitiques de l'IA, son impact sur des sujets comme la santé, le sport, l'environnement… et, ainsi, encourager une réflexion collective sur les futurs possibles pour notre humanité, tels sont les objectifs de cet anti-manuel de l'IA.

Contexte historique

Comprendre la révolution technologique qui sous-tend l'émergence de ChatGPT ou Midjourney nécessite un bref retour dans l'histoire récente de l'IA. Au début des années 2010, l'IA connaît un tournant décisif. Les chercheurs mettent en évidence la supériorité du deep learning pour un grand nombre de tâches traitant du texte ou des images en s'appuyant sur l'arrivée des données massives et des nouveaux moyens de calculs. On entraînait alors un modèle pour une tâche donnée : classer des images, trier du texte, traduire du texte, jouer au go, etc. Au-delà de la tâche pour laquelle ces modèles étaient entraînés, ils s'avéraient totalement inutiles. ChatGPT est l'incarnation d'un changement de paradigme qui s'opère avec l'IA générative : un modèle est désormais performant sur un grand nombre de tâches[1]. Il devient performant pour effectuer des tâches sur lesquelles il n'a pas été entraîné a priori et est ainsi capable de généraliser à partir de simplement quelques exemples écrits en langage naturel (en français ou anglais par exemple). En outre, ses performances s'améliorent lorsqu'on lui donne un exemple de raisonnement logique aidant à la résolution. À cet égard, il joue un rôle semblable à celui d'un compilateur informatique : traduire un programme haut niveau en instruction exécutable par une machine. En clair, on peut dialoguer avec la machine en langage naturel pour lui faire exécuter des tâches très diverses. Ce nouveau paradigme devient possible grâce à l'émergence concomitante de plusieurs facteurs. D'abord, le passage à l'échelle de la taille des modèles (qui a augmenté de près de 15 000 fois entre 2017 et 2022[2]) a permis des gains de performances inégalés. La puissance de calcul des cartes graphiques utilisée pour l'entraînement a été multipliée par 7 000 depuis 2003 et le prix d'une unité de calcul a drastiquement baissé[3] dans le même temps. Avant l'ère de l'apprentissage profond, la quantité de calcul utilisée par les modèles d'IA doublait tous les 21 mois environ ; depuis que l'apprentissage profond s'est imposé vers 2010, la quantité de calcul utilisée par les modèles a commencé à doubler tous les 4-6 mois[4]. Enfin, la donnée sur laquelle ces

1 : https://aclanthology.org/2022.emnlp-main.340.pdf

2 : https://d1.awsstatic.com/events/Summits/reinvent2022/AIM405_Train-and-deploy-large-language-models-on-Amazon-SageMaker.pdf

3 : https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

4 : https://arxiv.org/abs/2202.05924

modèles sont entraînés est accessible aux développeurs en quantité toujours plus importante. Ces trois facteurs de base sont centraux dans la fabrication des modèles de fondation.

La fabrication des modèles de fondation, au cœur de l'IA générative

Les foundation models ou « modèles de fondation » en français forment une classe émergente de modèles à la base de la révolution de l'IA générative. Avec l'avènement des foundation models, les chercheurs ont conclu que le passage à l'échelle de ces systèmes (en termes de taille des modèles et des jeux de données et donc des capacités de calculs) était la clé pour repousser la frontière technologique. Ces modèles sont fabriqués en plusieurs étapes. Ces modèles peuvent traiter un type de données en particulier (texte, image, audio, séquences de protéines, etc.) ou être aussi multimodaux, c'est-à-dire combiner plusieurs types de données pour produire leur analyse. Notons que les large language models (LLMs) ou grands modèles de langage sont une sous-classe des modèles de fondation. Pour illustrer leur fabrication, nous prendrons l'exemple du texte.

Le modèle de base est souvent formé par apprentissage autosupervisé sur des données à grande échelle (« web scale data »). On prédit par exemple un mot manquant dans une phrase compte tenu de son contexte environnant (c'est-à-dire les mots précédant et/ou suivant le mot masqué). Ainsi, l'apprentissage autosupervisé ne nécessite pas d'exemples annotés, mais construit automatiquement ses propres tâches de prédiction. Cette phase est appelée la phase de pré-entraînement et le modèle qui en résulte est essentiellement une machine à prédire le prochain mot (par exemple, proposer une suite à un prompt d'un utilisateur de chatbot). À cette étape, les LLM peuvent être vus comme une « infrastructure », une brique de base pour traiter de l'information. Ce ne sont pas encore des produits capables de respecter une politique de modération de contenu du type de ChatGPT. Ces modèles sont en général modifiés et affinés après cette phase de pré-entraînement initiale.

La phase de fine-tuning ou d'adaptation consiste à modifier les LLM pré-entraînés pour qu'ils produisent de manière plus fiable les résultats souhaités. Cela peut consister à améliorer l'alignement d'un modèle sur des préférences utilisateurs ou à le spécialiser pour un certain domaine ou un certain type de tâches. Parmi les approches principales, on compte :

- L'apprentissage par instructions, qui consiste à montrer des exemples de réponses possibles pour diverses tâches au modèle (cet entraînement est donc supervisé).

- L'apprentissage par renforcement à partir de feedback humain. La plupart des nouveaux modèles ont été affinés/spécialisés par leurs concepteurs pour être plus performants sur certaines tâches (et domaines) ou agir comme des filtres de sécurité afin de contraindre certains types de réponses à *« rester dans les rails »*, c'est-à-dire à se conformer à certaines politiques de modération de contenu. On utilise des données humaines collectées par les grands laboratoires d'IA qui externalisent l'annotation à des plateformes tierces[5] (les annotateurs peuvent être localisés par exemple au Kenya et payés 2 dollars de l'heure dans le cas de OpenAI[6]). Ces annotateurs vont noter les « bonnes » et les « mauvaises » réponses du modèle en fonction de règles qui leur sont données par les concepteurs des modèles. Cette phase a pour but d'aligner le modèle sur des préférences utilisateurs (avoir une bonne expérience de dialogue fluide) ou de contraindre le modèle à ne pas répondre à certaines questions ou répondre d'une certaine manière. Cette phase est facilement cassable (pour un développeur malveillant qui aurait accès au modèle) et peu robuste.

- Il existe d'autres techniques moins coûteuses que l'adaptation des LLM, ne nécessitant pas l'emploi de ces plateformes d'annotation. C'est le cas de l'apprentissage en contexte (incontext learning) qui ne requiert de la part de l'utilisateur que de donner l'information nécessaire au modèle via un prompt contenant lui-même des exemples ou un contexte important pour traiter sa requête. L'architecture RAG (Retrieval Augmented Generation) permet aussi au modèle d'aller trouver de l'information dans des documents stockés dans une base de données qu'on lui met à disposition en plus de l'information déjà encodée dans ses poids[7].

Les jeux de données de l'IA générative

L'entraînement des foundation models nécessite d'énormes quantités de données (notamment dans la phase de pré-entraînement). Un humain aurait mis 20 000 années à lire les jeux de données utilisés pour l'entraînement des LLMs publiés fin 2023. Bien que les développeurs cherchent généralement à maximiser la qualité, la taille et la diversité des données d'apprentissage, de facto, internet est devenu la source privilégiée pour compiler ces énormes jeux de données. Cette situation pose un problème de représentativité et de biais pour des modèles censés pouvoir constituer demain une nouvelle interface dans notre accès à l'information. En effet, 63,7 % des communications sur internet se font en anglais mais seulement 5 % de la population mondiale parle l'anglais comme langue maternelle à la maison. Près de 3 milliards de personnes — soit 37 % de la population mondiale — n'ont pas accès à internet, selon les Nations unies. Enfin, l'accès à l'internet est inégalement réparti en fonction de l'âge, du sexe, du revenu, du niveau d'éducation, etc. La taille si volumineuse des données fait qu'il n'est pas

5 : Exemple de telles plateformes : https://www.mturk.com/

6 : https://time.com/6247678/openai-chatgpt-kenya-workers/

7 : Les poids (weights) sont les coefficients ajustés post-entraînement d'un modèle d'apprentissage automatique, notamment dans les réseaux de neurones.

possible de pouvoir, un à un, examiner la qualité des exemples que l'on montre au modèle. Les jeux de données utilisés sont et resteront biaisés et la qualité du texte varie de façon importante. Ainsi, des chercheurs ont trouvé qu'un des jeux de données d'images devenu presque incontournable pour l'IA contenait des images d'abus sexuels d'enfants[8].

L'évaluation des foundation models

L'évaluation est le processus qui consiste à mesurer les performances d'un modèle selon diverses métriques, c'est-à-dire différents critères. Les modèles de fondation et leurs fonctionnalités (raisonnement, planification, suivi d'une instruction, etc.) sont difficiles à évaluer pour de nombreuses raisons. Ils peuvent accomplir un grand nombre de tâches différentes : l'évaluation (comprendre la performance des modèles) devient de plus en plus complexe à mesure que les systèmes d'IA sont à usage général, c'est-à-dire utilisables pour des tâches de plus en plus variées. Une autre couche de complexité de l'évaluation de la performance est qu'il n'y a pas toujours de bonne réponse à une question ou une instruction donnée. Enfin, les évaluations peuvent demander beaucoup de ressources (par exemple l'intervention de médecins pour superviser les évaluations pour des modèles d'IA appliqués à la biologie, etc.).

Dans la pratique, les types d'évaluation les plus courants sont les suivants :

▪ Benchmarks académiques : un benchmark académique est constitué d'une ou plusieurs tâches à évaluer, d'un ou plusieurs jeux de données servant d'exemples pour l'évaluation et d'une ou plusieurs métriques servant à mesurer la performance du modèle. Les benchmarks académiques sont développés par des chercheurs dans le but de fournir des mesures rigoureuses et reproductibles de la capacité des modèles. Ils sont souvent publiés dans des documents tels que les documents techniques ou les papiers de recherche sur les modèles.

▪ Préférences humaines : pour certaines tâches, il n'y a pas de bonne réponse, par exemple pour savoir quel modèle est le meilleur pour écrire des poèmes. Pour cela, les développeurs évaluent les préférences humaines, par exemple en fonction de multiples critères (ton de la réponse, longueur de la réponse, formatage des réponses ou simplement de l'utilité de la réponse).

▪ Red teaming : le « red team » est similaire aux crash-tests des voitures. Bien que personne n'espère jamais un accident, on veut tester la résilience de la voiture dans les pires scénarios imaginables. Les modèles sont souvent testés avec une présomption de bonne intention, c'est-à-dire que les tests ne sont pas de nature adverse. Le red teaming est une forme de test où le modèle est spécifiquement incité à dérailler, avec des prompts destinés à susciter des réponses non désirées.

Les gouvernements ont commencé à encourager la pratique du « red teaming » pour l'IA[9].

▪ Le LLM juge : cette méthode d'évaluation est relativement récente. L'idée est d'entraîner, puis d'utiliser un modèle spécifiquement pour évaluer les résultats d'un autre modèle. Cela a pour avantage d'automatiser l'évaluation sans avoir à recourir à des annotations humaines. En revanche, cette méthode reste pour l'instant moins précise que le recours à un humain dans de nombreuses situations.

Aucune de ces méthodes n'est parfaite. L'évaluation est aujourd'hui à la fois un problème pratique et un problème de recherche car on observe des écarts importants entre ce que mesurent les benchmarks académiques et les performances réelles des modèles déployés. □

8 : https://www.theverge.com/2023/12/20/24009418/generative-ai-image-laion- csam-google-stability-stanford

9 : https://www.whitehouse.gov/ostp/news-updates/2023/08/29/red-teaming- large-language-models-to-identify-novel-ai-risks/

▪ Auteur(s) : Victor Storchan, Vladimir Atlani
▪ Editeur(s) : Eyrolles
▪ EAN13 : 9782416017230
▪ ISBN10 : 2416017233 ▪ Prix : 22 €

Centre de données

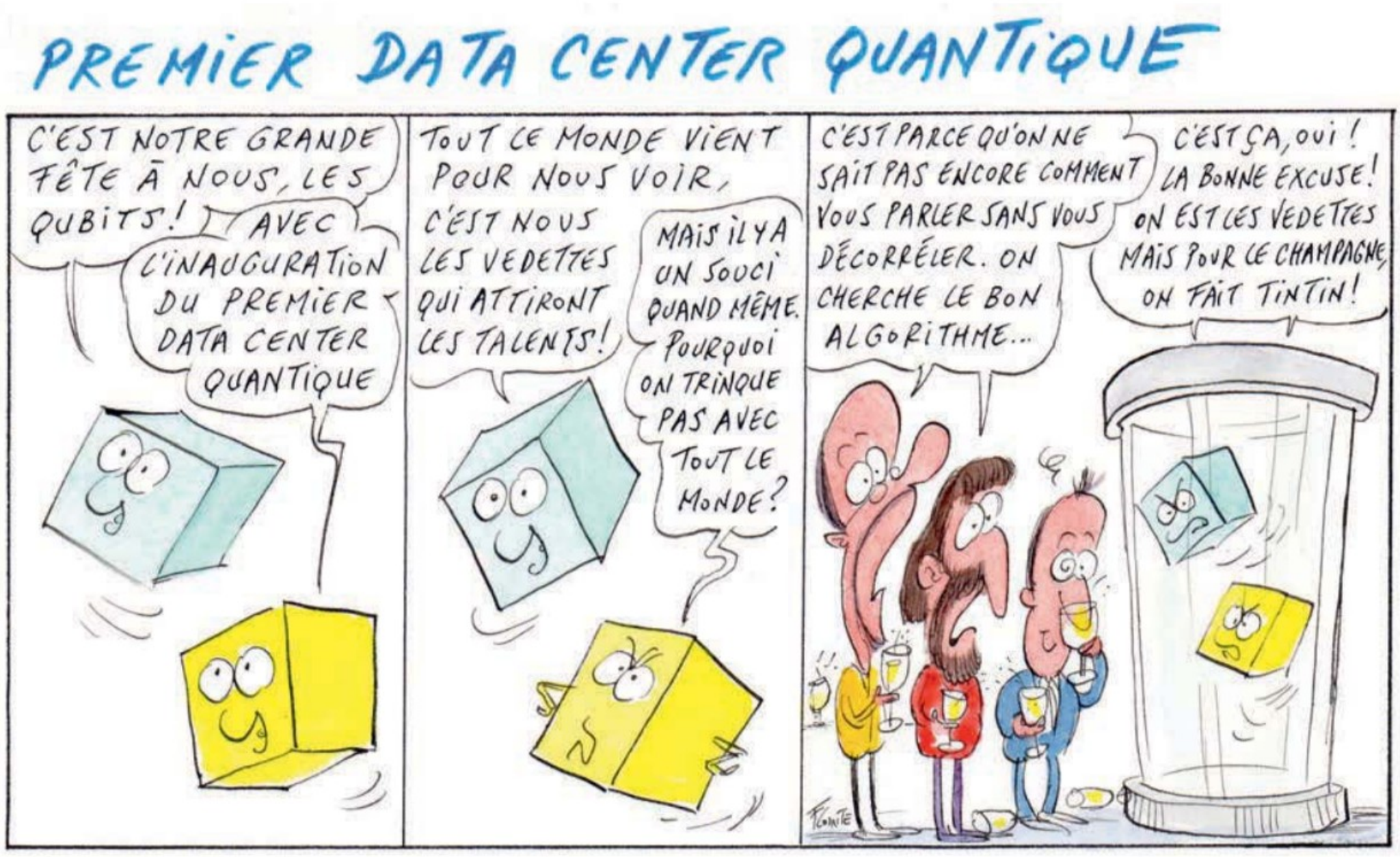
Un premier data center quantique en Europe

IBM a inauguré son premier centre de données quantiques (IBM Quantum Data Center) en dehors des États-Unis, en présence d'une kyrielle d'officiels.
Ce deuxième datacenter quantique de la compagnie dans le monde marque une étape clé dans l'expansion de sa flotte de systèmes quantiques avancés, accessibles à grande échelle via le Cloud.

Le premier datacenter quantique d'IBM en Europe comprend deux nouveaux systèmes IBM Quantum à une échelle utile basés sur le processeur Eagle, et il sera bientôt doté d'un nouveau système IBM Quantum basé sur le processeur Heron. Ces systèmes sont capables d'effectuer des calculs dépassant les capacités de simulation par « brute-force » des ordinateurs classiques.

Présentée pour la première fois à la fin de l'année dernière, IBM Heron est la puce quantique la plus performante de la compagnie à ce jour. Elle fait progresser la mission d'IBM, consistant à mettre l'informatique quantique utile à la portée du monde entier, en permettant aux utilisateurs d'accroître la complexité des algorithmes qu'ils explorent sur du vrai matériel quantique. Il s'agira du troisième IBM Heron installé dans la flotte de systèmes quantiques d'IBM auxquels le réseau quantique mondial de la compagnie (IBM quantum network), composé de plus de 250 entreprises, universités, instituts de recherche et organisations, peut accéder. IBM Heron permet de multiplier par 16 les performances et par 25 la vitesse des précédents ordinateurs quantiques d'IBM, tels qu'ils ont été mesurés il y a deux ans. Lorsqu'il sera déployé aux côtés des systèmes à une échelle utile désormais disponibles, installés dans le nouveau datacenter quantique d'IBM, le système basé sur IBM Heron viendra s'ajouter à la douzaine d'ordinateurs quantiques qu'IBM propose actuellement via le Cloud, soit la plus grande flotte de ce type dans le monde. Le datacenter quantique d'IBM en Europe est accessible via la plateforme IBM Quantum, poursuivant la mission d'IBM consistant à permettre le développement de cas d'usage de l'informatique quantique et à aider les clients à progresser dans la découverte d'algorithmes à l'ère de l'utilité quantique et vers l'avantage quantique. □

B.G



IA Les acquisitions déguisées, nouvelle astuce des géants du numérique

Google, Amazon et Microsoft ont tous récemment investi dans de jeunes pousses de l'IA dont ils ont aussitôt capté les meilleurs talents.

