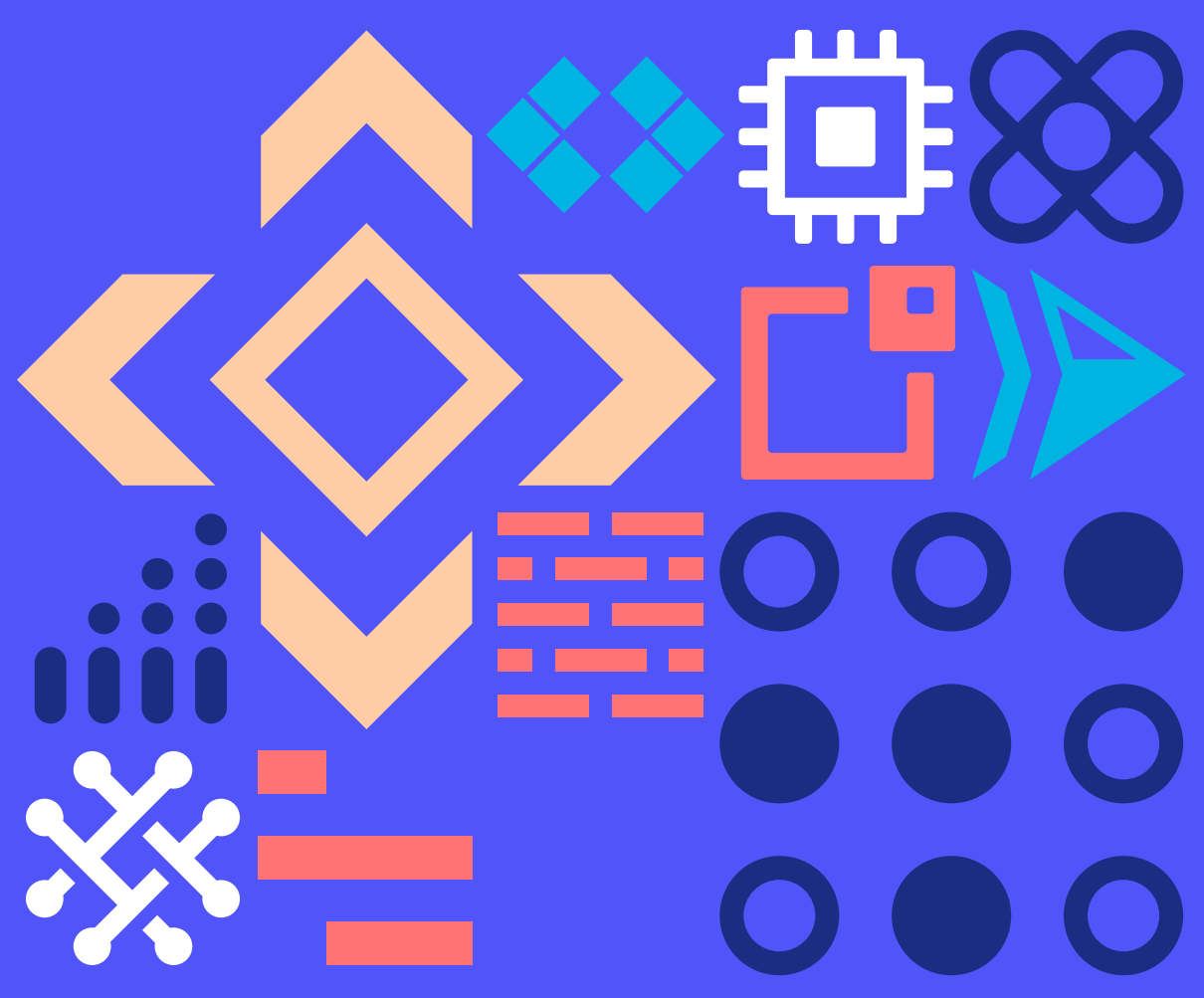




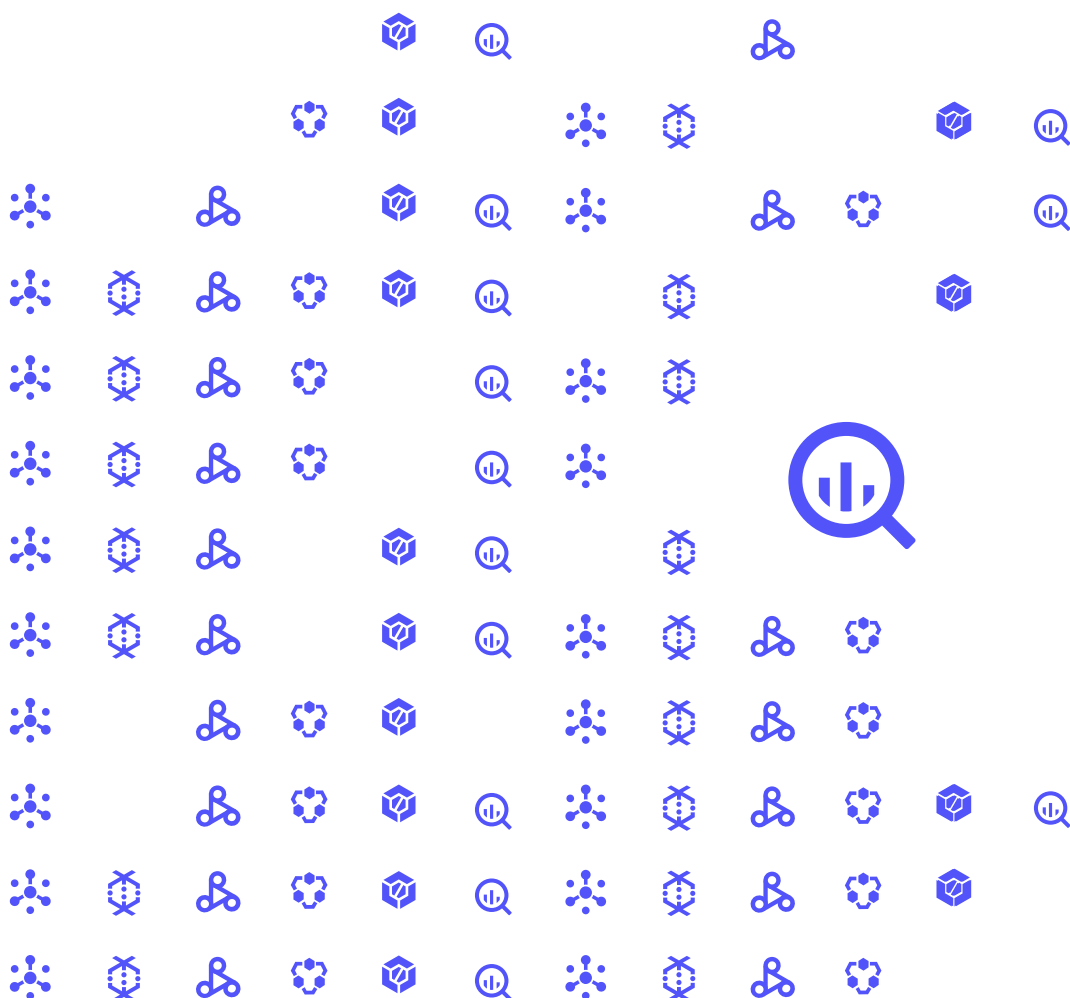
Google Cloud

Comment bien débuter sur Google Cloud ?

Nos conseils et avis d'experts pour s'y retrouver

[sf≡ir]

Sommaire



3 Introduction

5 3 questions à...

7 Préambule

8 Cloud & infrastructure

9 3 questions à...

11 Tendances infrastructure

23 Cloud & Data

24 3 questions à...