Une tendance prend de l'ampleur depuis quelques mois au sein de la Silicon Valley : un géant de la tech investit lourdement dans une jeune pousse, généralement spécialisée dans l'IA. Peu de temps après, il recrute ses fondateurs, ses meilleurs ingénieurs, et récupère sa technologie, ne laissant dans son sillage qu'une coquille vide.

Ces acquisitions déguisées, ou « acqui-hires » en anglais, connaissent une popularité croissante alors que les big techs se livrent une compétition acharnée sur l'intelligence artificielle (IA), domaine dans lequel les meilleurs talents se monnaient à prix d'or. Mais elles commencent aussi à attirer l'attention des autorités anti-monopole, et ce des deux côtés de l'Atlantique.

Tout acheter, sauf l'entreprise

En mars dernier, Microsoft a ainsi investi plus de 650 millions de dollars dans la jeune pousse Inflection AI, en échange d'un accord de licence. Le géant de l'informatique a ensuite recruté la quasi-totalité du personnel de la start-up, y compris Mustafa Suleyman, l'un des cofondateurs de DeepMind, pépite britannique rachetée par Google en 2014, et devenue le laboratoire de recherche du groupe sur l'IA après sa fusion avec Google Brain. Mustafa Suleyman dirige désormais la stratégie IA de Microsoft.

En juin, Amazon a mené une opération similaire auprès d'Adept AI, une autre jeune pousse de l'IA basée à San Francisco. Le géant du commerce en ligne a versé 330 millions de dollars à la start-up en échange d'une licence pour sa technologie, avant de débaucher le dirigeant, David Luan, un ancien d'OpenAI et de Google, ainsi que les chercheurs en IA de la jeune pousse.

Un mois plus tard, Google déboursait à son tour trois milliards de dollars pour la technologie de Character. AI, une jeune pousse construisant un chatbot basé sur l'IA générative. Lancée en 2022 par deux anciens ingénieurs de… Google, Character. AI avait déjà conquis plusieurs millions d'utilisateurs, levé près de 200 millions de dollars, et atteint une valorisation estimée à environ un milliard. Noam Shazeer et Daniel De Freitas, les deux fondateurs, ont dans la foulée réintégré Google, en compagnie d'environ 20 % de leurs effectifs, probablement les ingénieurs chargés de concevoir le LLM de la jeune pousse. Les effectifs restants au sein de celle-ci ont, dans la foulée, abandonné l'idée de construire un LLM et ont décidé de lancer rapidement un produit commercial en s'appuyant sur des modèles existants sur le marché.

Un moyen de passer sous les radars des autorités antitrust

Selon Mark Lemley, professeur de droit des technologies à Stanford, l'objectif est de capter les meilleurs talents et technologies du marché sans trop attirer l'attention des autorités anti-monopole avec une acquisition en bonne et due forme. *« Il s'agit de mettre la main sur quelques individus exceptionnels. Ce type d'arrangement n'est en soi pas nouveau, mais a tendance à se multiplier dans le secteur de l'IA, à l'heure où les fusions-acquisitions sont passées au peigne fin par les autorités. »*

Une vue de la manière dont Adept AI reconfigure un site web comme celui d'AirBnB pour afficher instantanément les résultats les plus pertinents pour chaque visiteur du site.

Les autorités anti-monopole ont en effet musclé leur approche vis-à-vis des géants technologiques au cours des dernières années. Aux États-Unis, d'abord, où Joe Biden a décidé de donner un tour de vis contre la domination des big techs. Il a ainsi nommé plusieurs personnalités fermement engagées dans la lutte contre les monopoles technologiques à des postes clef : en particulier, Lina Khan à la tête de la FTC, gendarme de la concurrence, et Jonathan Kanter pour diriger la division anti-monopole du Département de la Justice.

Les agences fédérales américaines se sont lancées, à la suite de cela, dans une série d'offensives à l'encontre des acquisitions menées par

les big techs, accusés de racheter les sociétés susceptibles de leur faire de l'ombre pour entretenir leur domination. La FTC est ainsi en train de passer au peigne fin le rachat de WhatsApp et Instagram par Facebook, ainsi que celui de deux entreprises publicitaires par Google, également ciblé par un procès anti-monopole pour sa domination sur le marché de la publicité en ligne. Conscientes d'avoir laissé des monopoles se former en ne régulant pas assez vite des domaines comme la recherche en ligne et les réseaux sociaux, les autorités antitrust entendent ne pas reproduire la même erreur avec l'IA. Dans un blog publié en juillet dernier, Marc Andreessen, codirigeant du célèbre fonds d'investissement en capital-risque Andreessen Horowitz et soutien de Donald Trump, écrivait, frustré, que *« les agences de régulation empêchent les start-ups de se faire racheter par les grandes entreprises »*.

La tendance n'est toutefois pas cantonnée aux États-Unis. En Europe, la Competition and Market authority (CMA) britannique et les autorités européennes ont, en effet, durci le ton contre les monopoles technologiques et ont multiplié les enquêtes. Google a par exemple récemment reçu une amende de 2,4 milliards d'euros de la Commission européenne pour avoir profité de son monopole sur la recherche en ligne afin d'étouffer la concurrence.

Mais la multiplication des acquisitions déguisées commence aussi à attirer l'attention des régulateurs. La FTC a ainsi lancé plusieurs enquêtes liées à des investissements et partenariats dans l'IA générative, dont le rapprochement entre Microsoft et OpenAI. Au Royaume-Uni, la CMA a également commencé à se pencher sur l'investissement de Microsoft dans Inflection AI, y voyant une potentielle acquisition déguisée visant à endormir la vigilance des autorités anti-monopole.

Un danger pour la créativité de la Silicon Valley ?

Ces acquisitions déguisées comportent des avantages pour les sociétés qui se font ainsi plus ou moins racheter. La course au développement de modèles d'IA de pointe coûte en effet très cher, à la fois pour recruter et payer les talents nécessaires, et pour louer de la puissance informatique via le cloud. Face à la trésorerie sans fond des big techs, les jeunes pousses ont rapidement du mal à rivaliser, et doivent tôt ou tard abandonner la recherche fondamentale pour se concentrer sur la mise sur le marché d'un produit commercial viable. En rejoignant Google, Amazon ou Microsoft, les chercheurs gagnent ainsi la possibilité de se focaliser sur leurs recherches et le développement de la meilleure technologie possible, sans impératif commercial à court terme.

Mustafa Suleyman, fondateur d'Inflection AI repris par Microsoft.

Les investisseurs ayant misé tôt sur la jeune pousse y trouvent également leur compte, puisque l'argent versé lors du rachat déguisé est généralement en partie utilisé pour leur offrir une compensation. Lors de son investissement dans Character. AI, Google a par exemple rémunéré les investisseurs en proportion de leur participation dans la jeune pousse, sur une base de 2,5 milliards de dollars de valorisation, soit bien plus que la valeur de Character. AI, estimée à un milliard seulement.

Ces « deals » d'un type nouveau ne font cependant pas que des heureux. En effet, les employés de la start-up qui ne sont pas invités à rejoindre le grand groupe se retrouvent ainsi sur le carreau, sans la compensation financière que leur aurait conféré un rachat en bonne et due forme. Ils créent ainsi une disproportion importante entre les fondateurs et chercheurs en IA d'une part, et le reste des employés de la start-up d'autre part.

Un déséquilibre qui pourrait, à terme, nuire à la capacité créatrice de la Silicon Valley. Cette dernière a en effet construit son succès sur le phénomène des serial entrepreneurs, des travailleurs de la tech qui, après un premier rachat, profitent des fonds récupérés pour lancer une nouvelle société, qui sera peut-être rachetée à son tour, et ainsi de suite. La célèbre PayPal mafia constitue l'un des exemples les plus célèbres. En grippant ce mécanisme, les acquisitions déguisées pourraient nuire au dynamisme de l'écosystème. □

G.R

Détection

Kineis et Dryad s'associent pour prévenir les incendies de forêts

Dryad Networks, un spécialiste dans la détection des incendies de forêts par des capteurs de gaz, s'associe avec Kineis, un opérateur français de satellites et fournisseur mondial de connectivité pour l'Internet des objets (IoT).

Les capteurs de gaz de Dryad, alimentés par l'énergie solaire, et dotés d'une intelligence artificielle, surveillent et détectent les signes précurseurs des incendies de forêt en analysant la qualité de l'air et en identifiant des schémas gazeux uniques indiquant la présence d'incendies. Une fois intégré à la constellation de satellites de Kinéis, ce réseau de capteurs s'affranchira des limites terrestres et assurera une connectivité mondiale en quasi-temps réel.

Cette synergie permet la détection précoce des incendies de forêt, même dans les zones reculées dépourvues de couverture terrestre, et l'envoi d'alertes critiques aux autorités via les 25 satellites de la constellation Kinéis, ce qui améliore considérablement les capacités d'intervention en cas d'urgence et contribue à la préservation des zones naturelles et à la protection des infrastructures essentielles. Auparavant, le déploiement d'une couverture complète de capteurs dans ces scénarios pouvait s'avérer difficile en raison des limites de la couverture des réseaux terrestres. L'intégration par Dryad de la technologie satellitaire de Kinéis relève ce défi, en permettant une communication directe capteur-satellite simplifiée et rentable pour une détection et une prévention inégalées des incendies de forêt le long des corridors critiques. En particulier, les installations d'infrastructures linéaires telles que les lignes électriques ou les voies ferrées, ou sur des terrains difficiles tels que les régions montagneuses, bénéficieront de la connectivité directe avec les satellites. Grâce à une installation considérablement simplifiée, les capteurs de feux de forêt Silvanet de Dryad deviendront encore plus rentables.

Lorsqu'elle est déployée en combinaison avec les passerelles LoRaWAN Mesh de Silvanet, qui offrent des capacités supplémentaires, la connectivité IoT spatiale

Un capteur Silvanet de Dryad.

de Kinéis garantira un service ininterrompu même si la connectivité terrestre est interrompue ou si les capteurs sont placés en dehors de la portée des passerelles. Dryad et Kinéis ont pour objectif de déployer des centaines de milliers de ces terminaux IoT dans le monde entier au cours des trois prochaines années. S'appuyant sur les chipsets avancés de Semtech, qui combinent LoRaWAN terrestre et la connectivité satellite de Kinéis en une seule solution, cette intégration garantit aux utilisateurs la possibilité de déployer des produits dotés à la fois d'une connectivité LPWAN terrestre et d'une connectivité satellite. Les premiers produits basés sur cette intégration devraient être disponibles avant la fin de l'année.□

B.G

MÉGA-INCENDIES

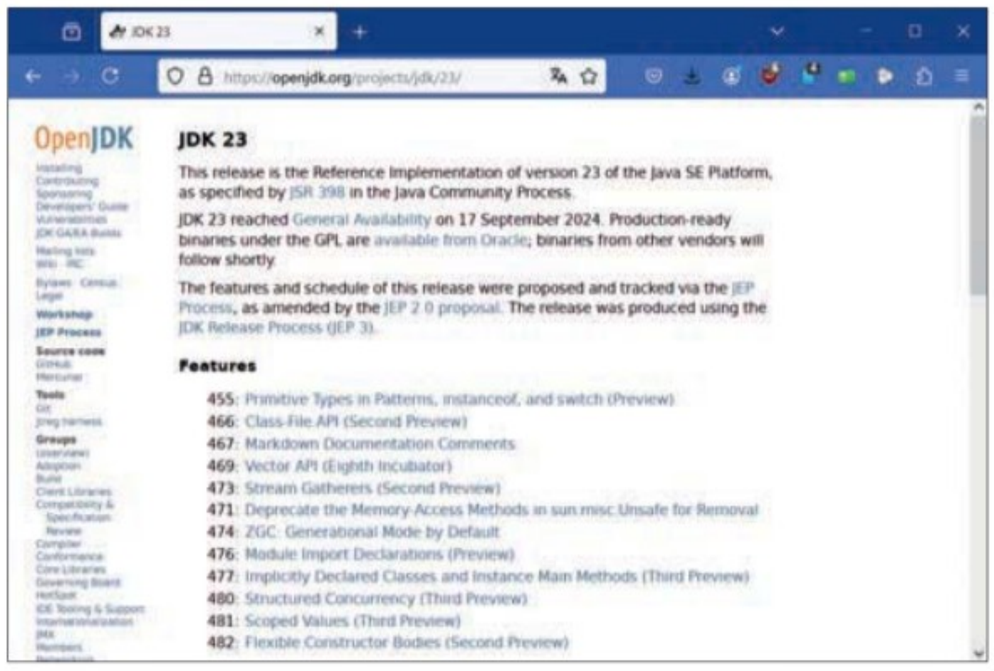
Depuis quelques années, les forêts, qui couvrent aujourd'hui 31% des terres émergées de notre planète, sont victimes de vastes incendies provoqués par le réchauffement climatique et l'activité humaine. Ces méga-incendies ravagent des milliers d'hectares de forêts à travers la planète, représentant chaque année 20% des émissions mondiales de carbone.

JAVA

Le nouveau JDK est arrivé

Oyez oyez braves développeurs, venez vous déhancher sur la 23ème Java !
La nouvelle danse à la mode est arrivée dans les salons. Nous allons voir dans cet article quels nouveaux pas vous devrez apprendre pour rester à la mode du jour.

Le JDK 23 vient tout juste de sortir le 17 septembre dernier. Il est accompagné notamment d'une seconde preview de l'API de fichier de classe, d'une deuxième preview de Stream Gatherers (collecteur de flux) dont la première preview avait été livrée avec le JDK 22, de l'API Vectorielle qui en est tout de même à sa huitième incubation ainsi que d'autres fonctionnalités plutôt intéressantes comme les types primitifs dans les patterns, instanceof et switch. Cette version fait suite à la LTS Java 21 et précède la prochaine LTS attendue, Java 25, prévue elle pour septembre 2025.



Vous trouverez tous les détails des nouvelles fonctionnalités de Java 23 ainsi que la documentation complète sur le site d'open JDK à l'adresse : https://openjdk.org/projects/jdk/23/

L'API vectorielle

L'API vectorielle a été incubée dans les précédentes versions de Java depuis la sortie du JDK 16 jusqu'à la version 22. Cette nouvelle version introduit une API aidant à mieux exprimer les calculs vectoriels qui sont compilés à l'exécution. Son objectif est de mettre à disposition des instructions vectorielles optimales, ce quelles que soient les architectures de CPU prises en charge.

La proposition pour le JDK 23 concernant cette API vectorielle comprend :

- la fourniture d'une API claire et concise : l'API doit être capable d'exprimer de manière claire et concise une large gamme de calculs vectoriels consistant en des séquences d'opérations vectorielles composées de boucles et éventuellement d'un flux de contrôle. Il doit être possible d'exprimer un calcul générique en ce qui concerne la taille du vecteur ou le nombre de voies par vecteur. Cela doit permettre à ces calculs d'être portables sur du matériel prenant en charge différentes tailles de vecteurs.
- la fourniture d'une plateforme agnostique : l'API doit être agnostique vis-à-vis de l'architecture du processeur afin de permettre des implémentations sur de multiples architectures supportant les instructions vectorielles. Comme c'est habituellement le cas pour les API Java, lorsque l'optimisation de la plate-forme et la portabilité sont en conflit, c'est la portabilité de l'API qui est privilégiée, quand bien même cela se traduirait par des instructions spécifiques avec du code non ou moins portable.
- une compilation et des performances fiables du runtime sur les architectures x64 et AArch64 : sur les architectures x64 compatibles, le runtime Java et plus particulièrement le compilateur HotSpot C2 devrait compiler les opérations vectorielles avec les instructions vectorielles idoines telles que celles prises en charge par les extensions SIMD en continu (SSE) et les extensions vectorielles avancées (AVX). Sur les architectures ARM AArch64 par exemple, HotSpot C2 compilera de la même manière ces opérations vectorielles pour travailler avec des instructions vectorielles supportées par NEON et REV.
- dégradation progressive : Cependant, il arrive parfois qu'un calcul vectoriel ne puisse pas être entièrement exprimé sous la forme d'une séquence d'instructions vectorielles. Une des raisons possibles peut-être que l'architecture ne prend pas en charge certaines de ces instructions. L'implémentation de l'API vectorielle devra alors, dans ce cas, se dégrader « avec parcimonie » et surtout continuer à fonctionner. Cela peut impliquer l'émission d'avertissements si un calcul vectoriel ne peut pas être compilé efficacement en instructions vectorielles. Sur les plates-formes dépourvues de vecteurs, cette « dégradation gracieuse » produira un code restant compétitif avec des boucles déroulées manuellement et pour lesquelles le facteur de déroulement sera le nombre de voies dans le vecteur sélectionné.
- Alignement sur le projet Valhalla ; l'objectif à long terme de l'API vectorielle est de tirer au maximum parti des améliorations apportées par le projet Valhalla au modèle d'objet Java

Collecteurs et API de flux

Les Stream Gatherers, qui avaient été présentés en preview dans la version 22 du JDK, ont été entièrement ajoutés au JDK 23. Ces collecteurs de flux ont pour but d'améliorer l'API Stream pour prendre en charge les opérations personnalisées. Ils permettent également aux pipelines de flux de transformer les données plus simplement qu'avec les opérations intégrées actuelles. L'objectif recherché est de rendre les pipelines de flux beaucoup plus flexibles et expressifs en vue de permettre également aux opérations personnalisées de manipuler des flux de taille infinie. L'API Stream introduite depuis Java

8 offre la possibilité d'effectuer des opérations de manière déclarative sur une séquence d'éléments.

Un Stream peut être découpé en trois étapes :

1. La création à partir d'une collection, d'un tableau ou en utilisant la méthode Stream.of().
2. Les opérations intermédiaires qui transforment le Stream comme filter, limit, map.
3. Les opérations terminales qui finissent le Stream telles que collect, reduce ou forEach.

L'API Stream existante offrait déjà un ensemble défini d'opérations intermédiaires et terminales et proposait d'étendre les opérations terminales via la méthode Stream : collect(Collector). Cependant, elle n'offrait pas la possibilité d'étendre les opérations intermédiaires. Les Streams Gatherers introduisent désormais la possibilité de définir ses propres opérations intermédiaires grâce à la méthode Stream : gather(Gatherer). De plus, la classe java.util.stream. Gatherers met à disposition plusieurs Gatherers prédéfinis, tels que fold, mapConcurrent et windowFixed.

CLASSES IMPLICITEMENT DÉCLARÉES

Les spécificités suivantes s'appliquent aux classes implicites :

- **Une classe implicite se situe toujours dans le package sans nom (tout comme une classe normale sans package/déclaration).**
- **Une classe implicite est toujours finale.**
- **Une classe implicite n'est pas capable d'implémenter d'interfaces ni d'hériter d'autres classes. De même, aucune classe n'hérite d'une classe implicite.**
- **Une classe implicite n'est pas accessible via le nom donné par le compilateur, c'est-à-dire que les autres classes n'instancient pas une classe implicite et n'y appellent pas de méthodes, pas même statiques.**

Cependant, une classe implicite est capable d'appeler des méthodes sur elle-même, comme dans l'exemple suivant :

```
void main() {
System.out.println(greeting()) ;
}
String greeting() {
return « Hello, World !";
}
```

Puisqu'une classe implicite n'est pas accessible de l'extérieur, elle doit contenir une méthode main().

Types primitifs dans les modèles

Le JDK 23 apporte une autre fonctionnalité en preview qui semble valoir la peine d'être mise en avant. Il s'agit des types primitifs dans les patterns, instanceof et switch. De nombreuses restrictions relatives aux types primitifs provoquent des problèmes lors de l'utilisation de la recherche de patterns avec les instructions instanceof et switch. L'élimination de ces restrictions rendrait le langage Java bien plus uniforme et expressif. La première restriction est que la recherche de patterns pour switch ne prend pas en charge les patterns de type qui spécifient un type primitif. Seuls les patterns de type qui spécifient un type de référence sont pris en charge, tels que case Integer i ou case String s. Depuis Java 21, les patterns d'enregistrement sont également pris en charge pour switch. Cette prise en charge des patterns de types primitifs dans switch permet donc d'améliorer l'expression de ces instructions conditionnelles (ou plutôt « embranchements »), comme ici :

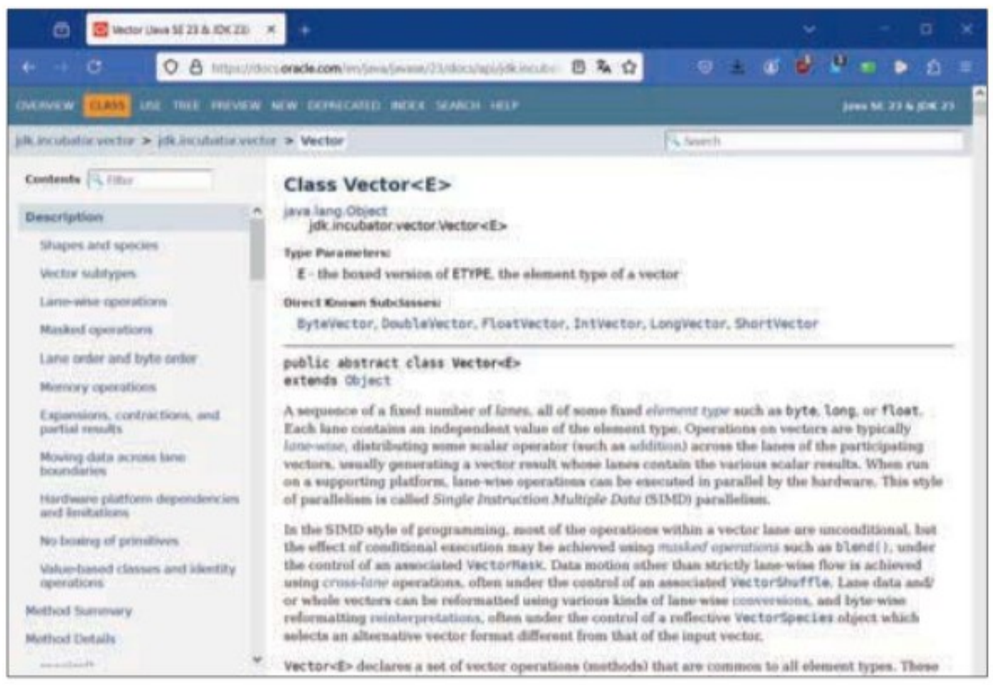
```
switch (x.getStatus()) {
case 0 -> "ok";
case 1 -> "avertissement";
case 2 -> "erreur";
default -> « statut inconnu : " + x.getStatus() ;
}
```

en transformant la clause default en une clause case avec un modèle de type primitif qui expose la valeur correspondante :

```
switch (x.getStatus()) {
case 0 -> "ok";
case 1 -> "avertissement";
case 2 -> "erreur";
case int i -> « statut inconnu : " + i ;
}
```

La prise en charge des patterns de type primitif permet également aux « gardes » d'inspecter la valeur correspondante :

```
switch (x.getYearlyFlights()) {
case 0 ->...;
case 1 ->...;
case 2 -> issueDiscount() ;
case int i when i >= 100 -> issueGoldCard() ;
case int i -> << action appropriée si i > 2 && i < 100 >>
}
```



L'API vectorielle aide à mieux exprimer les calculs vectoriels qui sont compilés à l'exécution. Son objectif est de mettre à disposition des instructions vectorielles optimales, et ce, quelles que soient les architectures de CPU prises en charge.

En résumé, cette fonctionnalité améliore considérablement la recherche de patterns en autorisant les patterns de type primitif. Elle fournit de plus des constructions faciles à utiliser tendant à éliminer le risque de perdre des informations en raison de casts mal construits et par conséquent, pas suffisamment sûrs. Faisant enfin suite aux améliorations qui avaient été apportées à switch dans Java 5 (switch enum) et Java 7 (switch string), cela va permettre de traiter des valeurs de n'importe quel type primitif avec cette instruction.

AUTRES FONCTIONNALITÉS

D'autres fonctionnalités présentées en preview dans le JDK 22 pourraient être intégrées progressivement dans le JDK 23. Il s'agit notamment :

- **des déclarations précédant l'instruction super() qui donneraient aux développeurs une plus grande liberté dans l'expression du comportement des constructeurs ;**
- **des modèles de chaînes qui faciliteraient l'expression de chaînes comprenant des valeurs calculées au moment de l'exécution ;**
- **des valeurs « scopées », qui permettraient le partage de données immuables au sein et entre les threads ;**
- **des classes et méthodes principales d'instance déclarées implicitement, qui permettraient aux développeurs débutants d'écrire plus facilement des programmes sans avoir besoin de comprendre les caractéristiques du langage conçues pour les programmes de grande taille.**

L'API Class-File

L'API class-file a pour fonction le traitement des fichiers de classe en suivant le format de fichier de classe défini par la spécification de la JVM (machine virtuelle Java). Son but est de permettre aux composants du JDK de migrer vers l'API standard afin de supprimer la copie de la bibliothèque ASM du JDK. Elle ajoute quelques ajustements, comme notamment la rationalisation de la classe CodeBuilder. Cette classe contient par défaut des méthodes d'usine pour les instructions de bytecode, y compris des usines (fabriques) de bas niveau, des usines de niveau moyen et des constructeurs de haut niveau pour les blocs de base. Voici un exemple de code d'un CodeBuilder ASM et d'un CodeBuilder Java. Supposons que vous souhaitiez générer la méthode suivante dans un fichier de classe :

```
void fooBar(boolean z, int x) {
if (z)
foo(x) ;
else
bar(x) ;
}
```

Avec la bibliothèque ASM, vous pourriez générer la méthode de cette manière :

```
ClassWriter classWriter =...;
MethodVisitor mv = classWriter.visitMethod(0, "fooBar", "(ZI)V », null,
null) ;
mv.visitCode() ;
mv.visitVarInsn(ILOAD, 1) ;
Label label1 = new Label() ;
mv.visitJumpInsn(IFEQ, label1) ;
mv.visitVarInsn(ALOAD, 0) ;
mv.visitVarInsn(ILOAD, 2) ;
mv.visitMethodInsn(INVOKEVIRTUAL, "Foo", "foo", "(I)V », false) ;
Label label2 = new Label() ;
mv.visitJumpInsn(GOTO, label2) ;
mv.visitLabel(label1) ;
mv.visitVarInsn(ALOAD, 0) ;
mv.visitVarInsn(ILOAD, 2) ;
mv.visitMethodInsn(INVOKEVIRTUAL, "Foo", "bar", "(I)V », false) ;
mv.visitLabel(label2) ;
mv.visitInsn(RETURN) ;
mv.visitEnd() ;
```

Dans ASM, le type MethodVisitor sert à la fois de « visiteur » et de constructeur. Les clients peuvent créer un ClassWriter directement, puis à partir de là générer un MethodVisitor à l'aide de la méthode visitMethod. L'API Class-File inverse ce principe. Au lieu de créer un constructeur avec un constructeur ou une fabrique, le client doit fournir une lambda qui accepte un constructeur :

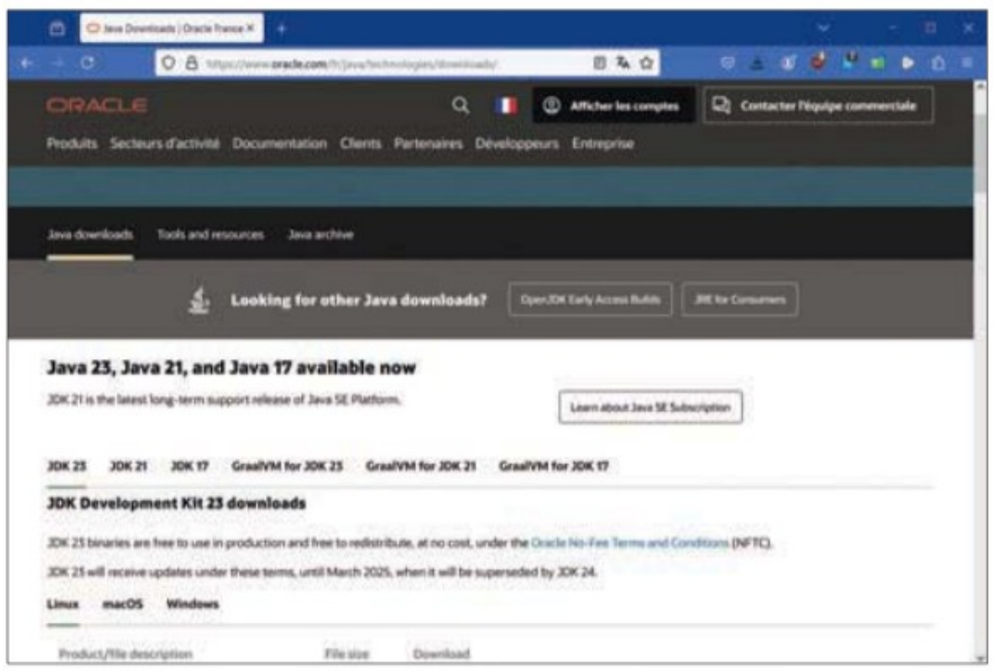
```
ClassBuilder classBuilder =...;
classBuilder.withMethod("fooBar", MethodTypeDesc.of(CD_void, CD_
boolean, CD_int), flags,
methodBuilder -> methodBuilder.withCode(codeBuilder -> {
Label label1 = codeBuilder.newLabel() ;
Label label2 = codeBuilder.newLabel() ;
codeBuilder.iload(1)
.ifeq(label1)
.aload(0)
.iload(2)
.invokevirtual(ClassDesc.of("Foo"), "foo", MethodTypeDesc.of(CD_
void, CD_int))
.goto_(label2)
.labelBinding(label1)
.aload(0)
.iload(2)
.invokevirtual(ClassDesc.of("Foo"), "bar", MethodTypeDesc.of(CD_
void, CD_int))
.labelBinding(label2) ;
.return_() ;
}) ;
```

Dans le JDK 23, le Java CodeBuilder supprime les méthodes de niveau intermédiaire qui font double emploi avec les méthodes de bas niveau ou bien qui, dans le pire des cas, sont très rarement utilisées. Les méthodes de niveau intermédiaire restantes sont renommées pour en améliorer la convivialité. Le modèle de classe ClassSignature a été, lui aussi, affiné. Il a été amélioré dans le but de modéliser avec une plus grande précision les signatures génériques des superclasses et des superinterfaces. D'après la proposition OpenJDK à l'origine de cette fonctionnalité, la plateforme Java devrait définir et mettre en œuvre une API de fichier de classe standard qui évoluera en même temps que le format de fichier de classe et qui pourra facilement changer ou évoluer tous les six mois approximativement.

Markdown Documentation Comments (JEP 467)

Jusqu'à présent, pour formater des commentaires JavaDoc, il était nécessaire d'utiliser HTML. Aujourd'hui, Markdown est devenu bien plus populaire pour rédiger de la documentation. Cette JEP apporte donc la possibilité d'écrire des commentaires JavaDoc avec un formatage Markdown. L'exemple ci-dessous présente un extrait de la documentation de la méthode java.lang. Object.hashCode.

```
/**
* Returns a hash code value for the object. This method is
* supported for the benefit of hash tables such as those provided by
* {@link java.util. HashMap}.
* <p>
* The general contract of {@code hashCode} is :
* <ul>
* <li>Whenever it is invoked on the same object more than once
during
*.....
.....
* <li>If two objects are equal according to the {@link
* #equals(Object) equals} method, then calling the {@code
*.....
.....
* <li>It is <em>not</em> required that if two objects are unequal
*.....
.....
* unequal objects may improve the performance of hash tables.
* </ul>
*
* @implSpec
* As far as is reasonably practical, the {@code hashCode} method
defined
* by class {@code Object} returns distinct integers for distinct
objects.
*
* @return a hash code value for this object.
* @see java.lang. Object#equals(java.lang. Object)
* @see java.lang. System#identityHashCode
*/
```



Le JDK 23 est disponible au téléchargement pour toutes les architectures matérielles depuis le site d'Oracle à l'adresse : https://www.oracle.com/fr/java/technologies/downloads/

Et voici la même documentation en version Markdown :

```
/// Returns a hash code value for the object. This method is
///.....
/// The general contract of `hashCode` is :
///
/// — Whenever it is invoked on the same object more than once
during
///.....
/// — If two objects are equal according to the
/// [equals][#equals(Object)] method, then calling the
///.....
/// — It is _not_ required that if two objects are unequal
///.....
/// @implSpec
/// As far as is reasonably practical, the `hashCode` method defined
///.....
/// @return a hash code value for this object.
/// @see java.lang. Object#equals(java.lang. Object)
/// @see java.lang. System#identityHashCode
```

Vous remarquerez quelques différences :

- L'utilisation de Markdown est indiquée par un nouveau format de commentaire de documentation : les lignes commencent par /// à la place de la syntaxe traditionnelle /**... */
- À la place de {@code...}, le code source est marqué par '...'
- Les liens, noté {@link... } en Javadoc HTML, sont désormais noté par [...]
- Le tag HTML <p> a été remplacé par une ligne vide
- Les tags d'énumérations <ul> et <li> sont remplacés par la liste à puces Markdown
- Les tags de détails spécifiques à la JavaDoc, comme @implSpec, @return, et @see restent inchangés

ZGC : Generational Mode by Default (JEP 474)

Cette JEP établit le mode générationnel comme mode par défaut pour le Z Garbage Collector (ZGC), tout en dépréciant le mode non-générationnel. Le mode générationnel étant généralement plus performant que le mode non-générationnel, il deviendra le principal axe des futurs développements.

Et Oracle dans tout ça...

Le « proprio » de Java, Oracle, a également dévoilé ses projets pour Java en 2024. Oracle a présenté des améliorations impliquant des projets OpenJDK allant d'Amber, pour le développement de petites fonctionnalités axées sur la productivité, à Babylon, pour l'extension de Java à des modèles de programmation étrangers tels que les GPU, en passant par Valhalla, pour l'augmentation du modèle d'objet Java avec des objets de valeur afin d'éliminer les goulets d'étranglement datant de longue date en matière de performances. □

T.T

Mainframe

La pierre de touche de l'IA et de l'hybride

Pour la deuxième année consécutive, Kyndryl a commissionné Coleman Parkes Research pour interroger 500 entreprises utilisatrices de systèmes centraux comme des mainframes. Les principaux résultats indiquent que celles-ci font du mainframe la plateforme de prédilection pour les charges hybrides et d'intelligence artificielle.