26 Tendances data

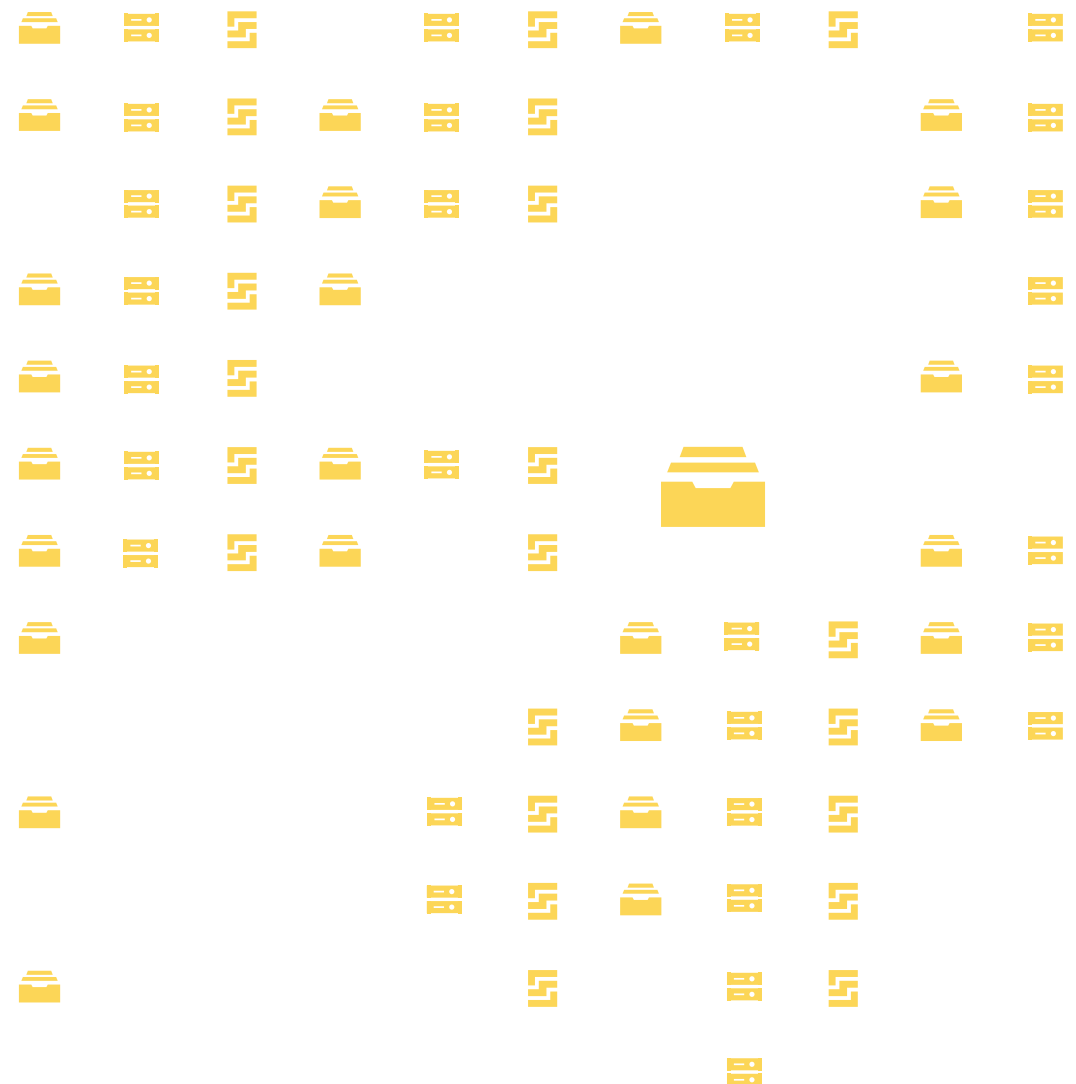
38 Méthodologie / Accompagnement

39 À vos côtés pour défendre le futur

45 Conclusion

50 Bibliographie

Introduction



Mettre à profit la puissance du cloud comme un catalyseur de l'intelligence et de l'efficacité des équipes.

Fonder toutes les stratégies sur la donnée pour mieux servir les clients. Dans un contexte de transformation permanente, la clé de la réussite tient en un mot : flexibilité.

Une flexibilité qui constitue le cœur d'une promesse appelée cloud.

3 questions à...



Anthony Cirot, Managing Director France Google Cloud, a accepté de partager sa vision de l'enjeu.

Chez SFEIR, nous incitons et aidons les entreprises à se mettre en mouvement. Cela implique des changements organisationnels et technologiques ; il faut devenir plus agile, être en mesure de s'adapter et d'innover en permanence. Quel est pour vous l'apport du cloud dans cette perspective ?

Le cloud est un accélérateur de la transformation numérique, on a tous vu son impact positif dès le début de cette crise sanitaire. Le cloud a ouvert le champ des possibles pour s'adapter à ce nouveau monde et bénéficier de nouvelles applications. Nos clients ont pu créer ex nihilo de nouveaux services comme le click & collect en quelques semaines voire en quelques jours, sans avoir à installer de serveurs.

Mais au-delà de cet aspect du cloud, ce qui est intéressant, c'est toute la vague axée sur la valorisation de la donnée et l'intelligence artificielle. Tout le monde veut en tirer profit, pour mieux connaître et servir ses clients, prédire leurs besoins, ou bien pour optimiser ses processus, sa chaîne logistique... Beaucoup de modèles existants ont été bouleversés par la crise sanitaire,

et les changements perdurent. Les entreprises doivent s'adapter et répondre à l'élasticité de la demande.