Premier enseignement de ce sondage, 86 % des entreprises souhaitent ou ont déployé des charges d'intelligence artificielle, même générative, sur des mainframes. 71 % indiquent qu'elles ont déjà mis en œuvre des enseignements tirés de l'intelligence artificielle générative comme élément de la modernisation de leur stratégie autour de ce type de serveur. Plus d'un tiers des répondants (36 %) voient dans cette technologie une opportunité d'investissement pour rajeunir leurs grands systèmes. En comparaison de la vague précédente de l'étude, ce choix de l'IA dans la décision d'investissement dépasse les thèmes précédents autour de la sécurité, de la résilience et de la performance. Dans l'utilisation de l'intelligence artificielle générative, 41 % des personnes interrogées créent des actions opérationnelles plus rapides, répétables et moins sujettes à l'erreur humaine. Un tiers (33 %) l'utilisent pour améliorer l'expérience client, comme par exemple avec une personnalisation plus poussée et plus complète de la relation avec lui. Avec la possibilité de proposer plusieurs réponses, 44 % indiquent une utilisation pour débloquer leurs données importantes et transformer leurs données non structurées en informations exploitables. Un tiers cite l'aide au développement de nouveaux produits et services. Il faut cependant tempérer ces affirmations alors que 80 % des entreprises interrogées ne sont qu'au début ou au milieu de déploiement de cas concrets. Plus d'un tiers (36 %) sont toujours en phase d'exploration sur les cas d'usages possibles. 14 % ne pensent pas utiliser la technologie dans un futur proche et citent comme raison des inquiétudes sur la sécurité (41 %) ou des questions de conformité réglementaire (35 %).

Appel à des ressources extérieures pour moderniser les mainframes.

Comme chaque année, Kyndryl a livré son étude sur la modernisation des mainframes.

L'Hybride est là pour rester

Si 89 % des personnes interrogées indiquent que le mainframe reste essentiel pour leurs opérations, quasiment toutes ont migré des workloads hors du système. À 36 %, les entreprises expliquent que c'est pour combiner à la fois les bénéfices du mainframe et du Cloud. Trois scénarii se dégagent : le premier est de sortir du mainframe, le second est d'intégrer

le mainframe avec d'autres plateformes, le troisième est de moderniser les workloads sur le mainframe. Le premier scénario s'accompagne parfois du maintien du langage utilisé sur le mainframe ou du refactoring des applications sur des langages plus modernes comme Java. Il existe certains exemples de réarchitecture des applications de réécriture de l'application et des migrations de données ultérieures. Le deuxième scénario est bien connu avec l'intégration du mainframe avec d'autres plateformes ou le Cloud. L'émergence de l'utilisation de l'intelligence artificielle est vue comme un moteur de ce type de scénario. Le troisième scénario est celui qui représente la principale tendance actuelle qui place au centre une rationalisation des applications présentes sur le système avec les possibilités de revoir le code des applications ou d'exploiter de nouvelles technologies comme l'intelligence artificielle ou la containerisation des applications. Cette modernisation passe par une intégration accrue avec le Cloud qui surpasse désormais le scénario de migration hors du mainframe.

Sécurité et conformité restent un bénéfice important

La réputation de la plateforme en termes de sécurité n'est plus à faire et reste un des points forts pour la prise de décision de continuer sur le mainframe. 49 % des répondants citent la sécurité comme un des critères importants dans leur décision d'investissement dans la plateforme. 66 % citent cette fonction comme la plus importante et 35 % la mettent en valeur dans la stratégie de modernisation du mainframe.

Cette fonction est de plus appréciée pour le respect des différentes contraintes réglementaires des clients (DORA, NIS2, et SEC). 28 % des personnes interrogées indiquent investir plus dans la modernisation de ce système pour y répondre. La majorité (45 %) intègre la démarche par intégration avec le Cloud et seulement 19 % migrent des tâches pour se conformer aux différentes règles.

Le frein des compétences

Le manque persistant de compétences sur le mainframe et les nouvelles technologies d'intelligence artificielle sont toujours un frein dans les opérations de modernisation de la plateforme dans les entreprises. 18 % pointent cette difficulté dans la réussite des projets. 28 % indiquent ne pas avoir les compétences d'un assez haut niveau pour effectuer cette modernisation. La moitié des répondants contournent le problème par des embauches ou des remises à niveau de leurs ressources internes. Les principaux investissements en formation sont réalisés sur la mise en conformité et sur l'intelligence artificielle. 43 % déplorent cependant ne pas avoir les compétences en IA pour bénéficier de la technologie dans leur stratégie de modernisation du mainframe. 45 % ont de plus connu des difficultés pour trouver des spécialistes en sécurité.

L'observabilité comme tour de contrôle

Devant les environnements hybrides largement présents, 92 % des entreprises ayant répondu indiquent qu'il est important d'avoir un seul dashboard pour suivre les opérations, mais 85 % trouvent cette possibilité difficile à réaliser. Cela permettrait cependant d'avoir une vue précise sur la performance de l'infrastructure (54 %) et des applications (45 %) ou de la gestion des services IT (49 %).

Après 60 ans d'existence, le mainframe reste un composant critique pour de nombreuses entreprises et sa modernisation est au cœur de la stratégie des utilisateurs de la plateforme visant à combiner les forces du mainframe avec celles des autres environnements comme le Cloud afin de profiter de plus de flexibilité et de mise à l'échelle par le Cloud. □

B.G

Le mainframe reste toujours un élément important dans les entreprises.

Elus locaux

Le Wagon forme les élus de Thiais

L'institut de formation aux métiers de la tech et de l'intelligence artificielle innove en lançant une initiative inédite : la formation des élus locaux de la commune de Thiais.

Une formation à l'IA à Thiais.

Cette démarche vise à préparer les décideurs locaux aux défis et opportunités que présente l'IA pour leur territoire, tout en mettant l'accent sur l'éthique et la responsabilité. Le Wagon a conçu un programme sur mesure pour les décideurs locaux. Cette formation se concentre sur l'aspect pratique de l'IA et son utilisation éthique, permettant aux élus et aux cadres de saisir pleinement les enjeux de cette technologie transformative pour leur territoire. Le programme débute par une immersion dans la pratique de l'IA, où les participants apprennent à maîtriser la rédaction de prompts efficaces. Ils découvrent comment formuler des requêtes précises pour obtenir des résultats pertinents des systèmes d'IA, une compétence cruciale pour l'utilisation optimale de ces outils dans la gestion municipale. Les élus et les cadres s'exercent sur des cas concrets, tels que l'analyse de données urbaines ou la génération de rapports synthétiques sur des problématiques locales, toujours en gardant à l'esprit les implications éthiques de ces pratiques. Cette approche permet aux décideurs locaux de développer un sens critique aigu face aux solutions d'IA proposées et de s'assurer que leur mise en œuvre respecte les valeurs fondamentales de leur communauté.

Une impulsion en interne

À l'origine de cette initiative, Alexandre Caussignac, adjoint au maire chargé de la Transformation Numérique, a joué un rôle clé dans la mise en place de cette formation. Sa vision d'une ville intelligente, adaptée aux défis du futur et respectueuse de ses citoyens a été le catalyseur de ce partenariat innovant avec Le Wagon.

La formation permet aux élus et cadres de Thiais de développer une compréhension holistique des implications de l'IA pour leur territoire. Ils apprennent à évaluer le potentiel de l'IA dans divers domaines de la gestion municipale, tout en restant attentifs aux considérations éthiques. Un accent particulier est mis sur la démystification de l'IA et son appropriation par les élus et les cadres, dans une logique de transmission aux administrés. Les participants acquièrent les compétences nécessaires pour expliquer clairement les avantages et les limites de l'IA à leurs concitoyens, favorisant ainsi une adoption éclairée et responsable de ces technologies au niveau local. Cette approche vise à créer un dialogue ouvert et transparent avec la population sur l'utilisation de l'IA dans la gestion municipale.

Cette initiative du Wagon marque une étape importante dans la préparation des villes françaises à l'ère de l'IA. En formant les élus et les cadres municipaux, Le Wagon contribue à créer un écosystème local capable de tirer le meilleur parti de ces technologies tout en restant vigilant sur leurs implications éthiques et sociétales. Cette approche proactive et responsable ouvre la voie à une adoption réfléchie de l'IA dans la gestion municipale, promettant des villes plus intelligentes, plus efficaces, mais surtout plus humaines et inclusives. □

B.G

PME Certif-ia forme les entreprises

L'intelligence artificielle est dans toutes les bouches et toutes les têtes. 85 % des dirigeants croient que l'IA leur offre un avantage concurrentiel durable sur le long terme. Il devient alors primordial de se former dans les entreprises sur cette nouvelle technologie.

Certif-ia, société fondée en 2024 par un trio de jeunes entrepreneurs, veut redéfinir la formation en IA générative pour maximiser la productivité et encourager l'innovation au sein des entreprises françaises. Les programmes de formation de l'entreprise sont spécialement conçus pour transformer les équipes en spécialistes de l'IA générative, optimiser la productivité grâce à l'adoption de nouvelles méthodologies de travail et libérer le personnel des tâches répétitives afin qu'il puisse se concentrer sur des activités à haute valeur ajoutée. Les cours sont prodigués par des experts issus des plus grandes sociétés technologiques françaises et internationales. Grâce à ses programmes de formation pensés pour chaque verticale métier. Dotée des labels La French Tech et Qualiopi, la startup est fière de proposer des formations adaptées aux besoins spécifiques des entreprises, dispensées par des experts qui transmettent les méthodes agiles et modernes de l'écosystème contemporain.

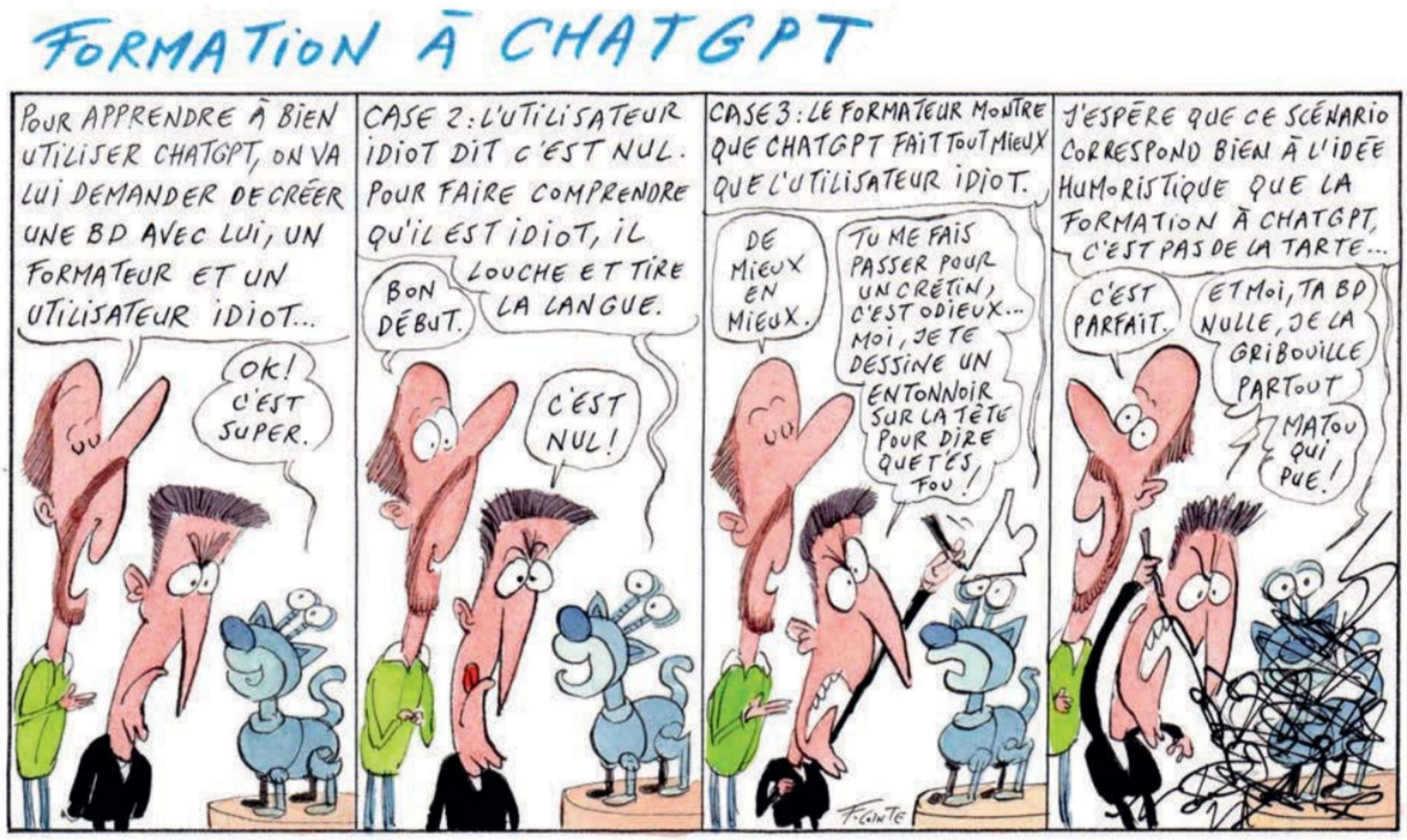
Un déroulé différenciateur

Certif.ia réalise une évaluation approfondie des besoins de chaque entreprise, la composition des équipes et l'infrastructure IT pour concevoir un programme de formation personnalisé. À la suite de cet audit, un formateur expert en IA se rend directement sur le lieu de travail pour renforcer les compétences des équipes, pour les former et intégrer l'IA dans leurs pratiques professionnelles quotidiennes. Les cours sont suivis en petits groupes de 8 à 10 personnes ou dans des masters classes de deux heures à une journée. Tout ceci est possible en présentiel ou en vidéo, en français ou en anglais. La formation ne s'arrête pas là, Certif-ia fournit à l'entreprise formée une bibliothèque de contenus qui inclut des mises à jour régulières sur les dernières évolutions en intelligence artificielle, permettant aux entreprises de maintenir un avantage compétitif. L'utilisation de l'IA par les collaborateurs de l'entreprise se solde par un gain de plus de 5h / semaine de temps gagné par participant.

De fortes ambitions

Certif-ia ne se contente pas de répondre aux besoins actuels des entreprises. Elle vise à anticiper et à former les entreprises aux cas d'usage futurs de l'IA, leur permettant ainsi de rester en tête dans un monde technologique en rapide évolution. L'ambition est claire : fournir des compétences essentielles et avancées pour tirer pleinement parti des possibilités offertes par les IA. □

B.G



Redémarrer après une cyberattaque : mythe ou réalité ?

Au-delà de la préparation, l'utilisation de la solution Dell PowerProtect Cyber Recovery vous permet de repartir suite à une cyberattaque.

Les cyberattaquants, boostés par l'IA, deviennent de plus en plus ingénieux en rendant inopérants les moyens classiques de prévention telles que les sauvegardes, causant ainsi des dégâts considérables. Leur objectif est de vous pousser à payer la rançon et revendre les informations collectées sur le Dark Web. La prévention est essentielle, mais le véritable enjeu est de pouvoir redémarrer après une attaque.

Notre technologie PowerProtect Cyber Recovery vous offre une solution unique de protection intégrant des mécanismes de sanctuarisation, d'immuabilité, et d'analyse avec un espace de redémarrage, assurant ainsi que vos données critiques soient sécurisées et récupérables.

Avec la solution Cyber Recovery de Dell Technologies, vous pouvez sanctuariser vos données et garantir une reprise efficace après un incident.

Rejoignez notre combat contre les cyberattaques.

Les postes de travail, une première ligne d'exposition criblée de brèches

Sommaire

Avec la montée en puissance du télétravail et des modèles hybrides, les terminaux, devenus nomades, cumulent les vulnérabilités, exposant davantage les organisations. Face à cette évolution du paysage, les cybercriminels adaptent leurs attaques en exploitant les failles des systèmes non gérés. Si les entreprises doivent renforcer leur arsenal de sécurité pour une protection efficace, la sensibilisation des collaborateurs est également un impératif essentiel.

Bien que le télétravail ne soit pas nouveau, il s'est démocratisé lors de la crise de la Covid-19, lorsque les entreprises ont été contraintes, du jour au lendemain, de gérer leurs activités à distance. La pandémie passée, les habitudes sont restées et nombre d'organisations ont adopté des modèles de travail plus souples pour leurs salariés. Selon l'Insee, en 2023, 47% des sociétés françaises ont instauré une part de télétravail, soit plus du double par rapport à la période prépandémie. Et si 25% des salariés le pratiquaient régulièrement en 2017, ils sont 36% en 2024, selon Statista.

Alerte : exposition critique

Cette transition n'est pas sans conséquences pour la cybersécurité des systèmes d'information. *« À l'époque, la sécurité était principalement périmétrique, ce qui permettait d'être un peu plus souple concernant la protection des postes de travail puisqu'ils restaient sur place »*, explique Éric Bohec, directeur général des opérations (COO) chez Nomios. Aujourd'hui, les postes sont, pour beaucoup, nomades. *« Ils peuvent se connecter à des réseaux d'entreprises, mais aussi à des réseaux non sécurisés ou publics. Leur exposition est devenue plus critique »*, poursuit Éric Bohec.
Le travail à distance a offert aux acteurs malveillants une surface d'attaque supplémentaire et plus vulnérable pour se frayer un accès au réseau des entreprises, notamment en usant de techniques de déplacement latéral. *« Les systèmes et réseaux des organisations ne sont pas systématiquement ciblés de façon directe, mais indirectement et de plus en plus souvent. L'un des moyens utilisés est la compromission des systèmes non gérés »*, analyse Richard de la Torre, directeur technique de Bitdefender Enterprise Solutions.
La politique du Bring Your Own Device (BYOD), qui consiste à autoriser les employés à travailler depuis leurs propres ordinateurs, smartphones ou tablettes, n'arrange rien. Les employés peuvent accéder aux données sensibles de l'entreprise et les stocker sur leurs appareils personnels, ce qui peut mener à une violation des données si ceux-ci sont perdus, volés ou compromis. *« Les organisations se soucient de mettre fréquemment à jour les microprogrammes et les logiciels des machines de leur réseau. En revanche, les systèmes domestiques sont souvent négligés, devenant alors une cible facile pour les attaques qui exploitent des failles logicielles »*, déplore Richard de la Torre.
La cohabitation sur un même appareil d'applications professionnelles et personnelles peut augmenter le risque d'exposition des données à des malwares et autres contenus malveillants. En outre, des utilisateurs non autorisés (famille, amis, etc.) peuvent utiliser cet ordinateur et accéder ainsi à des données potentiellement sensibles.

Une culture de la sécurité qui peine à s'installer

C'est donc une armée de soldats très exposés qui se dresse en première ligne. D'autant que l'acculturation des collaborateurs complique encore la problématique. Un enjeu majeur, et pour cause. En 2021, Verizon avait dévoilé les résultats d'un rapport international annuel sur les violations de données (Data Breach Investigations Report) : il y était calculé que 85% des compromissions étaient imputables à des « erreurs humaines ». Dans son rapport de 2024, il estime qu'encore plus des deux tiers (68%) des violations dans le monde impliquent une action humaine non malveillante, un pourcentage stable par rapport à 2023. La sensibilisation reste donc essentielle.
Certes, elle progresse mais lentement, alors que le risque, lui, croît de façon plus rapide. Verizon note que l'exploitation des vulnérabilités a connu une croissance de 180% par rapport à 2023. Sans même parler des disparités entre les organisations. Présentée début octobre 2024, une étude réalisée par Opinion Way pour Cybermalveillance.gouv.fr, en collaboration avec le Club Ebios, la CPME, le Medef et l'U2P, évalue le niveau de maturité cyber des TPE et PME en France. Concernant la sensibilisation des collaborateurs, l'étude met en lumière d'importantes disparités : *« Plus de la moitié (55%) d'entre elles sensibilisent leurs collaborateurs, davantage encore dans les grandes entreprises (79% des structures de 50 salariés et plus et 71% des sociétés de 10 à 49 salariés). »* Paradoxalement, bien que conscientes a priori des risques, 62% des entreprises se pensent faiblement exposées aux risques. Alors que 53% d'entre elles déclarent que leurs salariés utilisent leurs propres matériels à des fins professionnelles, tels que leur téléphone portable (95%), leur ordinateur (34%) et leur messagerie personnelle (28%).

Piéger ses collaborateurs

Les organisations disposent de plusieurs leviers pour améliorer le niveau de maturité de leurs collaborateurs. Dans son guide intitulé Sensibilisation du collaborateur, le maillon essentiel de la cybersécurité, paru en novembre 2023, le CLUSIF (Club de la sécurité de

Nicolas Cote,
responsable informatique TEHTRIS

Pour les ordinateurs personnels utilisés dans le cadre professionnel, la chose est un peu plus complexe. Il explique : *« mon expérience en tant qu'ancien RSSI et SOC manager m'a montré qu'une connexion permanente avec l'entreprise est essentielle pour assurer la sécurité. Lorsqu'un appareil est allumé et connecté à Internet, il établit automatiquement un lien avec l'organisation, et si ce lien est rompu, l'attaquant est également déconnecté. Cependant, pour les travailleurs indépendants utilisant leurs ordinateurs personnels, la responsabilité de maintenir cette connectivité et de garantir la sécurité leur incombe. Ils doivent s'assurer que leur poste de matériel est à jour, et suffisamment bien protégé. »*

« Il ne faut pas non plus oublier de chiffrer les données des postes de travail. On parle ici du chiffrement du disque dur, de manière à rendre impossible, en cas de vol du PC, toute récupération et d'exploitation des données. »

Éric Bohec,
Directeur Général des Opérations (COO) chez Nomios.

l'information en français) recommande de faire assimiler aux collaborateurs, direction comprise, la politique de sécurité du système d'information (SSI) de leur entreprise, ainsi que les conséquences des négligences, à grand renfort de vulgarisation. Le Club estime, en outre, que le salarié doit faire l'objet de sessions de sensibilisation adaptées aux risques inhérents à ses fonctions (direction, profils à privilèges, contrats temporaires, etc.) et aux flux d'informations auxquels il a accès. *« Le collaborateur a aussi une vie numérique personnelle : messagerie, réseaux sociaux, terminaux [...]. Une mauvaise utilisation des outils personnels peut entraîner une compromission de l'environnement professionnel si le même terminal est utilisé »*, souligne l'association.

Le CLUSIF propose une liste de méthodes et d'outils de sensibilisation dans son guide, allant de l'email ponctuel au e-learning et quiz, en passant par des tests et l'éducation aux bonnes pratiques, des campagnes de simulation de phishing. Les campagnes de sensibilisation à la cybersécurité, telles que les simulations d'hameçonnage, sont-elles réellement efficaces ? Mimecast a mené une étude interne auprès de 42 000 clients et a constaté que ces initiatives ne réduisaient les incidents que de 0,00015 %, un impact presque nul. Selon Sébastien Baron, Directeur Technique de Mimecast France, ces formations ciblent souvent tous les employés, alors que seules 8 à 10 % des personnes sont à risque. Il plaide plutôt pour des sensibilisations au cas par cas, se basant sur les données réelles issues des outils en production, en ciblant spécifiquement les utilisateurs à risque. Cette approche remet en question l'efficacité des méthodes traditionnelles. *« Notre solution d'Awareness Training, comme celle des autres, a obtenu ces résultats. Mais il faut avoir le courage de se dire que, même si les formations sont bien conçues, les simulations optimisées, et que le bouton de signalement est en place, le résultat final reste décevant »*, nous expliquait Sébastion Baron aux dernières Assises de la Cybersécurité. Cette remise en question du marché de la formation en cybersécurité est un « sujet Voldemort » selon les mots de Sébastien Baron, *« un sujet que tout le monde connaît, mais que personnes n'osent aborder »*. Ce dernier précise toutefois que ces chiffres ne concernent que les clients Mimecast. D'autres études plus approfondies devront sans doute être réalisées pour juger de l'efficacité réelle des méthodes de sensibilisation actuelles.

Il existe un autre risque : celui de sursolliciter les collaborateurs, lesquels pourraient alors ne plus prêter attention aux messages et campagnes. Sans donner de fréquence précise, le CLUSIF invite à pondérer les campagnes de sensibilisation. *« Les entreprises peuvent modifier la façon dont elles organisent leur formation à la sécurité. Elles peuvent choisir de gamifier la formation et d'offrir des récompenses pour le respect des règles »*, suggère de son côté Richard de la Torre.

Les organisations connaissent-elles seulement toutes ces solutions disponibles ? 46 % des TPE-PME interrogées dans l'étude d'Opinion Way soulignent les nombreux obstacles à l'atteinte d'un bon niveau de maturité : le manque de temps (60 %), le manque de connaissances/expertise (56 %), le manque de budget (53 %), et la non-identification des interlocuteurs à qui s'adresser (34 %).

Une défense multiniveau

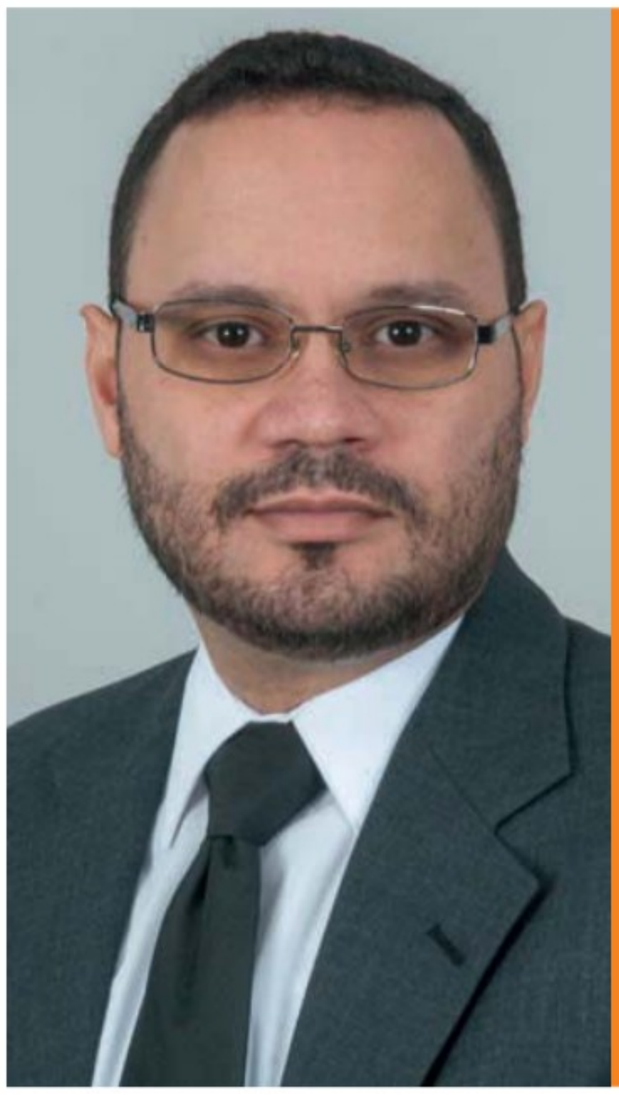
Bien que la maturité des collaborateurs soit cruciale, elle ne suffit pas, et même les employés les plus vigilants peuvent être trompés. Si des acteurs malveillants parviennent à franchir la première ligne de défense, une série d'outils de sécurité doit prendre le relais en arrière-plan. La protection des postes de travail s'appuie

Le temps de **la convergence**

EDR, EPP, MTD... tous convergent de plus en plus vers des solutions dites UES (Gestion unifiée des terminaux), qui combinent dans une interface de gestion unique les caractéristiques de chacun, parfois même avec des briques SOAR (Orchestration, automatisation et réponse). La XDR (Détection et réponse étendues) est, quant à elle, une technologie multiniveau, qui surveille tout le système d'information en supprimant les barrières entre les périmètres de détection. Il protège l'infrastructure informatique en collectant et corrélant des données provenant de divers points de sécurité, tels que les terminaux, les applications, les e-mails, le cloud et les réseaux. Cette approche est censée offrir une visibilité améliorée de l'environnement technologique d'une entreprise, permettant aux équipes de sécurité de détecter, d'enquêter et de répondre plus rapidement aux cybermenaces.

sur plusieurs piliers : l'authentification ; une gestion sécurisée des accès à distance ; la détection et la prévention des infections par des virus, vers et autres logiciels malveillants ; l'analyse comportementale pour repérer les activités suspectes ; le filtrage des URL ; l'application des politiques de sécurité de l'entreprise ; le chiffrement des données ; la segmentation du réseau pour limiter la portée des attaques ou contrôler le trafic selon des règles de sécurité spécifiques.

L'une des premières étapes consiste à appliquer les bases de la gestion des identités et des accès. Et là encore, on part de loin. Selon une enquête de NordPass, un utilisateur d'Internet possède en moyenne 168 mots de passe pour un usage personnel et 87 pour le travail. Les utilisateurs privilégient souvent la commodité à la sécurité lorsqu'ils créent leurs mots de passe. En France, par exemple, les trois mots de passe les plus utilisés sont «123456», «12345678» et «azerty». Si la complexité du mot de passe dépend en partie de la sensibilisation des collaborateurs, il incombe aussi aux organisations de mettre en place des mesures strictes de gestion des identités et des accès. Sans nécessairement aller jusqu'à l'authentification unique (Single Sign-On) ou aux passkeys, des actions simples peuvent faire une grande différence. Par exemple, *« s'assurer que les employés n'utilisent pas le même mot de passe que celui qu'ils utilisent pour le réseau d'entreprises et les plateformes de réseaux sociaux, ou utiliser, dans la mesure du possible, l'authentification à deux facteurs »*, développe Richard de la Torre. [L'authentification à deux facteurs, 2FA, à ne pas confondre avec l'authentification multifactorielle (MFA) qui consiste à utiliser plus de deux étapes de vérification pour accéder à un compte, ndlr.]



« Le plus grand défi pour les équipes de sécurité se situera toujours entre la chaise et le bureau. »

Richard de la Torre, Directeur Technique de Bitdefender Enterprise Solutions.

L'indispensable Zéro Trust ?

Eric Bohec souligne l'importance de penser à la sécurité des accès lorsque l'on travaille en mobilité : *« Quand je suis en dehors du bureau, comment puis-je accéder aux données de l'entreprise ? C'est là qu'interviennent des solutions ou des agents permettant de se connecter via une passerelle web contrôlée par l'entreprise. »* Dans ce contexte, le modèle Zéro Trust est souvent loué pour son efficacité. Il repose sur deux principes fondamentaux : ne jamais faire confiance et toujours vérifier. Chaque utilisateur, appareil et flux de données, est considéré par défaut comme potentiellement compromis et tout accès ou permissions exige, avant accord, une validation. Pour diminuer la confiance implicite accordée aux collaborateurs dans un périmètre classique, l'accès aux ressources doit être limité en fonction des besoins et en adoptant le principe du moindre privilège nécessaire à l'exécution d'une tâche. Les accès et les demandes doivent faire l'objet de réévaluations et de contrôles réguliers.

Dans un précédent entretien, John Kindervag, le théoricien du modèle Zéro Trust, affirmait que cette architecture était presque infaillible si correctement mise en place. *« Nous n'observons presque aucune cyberattaque réussie dans des environnements Zéro Trust bien conçus et dotés de politiques préventives. En deux ans et demi de gestion des services Zéro Trust, avant de rejoindre Illumio, nous n'avons enregistré qu'une seule petite attaque de ransomware, que nous avons rapidement détectée. Elle avait visé une acquisition qui avait refusé d'adopter le Zéro Trust »*, expliquait-il. Plus que la maturité des collaborateurs, c'est celle des entreprises qui est mise à l'épreuve avec le Zéro Trust. Bonne nouvelle, le modèle semble convaincre. Un rapport d'Okta intitulé *« L'état de la sécurité Zéro Trust 2023 »*, basé sur les réponses de 860 responsables de la sécurité de l'information dans le monde, révèle qu'en seulement deux ans, 61 % des organisations ont mis en place une initiative Zéro Trust, et 35 % envisagent de le faire prochainement.

Surveiller et protéger

Zero Trust ne se substitue pas aux autres cadres de sécurité et technologies. Les entreprises doivent également déployer des logiciels directement sur les postes de travail. L'implémentation d'une plateforme de protection des postes (EPP) utilisant une approche basée sur les signatures pour identifier et bloquer les menaces connues demeure un élément important pour se protéger des menaces anciennes ou identifiées. L'EPP intègre généralement la protection contre les logiciels malveillants traditionnels, la gestion des pare-feux locaux, du chiffrement des terminaux, le contrôle de l'intégrité des fichiers, le filtrage d'URL, ainsi que des capacités d'analyse comportementale, de prévention des intrusions et de gestion des vulnérabilités. Toutefois, à lui seul, il ne peut détecter toutes les menaces potentielles et émergentes dans un environnement cyber en constante évolution. Pour une action efficace en matière de prédiction, de

prévention, de détection et de réaction, les éditeurs recommandent d'opter pour des produits alliant EPP et EDR (détection et réponse sur les terminaux). Cette approche est de plus en plus plébiscitée. Selon des données de Gartner datant de 2021, 60 % des entreprises devraient remplacer leurs anciennes solutions de sécurité par celles combinant ces deux technologies. Autre enseignement, dans son baromètre 2024, réalisé auprès de ses membres, directeurs de la cybersécurité et responsables de la sécurité des systèmes d'information (RSSI), le Club des experts de la sécurité de l'information et du numérique (CESIN) révèle que l'EDR a progressé de 9 points, atteignant 90 % de solutions déployées.

Plus proactif, il est capable de détecter des menaces complexes ou inconnues. Il offre une visibilité en temps réel sur les activités des terminaux et une surveillance en continu. Il analyse le réseau pour détecter les comportements suspects et indicateurs de compromission. Lorsqu'une menace est identifiée, il génère des alertes, qualifie les événements et peut bloquer divers incidents (ransomwares, exploits de failles zero day, vol de données sensibles, etc.). Cependant, le nombre d'alertes et le taux de faux positifs pouvant être élevés, l'EPP effectue un premier filtrage des menaces, réduisant ainsi le volume des alertes à traiter.

Pour une protection optimale, les éditeurs ne manquent pas de vanter l'utilisation de l'intelligence artificielle (IA) qui vient renforcer les capacités de leurs produits. *« L'utilisation de l'IA dans les outils de sécurité devrait s'accélérer à l'avenir. Toutefois, les organisations doivent rester prudentes et ne pas se fier uniquement à ces outils, encore sujets à des hallucinations »*, prévient Richard de la Torre. Il incombe également aux entreprises de s'assurer des réelles capacités d'IA en prenant du recul par rapport aux discours marketing des éditeurs.