La nécessité de travailler à distance a aussi mis en lumière le besoin d'un mode de travail hybride, favorisant la collaboration entre collègues sur site et à distance, en équipes distribuées. Ce sont les solutions cloud qui fournissent cet environnement le plus immersif possible pour que les gens puissent collaborer au sein d'un document sans que les personnes qui travaillent à distance se sentent exclues.

Les collaborateurs doivent sentir qu'ils font partie d'un même espace de travail, sécurisé. Un cloud de confiance - indépendamment des lois - où les données sont protégées, et dont la localisation ou les mécanismes de manipulation sont expliqués, transparents.

Le marché du cloud public est aujourd'hui énorme, porté par de très nombreux acteurs. Qu'est-ce qui distingue aujourd'hui l'offre de Google Cloud sur le marché ?

Tout d'abord, je répondrais l'innovation, c'est l'ADN même de Google. Nous

avons réalisé des avancées technologiques majeures sur l'IA, l'infrastructure, la sécurité... Et nous mettons ces innovations à la disposition des entreprises, en contribuant fortement à la communauté Open Source : par exemple Kubernetes, Chromium, Open Telemetry...

Ensuite, nous adoptons systématiquement une approche de partenariat avec nos clients. Notre ambition n'est pas de vendre du serveur ou du stockage, mais bien d'accompagner nos clients dans leur transformation. C'est le sens des partenariats que nous avons signés avec LVMH, Orange, Carrefour ou Casino ; il s'agit de transformer le business model des entreprises. Le partenariat avec Fnac-Darty, que nous venons d'annoncer, relève quasiment du co-développement : nous redéfinissons ensemble la façon de chercher un produit sur un site de vente en ligne. Pour chaque client, petit ou gros, nous cherchons comment faire la différence en puisant dans tout ce que Google peut apporter. L'objectif est toujours d'appliquer le mode de pensée 10X : comment faire dix fois plus que ce qu'on faisait avant.

Le troisième grand différenciateur que je vois touche à la responsabilité sociétale. Google a atteint la neutralité carbone dès 2007, notre énergie est à 100 % compensée en énergies renouvelables et nous visons le zéro carbone en 2030 partout et à tout moment. Cet engagement, nous le mettons aussi à disposition des capacités applicatives des clients, au travers des infrastructures, mais aussi au-delà, avec des conseils pour développer des applications frugales.

Comment décririez-vous le positionnement et la stratégie de Google Cloud en France ?

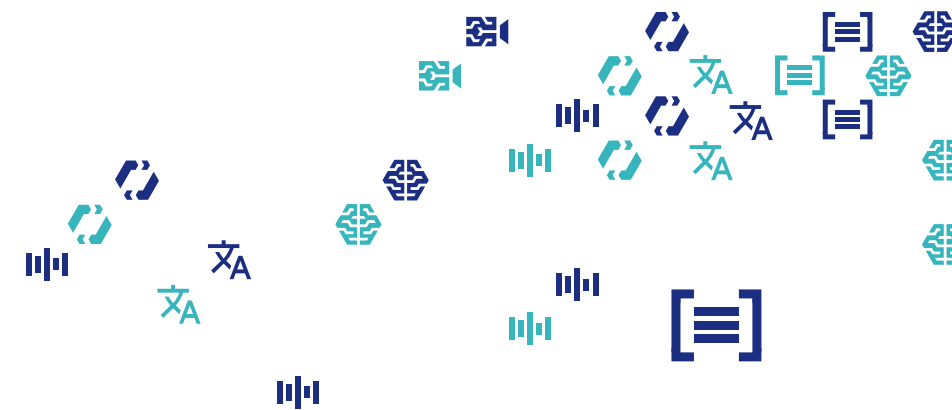
Le point le plus important à nos yeux est notre écosystème de partenaires. Notre ambition est d'aider les entreprises à se transformer et innover. Au vu de la demande du marché, nous avons besoin de nos partenaires pour tenir cette promesse et aider nos clients à bénéficier de l'accélération que peut apporter un fournisseur de plateformes comme Google Cloud. Cette focalisation sur les partenaires est valable pour Google Cloud dans le monde entier, mais je dirais qu'elle est amplifiée en France. L'autre point essentiel, c'est notre volonté de développer localement les infrastructures mondiales. L'ouverture de la région France en 2022 est importante pour soutenir l'activité du pays et pour répondre à des enjeux de souveraineté numérique. Nous irons même plus loin avec la création d'une société commune avec Thalès, qui aura pour but de proposer un cloud labellisé SecNumCloud piloté par une entité française. Ce cloud sera mis à disposition de l'écosystème, comme une région France mais souveraine.