Au-delà des seules considérations techniques, l'efficacité de l'EDR dépend également du facteur humain. Déjà, ils exigent de se fier aux analystes chargés de gérer les alertes remontées. Or, dans le cadre d'un service managé, par exemple, son efficacité dépendra aussi en partie des compétences du Centre d'opérations de sécurité (SOC). En ce qui concerne le déploiement, c'est là encore l'humain qui a la main. *« Si l'on observe ce que vous achetez et ce que vous installez, il existe des écarts parfois considérables [...]. Il n'est pas rare de voir des entreprises dont le parc est équipé à seulement 60 % d'un EDR qui n'est pas toujours maîtrisé ou correctement configuré »*, déplore Baptiste David, ancien RSSI et responsable de la stratégie de marché chez Tenacy, un éditeur qui développe une plateforme SaaS de gestion de la cybersécurité. Certaines organisations choisissent, en outre, de restreindre les solutions EDR aux postes critiques (ordinateurs des dirigeants, postes administrateurs, serveurs) pour des raisons budgétaires et/ou manque de ressources IT.

Les mobiles, parent pauvre de la cybersécurité ?

Il ne s'agirait pas non plus d'oublier téléphones et tablettes. *« Le mobile est devenu un poste de travail à part entière dans certaines entreprises et nécessite des outils spécifiques, comme le MDM (Mobile device management) et le MTD (Mobile Threat Defense) »*, décrit Jean-Baptiste Guglielmine, ingénieur systèmes senior chez Cybereason. Le MDM sert à la gestion d'une flotte d'appareils mobiles qui fournissent aux administrateurs un contrôle externe afin de déployer des configurations, assurer la sécurité des données, surveiller l'utilisation des mobiles et appliquer des politiques de conformité. Le MTD protège, quant à lui, les mobiles contre les menaces qui leur sont spécifiques. Il utilise plusieurs technologies allant de l'identification d'applications malveillantes et de malwares à l'analyse comportementale, ainsi qu'à la protection contre le phishing, entre autres. Cette approche est devenue cruciale avec l'augmentation de l'utilisation des appareils mobiles dans le milieu professionnel, notamment dans le cadre de la tendance Byod, et dans un contexte d'augmentation des risques liés à ce type d'appareil. Un rapport de 2024 réalisé par Zimperium, entreprise spécialisée dans la sécurité mobile, estime que 82 % des sites de phishing ciblent les appareils mobiles des entreprises.

« Si les grandes entreprises en sont souvent dotées, beaucoup d'autres restent sous-équipées dans ce domaine, faisant de la sécurité mobile un "parent pauvre" dans leur organisation », estime Jean-Baptiste Guglielmine. Il reste que les forces pourraient bientôt être rééquilibrées. Le marché des MTD est, en effet, en pleine expansion et devrait atteindre 4,66 milliards de dollars en 2024, et 12,21 milliards de dollars d'ici 2029, selon des données de Horizon Market Strategies.

Pas simple de se retrouver dans cet éventail de possibilités, qui exige un travail de prospection important de la part des organisations, pas toutes matures en matière de cybersécurité. C'est pourquoi, il s'agit avant tout, selon Baptiste David, *« de bien comprendre son système d'information et d'identifier les besoins en sécurité »* afin d'adopter les solutions techniques et organisationnelles les plus adaptées. *« La deuxième étape consiste à les mettre correctement en œuvre pour atteindre le niveau de sécurité souhaité »* et s'assurer que les mesures prévues sont bien appliquées.

V.M

Gare à **la multiplication des agents**

Adrien Merveille, directeur technique de Check Point France, alerte de son côté sur la multiplication des agents sur les postes. *« Aujourd'hui, il y a beaucoup d'agents différents installés sur les postes de travail. Parfois, vous allez même séparer EPP et EDR. Vous ajoutez un client VPN pour accéder aux ressources internes, ce qui fait un troisième agent, puis un autre pour l'identité. Si on ajoute en plus un agent pour chiffrer le disque, et encore un autre, puis encore un autre, cela peut ralentir le poste. Et le pire qui puisse arriver quand on fait de la cybersécurité, c'est de dégrader l'expérience utilisateur. »* D'où l'importance de bien analyser ses besoins en amont, de déterminer ses priorités et de limiter le nombre de solutions et d'éditeurs pour réduire le nombre de consoles de gestion.

CrowdStrike dévoile des innovations IA et lance des services financiers

CrowdStrike a annoncé plusieurs innovations lors de sa convention Fal.Con 2024 (16 au 19 septembre à Las Vegas), visant à unifier la sécurité et les opérations IT avec sa plateforme Falcon. Parmi les nouveautés, on retrouve « Project Kestrel » pour simplifier la gestion des menaces et « CrowdStrike Signal » pour améliorer la détection grâce à l'IA. L'entreprise a également lancé « Charlotte AI » pour automatiser les réponses aux incidents, ainsi qu'une filiale de services financiers pour faciliter l'accès à ses technologies.

CrowdStrike a organisé sa convention Fal.Con du 16 au 19 septembre à Las Vegas, l'occasion pour l'entreprise de faire plusieurs grosses annonces autour d'innovations pour sa plateforme de cybersécurité Falcon. L'objectif est d'unifier la sécurité et les opérations informatiques pour mieux répondre aux menaces et simplifier la gestion de la sécurité. Ces innovations reposent sur une architecture cloud et IA-native, avec un agent unique qui consolide plusieurs outils de sécurité, permettant ainsi d'éliminer la complexité et d'améliorer les résultats en matière de sécurité.

Parmi les nouveautés, l'entreprise américaine met en avant plusieurs points comme « Project Kestrel », à savoir une nouvelle interface utilisateur qui unifie les données de sécurité pour éliminer la complexité, améliorer la collaboration et accélérer la réponse aux menaces. Il faut aussi citer « CrowdStrike Signal », un moteur d'IA qui regroupe les événements pour fournir des alertes prioritaires et améliorer la détection des menaces. Cela permet aux analystes de se concentrer sur les menaces les plus importantes, augmentant ainsi l'efficacité et réduisant le risque de détections manquées.

Protection de l'identité et assistant d'IA

L'éditeur a également introduit une amélioration de sécurité cloud, permettant une gestion de la sécurité de bout en bout à travers l'ensemble de l'infrastructure cloud. Sur ce point, une des innovations notables est la gestion de la posture de sécurité de l'IA (AI-SPM), qui surveille les services d'intelligence artificielle et les modèles de langage déployés dans le Cloud pour détecter les erreurs de configuration et les vulnérabilités. Le développeur a aussi axé ses efforts sur la protection des identités afin d'imposer des accès basés sur les risques avec une approche Just-in-Time (JIT). Cela réduit la surface d'attaque liée aux identités dans des environnements hybrides. Une protection en temps réel est aussi disponible pour Microsoft Entra ID pour stopper les vols de mots de passe et les tentatives de phishing.

© CrowdStrike

George Kurtz, PDG et fondateur de CrowdStrike, a expliqué que « les nouvelles innovations de la plateforme Falcon simplifient les opérations de sécurité et informatiques, permettant aux entreprises de prendre des décisions plus rapidement, de collaborer efficacement et d'adopter une approche proactive pour prévenir les violations de sécurité ».

Comme de nombreuses autres entreprises, CrowdStrike arrive avec son propre assistant d'IA, baptisé « Charlotte AI ». Cet agent intervient pour automatiser le tri des détections, accélérant ainsi les enquêtes et les réponses aux incidents. Formée avec l'expertise de l'équipe Falcon Complete, « Charlotte AI » permet aux entreprises de bénéficier de meilleures pratiques en matière de sécurité, en réduisant considérablement le temps nécessaire

à l'analyse et à la résolution des incidents. Enfin, la solution a intégré des fonctionnalités pour simplifier les opérations informatiques, en utilisant des charges de travail alimentées par l'IA. La plateforme Falcon peut désormais interroger des actifs en temps réel pour obtenir des informations détaillées sur les configurations informatiques et appliquer des correctifs ou résoudre des problèmes de conformité de manière automatisée. Selon George Kurtz, PDG et fondateur de CrowdStrike, *« les nouvelles innovations de la plateforme Falcon simplifient les opérations de sécurité et informatiques, permettant aux entreprises de prendre des décisions plus rapidement, de collaborer efficacement et d'adopter une approche proactive pour prévenir les violations de sécurité »*.

© CrowdStrike

« Nous entamons un nouveau chapitre ensemble : un chapitre fondé sur l'adoption de la plateforme Falcon, la résilience et le succès de l'écosystème », a expliqué Daniel Bernard, le directeur commercial de CrowdStrike durant Fal.Con 2024.

Introduction de CrowdStrike Financial Services

Lors de la convention, CrowdStrike a également annoncé la création de sa filiale CrowdStrike Financial Services, dont l'objectif est de fournir des solutions de financement sur mesure destinées à faciliter l'adoption de la plateforme CrowdStrike Falcon. Selon l'entreprise, cette initiative vise à simplifier l'accès à la technologie en éliminant les complexités liées à l'approvisionnement et à permettre aux clients de consolider leurs outils de sécurité plus rapidement.

CrowdStrike Financial Services s'inscrit dans une stratégie plus large visant à rendre la technologie de CrowdStrike plus accessible aux entreprises qui cherchent à se protéger contre des menaces de plus en plus sophistiquées. En complément de ce nouveau service, CrowdStrike avait déjà lancé Falcon Flex, un modèle de licence flexible qui permet aux clients d'accéder à l'ensemble du portefeuille de modules de la plateforme et d'utiliser ceux dont ils ont besoin. Désormais, avec cette nouvelle entité financière, les clients peuvent utiliser CrowdStrike Financial Services en combinaison avec Falcon Flex ou les modèles de licence standard. *« Nous souhaitons supprimer la complexité, non seulement au niveau technologique, mais également dans tous les aspects liés à la sécurité des entreprises, y compris l'acquisition de la technologie »*, a détaillé George Kurtz.

Selon lui, CrowdStrike Financial Services offre plusieurs avantages significatifs : un financement complet et simplifié qui permet aux clients de bénéficier de solutions de financement internes avec des conditions faciles à comprendre et d'avoir le choix entre une grande variété de modalités de paiement (mensuels, trimestriels, annuels), ainsi que des options de report ou d'échelonnement. On peut aussi noter que les utilisateurs auront accès à un service client dédié composé *« de professionnels expérimentés en financement »*.

Partenariat avec AWS et Nvidia

Enfin, parmi les dernières annonces faites durant Fal.Con, CrowdStrike a annoncé le lancement de la deuxième édition de son programme d'accélérateur de startups en cybersécurité, cette fois en partenariat avec AWS et Nvidia. Ce programme vise à soutenir des startups innovantes aux États-Unis ainsi que dans la zone EMEA en leur fournissant des ressources essentielles pour accélérer leur croissance dans le domaine de la cybersécurité. Ce programme, qui est ouvert aux candidatures du 16 septembre 2024 au 10 janvier 2025, offre une expérience immersive de huit semaines. Les startups sélectionnées bénéficieront d'un accompagnement gratuit qui inclut du mentorat d'experts de l'industrie, des conseils techniques, des opportunités de mise sur le marché, ainsi que des ressources financières. De plus, elles pourront profiter d'opportunités de réseautage avec des investisseurs internationaux en cybersécurité et de sessions d'accompagnement pour accélérer leur développement.

Nvidia rejoint le programme pour la première fois, renforçant l'accélérateur avec son expertise dans l'intelligence artificielle (IA) et le calcul accéléré. À travers son programme Inception, Nvidia offre un soutien aux startups en phase de développement, en leur fournissant des ressources comme des crédits pour des formations en deep learning, des tarifs préférentiels sur ses produits matériels et logiciels, et une assistance technique approfondie. À la fin du programme, les startups présenteront leurs innovations lors d'une journée de démonstration à San Francisco, durant l'AWS Startup Loft en avril 2025. Les meilleures présentations pourraient être éligibles à un financement du CrowdStrike Falcon Fund, un fonds d'investissement qui soutient des entreprises de sécurité innovantes et intégrées à la plateforme CrowdStrike Falcon. La première édition du programme a été un succès, avec des participants ayant levé plus de 150 millions de dollars auprès d'investisseurs de premier plan.

Michel Chotard

Le faux CFO de Hong-kong

Décryptage d'une deepfake ultra sophistiquée

Au début de l'année, une multinationale basée à Hong Kong a été victime d'une escroquerie digne d'un film de science-fiction. L'un de ses employés a été piégé par des escrocs utilisant des deepfakes pour se faire passer pour un important directeur financier ainsi que plusieurs autres collaborateurs lors d'une visioconférence. Cette fraude leur a permis de dérober pas moins de 25,6 millions de dollars.

Les attaques de type deepfake exploitent diverses méthodes et l'intelligence artificielle pour créer des images, des vidéos ou des fichiers audio falsifiés imitant parfaitement l'apparence ou la voix de personnes réelles dans le but d'extorquer des actifs financiers ou stratégiques. Ils peuvent notamment utiliser des encodeurs automatiques pour transférer des images et des mouvements d'une image à une autre. Grâce aux progrès fulgurants de l'IA, les attaquants peuvent créer facilement du contenu audio et vidéo hyperréaliste en temps réel. Communément appelées « Attaques au président », elles ne cessent de monter en sophistication à l'instar de l'attaque du *« Faux CFO de Hong Kong »*. Pour cette dernière, la victime (une employée du département financier de l'entreprise) a été ciblée par un email de phishing semblant provenir du directeur financier (CFO) basé au Royaume-Uni. Selon le rapport d'enquête, le salarié a tout d'abord soupçonné qu'il s'agissait d'un phishing, car celui-ci mentionnait la nécessité d'effectuer une transaction secrète. Malgré ses soupçons, il s'est ensuite fait piéger en rejoignant une visioconférence au cours de laquelle il a reconnu le CFO et plusieurs autres cadres de l'entreprise.

Une visioconférence entièrement truquée

En confiance, l'employé a rejoint sans le savoir une visioconférence dans laquelle tous les participants, en dehors de lui-même, étaient faux ! Les escrocs auraient téléchargé des vidéos à l'avance et manipulé ensuite les images et les voix à l'aide d'une intelligence artificielle sophistiquée pour leurrer la victime. Les fraudeurs ont réussi à lui extorquer 200 millions de dollars hongkongais (25,6 millions de dollars) répartis sur quinze transferts sur cinq comptes bancaires. L'arnaque n'a été découverte que plusieurs jours plus tard, lorsque l'employée a effectué des vérifications auprès du siège social de l'entreprise. L'enquête est toujours en cours… De plus en plus sophistiquées et difficiles à détecter, les deepfakes représentent une menace croissante pour les entreprises. **JC**

© ArcadEddie

Grâce à l'IA, la techologie deepfake permet de superposer facilement le visage d'une personnalité sur celui d'un inconnu.

« Les deepfakes sont passées à un tout autre niveau grâce aux progrès de l'intelligence artificielle »

Babar Rashid, Solutions Engineer chez BeyondTrust.

Pouvez-vous nous présenter BeyondTrust ?

BeyondTrust est une société américaine de cybersécurité spécialisée dans la gestion des comptes à privilèges (PAM pour Privileged Access Management en anglais). Notre rôle consiste à sécuriser les comptes à privilèges des entreprises, c'est-à-dire les comptes administrateurs, les comptes root, les comptes admin des bases de données, etc. Tous les comptes ayant des droits plus élevés que les comptes standards. On travaille avec plus de 20 000 clients dans le monde, dont plus de 60 % des entreprises grands comptes du CAC 40 en France.

Le braquage numérique du faux CFO de Hong Kong qui a eu lieu en début d'année est particulièrement impressionnant. Quels sont, selon vous, les facteurs qui ont permis à ce type d'attaques de gagner en sophistication et en efficacité ?

Parmi les différents facteurs, il y a tout d'abord l'intelligence artificielle. Les progrès assez énormes réalisés dans le domaine de l'IA ces dernières années permettent aux attaquants de créer des deepkfakes de plus en plus élaborées et réalistes. Lorsqu'on parle de deepfakes, on pense souvent aux vidéos truquées, mais il ne faut pas oublier que cela peut aussi être des trucages audio. Ces derniers ne sont peut-être pas aussi impressionnants que l'attaque de Hong Kong, mais on a vu pas mal d'exemples d'attaques réussies via de faux appels téléphoniques. La disponibilité croissante de données (photos, vidéos, données personnelles...) sur le web et les réseaux sociaux permet également aux fraudeurs de concevoir des deepfakes qui semblent de plus en plus réalistes. Il y a quelque part aussi une faille chez l'être humain qui a tendance à faire confiance à ce qu'il voit dans les vidéos diffusées sur Snapchat, TikTok ou autres. Pour finir, on constate également une augmentation des attaques ciblées avec les données personnelles disponibles sur les réseaux sociaux dont je parlais. Au lieu de lancer des attaques aléatoirement, les fraudeurs exploitent ces données pour réaliser des deepfakes personnalisées et ciblées.

Les deepfakes sont de plus en plus réalistes et difficiles à détecter. Est-ce qu'il existe des technologies ou stratégies pour identifier et contrer ces attaques durant des vidéoconférences à priori parfaitement légitimes ?

Certaines deepfakes peuvent être effectivement très difficiles à détecter. Cette entreprise anglaise qui a des bureaux à Hong Kong s'est fait attaquer par le biais d'une fausse visioconférence. Il s'agissait en réalité d'une vidéo enregistrée que les hackers ont lancée durant cette visioconférence. La victime a fait confiance à ce qu'elle voyait à l'écran... Tous les cadres de l'entreprise présents dans la vidéo étaient des simulations générées par la technologie deepfake. Invité en tant que simple participant à la fausse visioconférence, l'employé victime de la supercherie pensait exécuter les ordres de son directeur financier pour faire l'acquisition d'une entreprise. Il a effectué plusieurs transferts d'un montant total de 25,6 millions de dollars. Pour détecter ce type de deepfakes, l'approche technologique va consister à utiliser des outils basés sur l'IA. C'est l'IA qui permettra de bloquer ces menaces à l'avenir. Il existe différents outils en ligne comme Deepware, Sensity AI, ou encore FakeCatcher d'Intel qui offre un taux de détection de 96 % des menaces deepfakes. Même si l'IA permet de réaliser des deepfakes de plus en plus sophistiquées, certaines limites techniques offrent encore la possibilité de les détecter assez facilement. Les IA ont en effet du mal à reproduire les personnes de profil. Une chaîne de télévision avait par exemple diffusé une vidéo truquée du président de l'Ukraine, mais son profil peu ressemblant avait immédiatement soulevé de nombreux doutes sur cette dernière. De ce fait, le FBI conseille aux grandes entreprises qui font des entretiens d'embauche à distance de demander aux candidats de montrer leur profil afin de s'assurer qu'ils sont bien des personnes réelles.

Avec la montée en puissance des deepfakes visant les entreprises, comment voyez-vous l'évolution des techniques de cyberdéfense dans les prochaines années ? Y a-t-il des innovations prometteuses qui pourraient changer la donne ?

C'est l'IA qui permet de faire ce type d'attaques, et c'est aussi elle qui permettra de les bloquer. Plus on aura de données, plus il deviendra simple et facile de détecter et contrer ce type de menaces. Les algorithmes vont s'améliorer au fur et à mesure pour détecter les vidéos truquées. Il y a également de plus en plus d'entreprises qui sont en train de travailler sur une approche Zero Trust et l'authentification continue basée sur le contexte. On va faire ce qu'on appelle de la biométrie comportementale. Au lieu de vérifier juste les empreintes digitales, la reconnaissance faciale, etc., on va pouvoir également analyser des aspects comme la manière de taper sur le clavier, la vitesse de la souris, ou même la position de l'utilisateur durant une visioconférence. Ce sont des choses qui sont en cours de développement et qui nous permettront de détecter des anomalies que l'on n'est pas encore capable de détecter aujourd'hui. À terme, les sociétés de cybersécurité pourront partager ce type de données en temps réel, comme elles le font actuellement avec les signatures de virus et les malwares pour leurs solutions antivirus. Il existera sans doute quelque chose de similaire pour détecter et bloquer les deepfakes.

Propos recueillis par Jérôme Cartegini

www.linformaticien.com

ABONNEZ-VOUS À L'INFORMATICIEN

linformaticien.com/abonnement

MAGAZINE

Recevez chaque mois (10 numéros par an) le magazine «papier» et accédez également aux versions numériques.

1 AN FRANCE : 72 €
2 ANS FRANCE : 135 €
1 AN UE : 90 €
2 ANS UE : 171 €
1 AN HORS UE : 108 €
2 ANS HORS UE : 207 €

NUMÉRIQUE

Accédez chaque mois (10 numéros par an) à la version numérique du magazine et retrouvez également via votre compte en ligne les versions numériques des dernières publications.

1 AN : 49 €
2 ANS : 89 €

ÉTUDIANT / ÉCOLE

Abonnez vos étudiants avec une formule dédiée à 60 % du prix normal de l'abonnement sous forme de PDF (10 numéros par an).
Possibilité abonnements groupés en contactant le service abonnements du magazine à abonnements@linformaticien.com.

ABONNEMENT 1 AN : 43, 20 €

Le RGPD : pierre angulaire de la protection des données en Europe

Par Patrick Blum, Délégué Général de l'AFCDP.

Le Règlement Général sur la Protection des Données (RGPD), entré en vigueur en 2018 dans l'Union européenne, a marqué un tournant décisif dans la gestion des données personnelles. Ce cadre juridique impose aux entreprises et organisations des règles strictes en matière de collecte, de traitement et de stockage des données personnelles des citoyens européens.

Les principes clés du RGPD incluent le consentement explicite et éclairé des utilisateurs pour la collecte de leurs données, la limitation de la collecte aux données strictement nécessaires (principe de minimisation), le droit d'accès et de rectification des données par les individus, le droit à l'effacement des données (ou « droit à l'oubli »), le droit à la portabilité des données, l'obligation de notifier les violations de données dans les 72 heures et la désignation d'un délégué à la protection des données (DPO) pour certaines organisations.

Ces dispositions visent à redonner aux individus le contrôle sur leurs informations personnelles et à responsabiliser les entreprises dans leur utilisation des données.

Le RGPD accorde une attention particulière au transfert de données personnelles en dehors de l'Union européenne. Le principe de base est que les données des citoyens européens ne peuvent être transférées que vers des pays tiers offrant un niveau de protection « adéquat ».

Le cas des États-Unis est particulièrement complexe. Après l'invalidation du « Safe Harbor » en 2015, puis du « Privacy Shield » en 2020 par la Cour de justice de l'Union européenne, un nouveau cadre a été mis en place : le Data Privacy Framework (DPF). Adopté en 2023, ce nouveau dispositif vise à limiter l'accès des services de renseignement américains aux données des Européens, établit un mécanisme de recours indépendant pour les citoyens européens et renforce les obligations des entreprises américaines en matière de protection des données.

Malgré ces avancées, le DPF fait l'objet de critiques, notamment de la part d'associations de défense des libertés numériques qui estiment que les garanties offertes restent insuffisantes. Du fait des recours dont il fait l'objet, ce cadre a un avenir incertain qui crée une situation instable pour les organismes, très nombreux, dont les activités s'appuient sur des transferts de données vers les États-Unis.

Les données de santé : un cas particulier hautement sensible

Les données de santé bénéficient d'une protection renforcée dans le cadre du RGPD qui les classe dans une catégorie spéciale de données nécessitant des mesures de sécurité accrues. Leur collecte et leur traitement sont soumis à des conditions strictes, notamment le consentement explicite de la personne concernée (sauf exceptions légales), la limitation du traitement à des finalités spécifiques (recherche médicale, santé publique, etc.), et l'obligation de mettre en place des mesures de sécurité renforcées.

La réutilisation des données de santé, notamment à des fins de recherche, pose la question cruciale de l'anonymisation. Même anonymisées, ces données peuvent parfois permettre de réidentifier les individus, ce qui soulève des inquiétudes quant à leur utilisation. Les techniques d'anonymisation doivent donc être constamment améliorées pour répondre à ces défis, tout en préservant la valeur scientifique des données.

La crise sanitaire liée au COVID-19 a mis en lumière les enjeux liés à l'utilisation des données de santé à grande échelle. La mise en place d'applications de traçage des contacts a soulevé de vifs débats sur l'équilibre entre santé publique et respect de la vie privée.

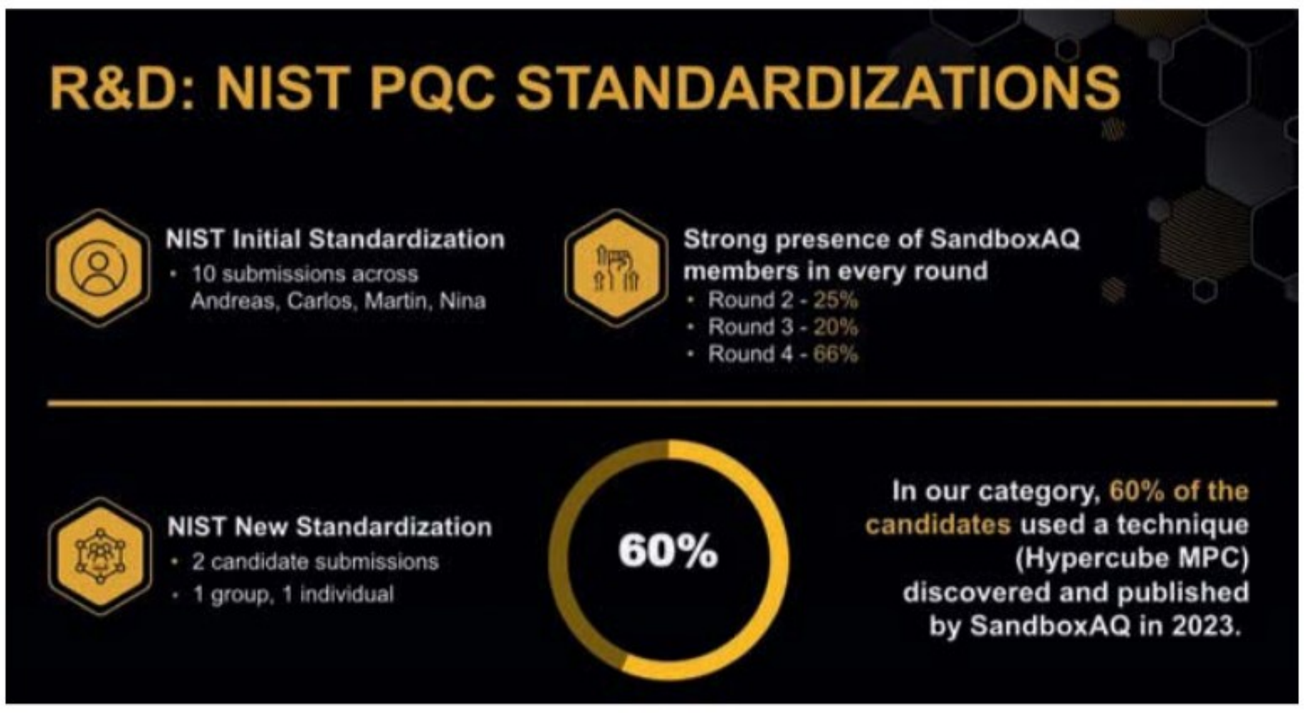
Ces outils ont montré à la fois le potentiel et les risques liés à l'utilisation massive de données de santé, et ont conduit à une réflexion approfondie sur les cadres éthiques et juridiques nécessaires pour encadrer de telles pratiques.

La réutilisation des données personnelles reste un sujet complexe et en constante évolution. Si le RGPD a posé un cadre légal solide en Europe, de nombreux défis persistent, notamment face aux avancées technologiques rapides et à la globalisation des échanges numériques.

La migration vers les nouveaux standards de cryptographie post-quantiques peut commencer

Avec la publication par le National Institute of Standards and Technology (NIST), des premiers standards d'algorithmes de chiffrement post-quantiques, les organisations américaines vont entamer leur migration et, pourquoi pas, entraîner le reste du monde avec elles. Une transition qui exigera un travail complexe d'inventaire et de priorisation des systèmes à faire migrer.

Face aux craintes que font peser les ordinateurs quantiques sur la cryptographie actuelle, le NIST a lancé, en 2016, un appel ouvert à propositions et soumissions d'algorithmes de chiffrement post-quantiques. Huit ans plus tard, après évaluation d'une quinzaine d'algorithmes prometteurs, l'institut a finalisé et publié un ensemble de standards de chiffrement conçus pour résister à une attaque d'ordinateur quantique : ML-KEM (ex-Crystals-Kyber), ML-DSA (ex-Crystals-Dilithium) et SLH-DSA (ex-Sphincs+). La norme pour un quatrième algorithme, baptisé FN-DSA (ex-Falcon), doit être publiée à la fin de l'année 2024.



© SandboxAQ

Des membres de SandboxAQ étaient présents à chaque tour de sélection des standards du NIST.

Des standards prêts au déploiement

Les algorithmes couvrent le chiffrement général, pour protéger les informations échangées sur un réseau public, ainsi que les signatures numériques pour authentifier des identités. ML-KEM repose sur un mécanisme d'encapsulation qui permet de générer une clé secrète partagée entre deux personnes sur un canal public. Ladite clé peut ensuite servir à réaliser des tâches de communication sécurisée, telles que le chiffrement et l'authentification, et sa sécurité repose sur un problème mathématique complexe. ML-DSA et SLH-DSA visent à créer et vérifier des signatures numériques, afin de garantir que les données n'ont pas été modifiées et que le signataire est bien celui qu'il prétend être. *« Les algorithmes annoncés aujourd'hui sont spécifiés dans les premières normes finalisées du projet de normalisation de la cryptographie post-quantique (PQC) du NIST et sont prêts à être utilisés immédiatement »*, explique l'institut dans son communiqué. L'agence enjoint donc les experts en cybersécurité à les intégrer dans leurs systèmes actuels et futurs. Cette publication des standards marque le début de la transition des systèmes vers les normes de cryptographie post-quantiques du NIST qui aura lieu entre 2025 et 2035, comme l'exige la feuille de route détaillée par le mémorandum de sécurité nationale des États-Unis, le *« NSM-10 and the Transition to Post-Quantum Cryptography »*.

Cette transition ne sera toutefois pas une mince affaire et exigera de répondre à un cahier des charges strict. *« Il y a un vrai sentiment d'urgence, mais il va d'abord falloir inventorier les systèmes, réaliser des tests, puis peu à peu tout faire migrer »*, détaille Carlos Aguilar Melchor, directeur scientifique chez SandboxAQ, une entreprise spécialisée dans divers domaines du quantique et qui a participé à la soumission de nombreux candidats au NIST et à la rédaction d'un des standards sortants.

Inventorier et identifier les systèmes vulnérables

Les organisations devront prendre des mesures afin de faciliter une transition en douceur. Mesures qui sont décrites dans un document du ministère de la Sécurité intérieure des États-Unis, repris

« Les attaques de types Harvest Now, Decrypt Later représentent un danger inacceptable, notamment pour les entreprises et les données sensibles, ou encore la propriété intellectuelle. D'où l'urgence de sécuriser le chiffrement des communications en transit. »

Carlos Aguilar Melchor,
Directeur scientifique de SandboxAQ.

en avril 2024 par le NSM-10. Ainsi, les organisations devront, si ce n'est pas déjà fait, identifier quelles données critiques pourraient être à risque et être déchiffrées lorsqu'un ordinateur quantique sera disponible. Elles devront ensuite mener un inventaire de tous les systèmes utilisant des technologies cryptographiques. Les responsables de la cybersécurité auront donc la charge d'identifier les normes de cybersécurité et de protection des données qui nécessiteront une mise à jour pour refléter les exigences de la cryptographie post-quantique. Ensuite, les systèmes devront être priorisés pour la transition en fonction de plusieurs critères, des fonctions, des objectifs et des besoins de l'entreprise. Enfin, après inventaire, un plan de transition des systèmes devra être élaboré.

Mais réaliser ces inventaires n'a rien d'une promenade de santé, car il n'existe pas de gestion centralisée de la cryptographie, qui *« est souvent implémentée de manière isolée par chaque développeur au sein de son application. En conséquence, elle se retrouve dispersée dans différents environnements, dans des codes binaires, des applications ou encore des systèmes d'exploitation »*, déplore Carlos Aguilar Melchor. Il est crucial de former dès maintenant les équipes de développement, afin qu'elles intègrent les bons algorithmes dès la conception. Ce qui implique de leur indiquer clairement quelles bibliothèques, versions et protocoles utiliser, et d'imposer certaines restrictions.

Automatisation des inventaires

Une fois ce travail mené à bien, *« il faudra commencer par sécuriser les communications, notamment via des protocoles comme TLS ou SSH, puis les systèmes de stockage, les infrastructures à clés publiques (PKI). Et enfin les logiciels utilisés par les entreprises, qui devront être mis à jour et testés pour s'assurer de leur conformité aux nouveaux standards »*, décrit Carlos Aguilar Melchor. D'ailleurs, selon lui, il y a urgence à protéger les communications chiffrées pour éviter des attaques de type *« Harvest Now, Decrypt Later »* (intercepter aujourd'hui, déchiffrer plus tard). Cette méthode consiste à intercepter maintenant des communications chiffrées, contenant de potentielles données sensibles, afin de les déchiffrer plus tard, lorsqu'un ordinateur quantique sera disponible.

Même si certains doutent de l'arrivée imminente des ordinateurs quantiques, il n'y a plus de débat : il faudra migrer. La question est de savoir si ces standards seront résistants à ces ordinateurs. Pour l'heure, il est impossible de le garantir à 100 %. Pourtant, les principes qui sous-tendent les ordinateurs quantiques sont connus depuis les années 1990 et, jusqu'à présent, *« aucune avancée majeure n'a été réalisée dans la résolution des problèmes mathématiques complexes qui servent de base à ces nouveaux algorithmes, ce qui confère une certaine confiance, mais pas de certitude absolue »*, nuance Carlos Aguilar Melchor. La prudence étant mère de sûreté, le NSM-10 prévoit que les plans de transition incluent *« la création d'une agilité cryptographique pour faciliter les ajustements futurs et permettre une flexibilité en cas de changements imprévus »*.

Les États-Unis entraîneront-ils les autres ?

Quid de la transition ailleurs qu'aux États-Unis ? Carlos Aguilar Melchor croit à l'effet domino. *« Si le gouvernement américain et les entreprises qui travaillent avec lui modifient leur cryptographie, alors tout le monde devra adopter ces nouveaux standards pour pouvoir interagir avec eux. »*

L'Union européenne, elle, tâtonne, encore. *« Il y a actuellement des discussions sur la manière de légiférer cette transition et nous avons été sollicités pour y contribuer. Mais je ne connais pas encore précisément la forme que prendra la réglementation »*, précise Carlos Aguilar Melchor. En France, jusqu'à présent, l'ANSSI a plutôt adopté une posture d'accompagnement en fournissant des recommandations sur la manière d'agir, sans imposer de directives strictes. Dans son avis révisé en janvier 2024 sur la migration vers la cryptographie post-quantique, elle *« encourage toutes les industries à inclure la menace quantique dans leur analyse de risque et à envisager des mesures de protection quantique dans les produits cryptographiques concernés »*. Cela contraste avec la méthode américaine, qui a rapidement adopté une approche plus directive, en publiant des documents détaillant explicitement les procédures à suivre. ▮

V.M

L'ANSSI défend l'hybridation

Dans son avis de janvier 2024 sur la migration vers la cryptographie post-quantique, l'ANSSI défend l'hybridation. Celle-ci consiste à combiner des algorithmes cryptographiques post-quantiques avec des algorithmes cryptographiques classiques, afin d'offrir une défense supplémentaire en cas de défaillance des premiers. *« Même si les algorithmes post-quantiques ont fait l'objet d'une grande attention, ils ne sont pas encore suffisamment matures pour garantir à eux seuls la sécurité. Par exemple, plusieurs algorithmes post-quantiques ont subi des attaques classiques au cours des dernières années. »*

Les attaques ciblant les pilotes Windows vulnérables en forte augmentation

Le spécialiste en cybersécurité, Kaspersky, tire la sonnette d'alarme sur l'augmentation significative des cyberattaques exploitant les vulnérabilités des pilotes Windows. Connue sous le nom de BYOVD (Bring Your Own Vulnerable Driver), cette méthode permet aux attaquants de désactiver les solutions de sécurité des systèmes informatiques pour mener différentes activités malveillantes.

Au deuxième trimestre 2024, les experts de Kaspersky ont signalé une augmentation de 23% des attaques exploitant des pilotes Windows vulnérables par rapport au premier trimestre. Des groupes de cybercriminels, tels que BlackByte ou AvosLocker, utilisent de plus en plus cette méthode pour désactiver les outils de détection et de réponse des systèmes, tout en élevant leurs privilèges. Cela leur permet de mener des attaques sophistiquées, comme le déploiement de ransomwares ou le renforcement de leur persistance à des fins d'espionnage ou de sabotage.

© Alexxsun

Les experts du GReAT (Global Research & Analysis Team) de Kaspersky font une veille constante pour rechercher des drivers Windows vulnérables.

Parallèlement, les experts ont constaté une recrudescence des outils permettant d'exploiter ces pilotes vulnérables. *« Bien que rien n'empêche réellement les acteurs de menaces de développer leurs propres outils privés, ceux disponibles publiquement éliminent le besoin de compétences spécifiques requises pour rechercher et exploiter les pilotes vulnérables. Rien qu'en 2023, nous avons identifié environ 16 nouveaux outils de ce type, soit une augmentation considérable par rapport aux quelques rares outils recensés les années précédentes. Compte tenu de cette augmentation, il est fortement recommandé de mettre en place des mesures de protection éprouvées »*, explique Vladimir Kuskov, responsable de la recherche anti-malware chez Kaspersky. Lutter contre ce type d'attaques capable de contourner les solutions de sécurité traditionnelles représente désormais un défi majeur pour les entreprises. ▮

« Il y a des groupes de ransomwares comme BlackByte qui utilisent fréquemment le BYOVD dans leurs techniques »

Robin Kwiatkowski, chercheur du GReAT de Kaspersky.

Robin Kwiatkowski, chercheur du GReAT de Kaspersky, est en première ligne pour traquer les groupes de cyberattaquants et documenter sur leurs techniques et leurs outils. Il nous explique comment la technique BYOVD permet à des hackers d'exploiter des pilotes de périphériques vulnérables pour contourner les protections de sécurité et ouvrir une porte d'entrée discrète dans les systèmes ciblés.

Pouvez-vous expliquer en quoi consiste exactement la méthode BYOVD ?

Le terme BYOVD (Bring Your Own Vulnerable Driver) est inspiré de l'expression courante BYOB (Bring Your Own Beer), utilisée lors de soirées où chaque participant apporte sa propre bière. Dans le contexte de la cybersécurité, cette technique tire son nom du fait qu'un attaquant, après avoir compromis une machine, cherche à obtenir des privilèges supplémentaires en identifiant des logiciels ou pilotes vulnérables déjà installés sur le système. Si aucun pilote vulnérable n'est présent, l'attaquant

amène alors son propre pilote vulnérable et l'installe pour obtenir ces privilèges. Il peut utiliser par exemple un pilote signé, comme celui d'une carte graphique, connu pour contenir une vulnérabilité. Étant donné qu'il est signé, ce pilote n'est pas bloqué par les antivirus. Ces derniers ne bloquent généralement pas les logiciels légitimes afin d'éviter d'altérer le bon fonctionnement du système. Cependant, une fois qu'un pilote spécifique est identifié comme vulnérable, il est possible de définir des règles plus précises pour surveiller son comportement. Si le pilote adopte un comportement suspect, cela peut indiquer qu'il est utilisé à des fins malveillantes. Cela nécessite toutefois une analyse préalable pour identifier la vulnérabilité et son exploitation.

Comment les attaquants se procurent-ils ces drivers ?

Ils effectuent essentiellement une recherche de vulnérabilités de la même manière que pour les autres logiciels. Ils ciblent des pilotes qui s'exécutent avec des privilèges élevés, puis cherchent des vulnérabilités. Cela peut inclure des failles de type buffer overflow, mais aussi une mauvaise gestion des contrôles du pilote permettant à un utilisateur d'accéder à des fonctionnalités auxquelles il ne devrait pas avoir droit. Concrètement, presque toutes les classes de vulnérabilités courantes peuvent être exploitées pour compromettre le pilote.

Robin Kwiatkowski, chercheur du GReAT de Kaspersky.

Est-ce facile pour les hackers de mener ce type d'attaques ?

Il existe plusieurs réponses à cette question. Tout d'abord, il devient de plus en plus crucial pour les attaquants de s'octroyer des privilèges élevés lorsqu'ils compromettent une machine, afin de désactiver les solutions de sécurité comme les antivirus et les outils d'EDR (Endpoint Detection and Response), qui sont de plus en plus répandus. Ils sont donc souvent contraints d'utiliser ce type de technique, bien que cela ne soit pas toujours indispensable. C'est cependant un moyen efficace pour contourner les logiciels de sécurité. De plus, ces techniques ont déjà été exploitées par des groupes APT (Advanced Persistent Threat). Une fois documentées, elles incitent d'autres attaquants à chercher de nouvelles vulnérabilités pour exploiter des méthodes similaires.

Comment les attaquants exploitent-ils les vulnérabilités pour compromettre les systèmes ?

Dans une enquête que l'on avait menée l'année dernière, un groupe avait développé un logiciel exploitant un pilote appelé « RTCore64 ». Ce pilote, intégré dans un exécutable, pouvait être chargé en mémoire, car il était signé numériquement. Une fois installée, une fonction du pilote permettait à l'attaquant d'écrire arbitrairement en mémoire, ce qui lui permettait de désactiver l'antivirus. Des groupes de ransomware, comme BlackByte, utilisent fréquemment la technique du BYOVD. Ils infiltrent un réseau, exploitent un pilote vulnérable pour obtenir des privilèges élevés, puis injectent un ransomware. Cette méthode fait désormais partie de l'arsenal des cybercriminels.

Quels conseils pratiques donneriez-vous aux entreprises pour se protéger contre les attaques basées sur les pilotes Windows vulnérables ?

Ce type de problème est relativement complexe, mais il est essentiel de limiter au maximum l'exécution de logiciels non validés au préalable par l'équipe IT. Idéalement, aucun logiciel ne devrait être exécuté sans autorisation, et il est recommandé de mettre en place des listes blanches pour contrôler ce qui peut s'exécuter. Cette approche peut grandement contribuer à réduire ce type de menace. Par ailleurs, une veille régulière sur les renseignements en matière de menaces (threat intelligence) est cruciale : dès qu'un composant est identifié comme étant utilisé de manière malveillante, il doit être spécifiquement désactivé ou interdit. Si un pilote est signalé comme vulnérable et potentiellement exploité, il est impératif de l'interdire sur le réseau.

Comment les solutions de sécurité doivent-elles évoluer pour lutter contre cette menace ?

Les spécialistes en cybersécurité se concentrent principalement sur la recherche proactive des composants et pilotes exploités par les attaquants. Cela permet parfois d'identifier un pilote vulnérable et de le bloquer avant qu'il ne soit utilisé de manière malveillante. Par ailleurs, l'objectif est aussi d'affiner les règles des solutions de sécurité pour détecter si le comportement d'un pilote est normal ou anormal, mais cela reste un processus complexe.

« Gérer une cyber-crise fait progresser toute l'organisation. »

Olivier Caleff, RSSI du groupe Erium, société française de conseils et expertise sécurité, membre du conseil d'administration du FIRST (Forum of Incident Response and Security Teams).

Olivier Caleff est très actif au CESIN (Club des Experts de la Sécurité de l'Information et du Numérique) et professeur à l'EGE (École de Guerre Économique).
Après avoir créé de nombreux centres de réponse à incidents IT (CERT/CSIRTs), il partage son expertise en gestion de crises cyber et aide les organisations régionales à gagner en résilience.

Quelle est la raison d'être de l'incubateur des CSIRTs que vous avez créé ?
Olivier Caleff : L'objectif est de mettre le pied à l'étrier d'équipes émergentes au sein de nouveaux centres de réponse aux incidents cyber, dans les entreprises ou au cœur des CSIRTs crées dans de la plupart des régions de l'Hexagone. Divers profils d'informaticiens peuvent construire ensemble une équipe tournée vers la résilience, pour fournir de la visibilité aux décideurs, avec un outillage technique, un référentiel juridique et des renseignements sur les menaces. Grâce à cet incubateur, créé avec Haude Costa et Vincent Nguyen, nous œuvrons pour que ces équipes coordonnent leurs activités et s'organisent autour d'un plan d'action pour gagner en maturité. Pour cela, nous nous appuyons sur deux outils reconnus : la RFC-2350 de l'IETF (Expectations for Computer Security Incident Response), et le modèle SIM3 (Security Incident Management Maturity Model).

Rencontrez-vous toujours des informaticiens hermétiques aux enjeux de cybersécurité ?
Olivier Caleff : Ils sont encore trop nombreux… mais il faut comprendre pourquoi. Les mentalités évoluent depuis le périple du référentiel BS7799, propagé en Europe depuis le Royaume-Uni, puis en France. La communication passe mieux autour des victimes d'attaques, mais l'anticipation ne fait pas encore partie de la culture de tous. Les grands groupes y sont plus sensibles que les PME.
Les organisations françaises ont dû se préparer aux menaces ciblant les J.O. de Paris 2024. Elles ont mené des exercices concrets et ont amélioré leur résilience grâce aux efforts de l'ARS, du CERT Santé, et de l'ANSSI. Le plan blanc s'est imposé aux établissements de santé, dès 2004, après un été meurtrier. Il est bien intégré à la culture de ces professionnels. La gestion de crise progresse dans d'autres secteurs, via les expériences partagées à l'InterCERT France et au CESIN notamment.

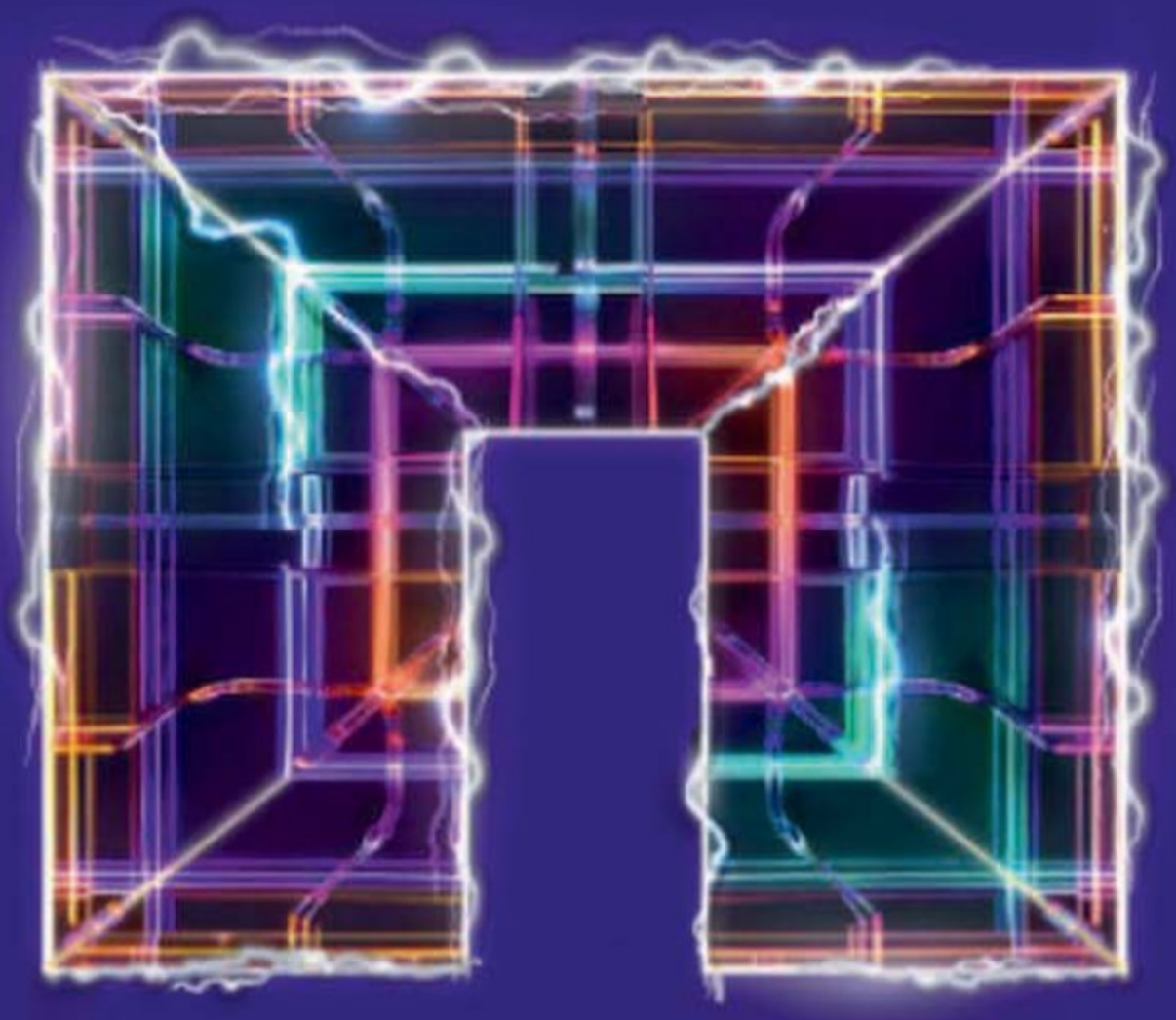
En matière de cyber-résilience, que faut-il anticiper en particulier ?
Olivier Caleff : Avant de reconstruire son parc IT, il faut penser à la survie de l'entreprise, avec très peu ou sans informatique car, quand l'infrastructure est tombée, rebâtir 100 ou 1 000 PC prendra du temps. Faute d'anticipation, se relever sera laborieux. Définir les priorités avec les métiers, en amont de la crise, est essentiel. Ce sont les métiers qui guident le séquencement du plan de redémarrage.
Disposer d'un annuaire de crise, d'inventaires et de cartographie est primordial. Il faut préparer la logistique, les RH, et la communication (interne, externe, technique, et de secours). Un bilan doit être fait, avec des leçons tirées pour ne pas répéter les erreurs du passé. Après la crise, on doit identifier les pierres d'achoppement, les procédures manquantes, les éléments qui ont fait défaut, et souligner ce qui a bien fonctionné. L'ensemble aide à construire un plan d'amélioration et de renforcement. Enfin, l'aspect financier doit soutenir la mise en œuvre de ce plan.

Quels traits de caractère doivent réunir les RSSI dorénavant ?
Olivier Caleff : La vie du RSSI n'est pas rose. Le stress fait partie de son quotidien. Nous devons traiter les événements avec lucidité, malgré la panique ou la sidération des victimes. Comme le font les services d'urgence du SAMU, nous devons appliquer des procédures et suivre des méthodologies (le protocole 6C), rassurer des professionnels en panique, surpris par l'ampleur et les dégâts provoqués par une crise cyber. Nous devons faire preuve de beaucoup de sang-froid et d'empathie.

Propos recueillis par Olivier Bouzereau

NetApp

INTELLIGENCE ARTIFICIELLE
CYBERSECURITÉ
CLOUD HYBRIDE
STOCKAGE UNIFIÉ



Les leaders de l'Infrastructure Intelligente de Données ont rendez-vous à :

NetApp INSIGHT Xtra Paris
10 décembre 2024, Cinémathèque de Paris

Inscrivez-vous sur insight.netapp.com/xtra

SCAN ME

LE NOUVEAU RENDEZ-VOUS
DES DÉCIDEURS IT

STOR'AGE

STOCKER, ARCHIVER,
SAUVEGARDER

5 NOVEMBRE 2024

ÉTOILE SAINT-HONORÉ, PARIS 8E

PURESTORAGE®

Hewlett Packard Enterprise

DELL Technologies

NUTANIX

HYCU®

NASUNI

NetApp

COMMUNICATIONS

L'INFORMATICIEN

DÉCIDEURS innovation

B SMART Change

storage-forum.com