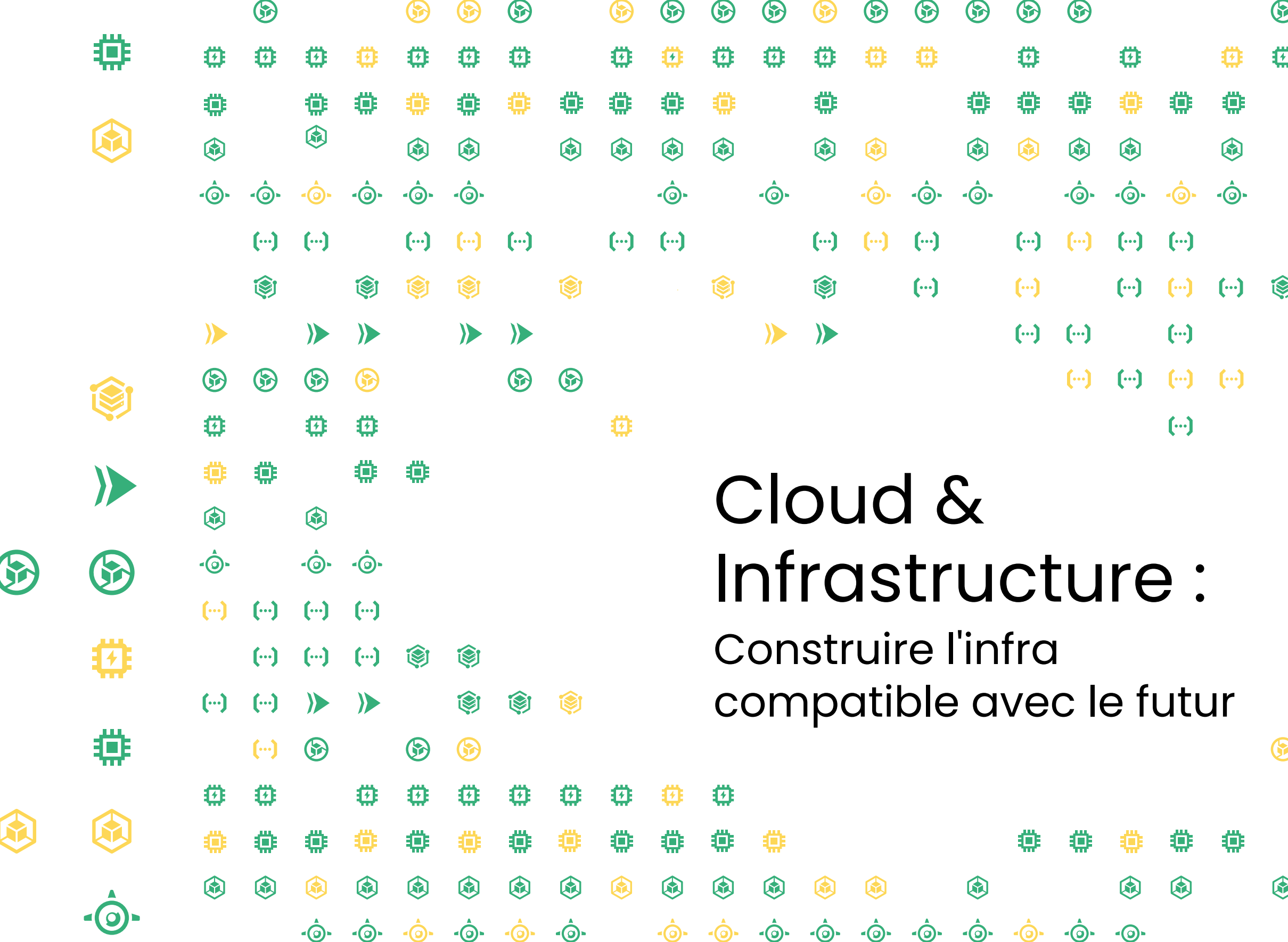
Le dernier axe fort est notre focus sur l'univers des start-up : nous sommes à Station F pour leur amener le meilleur de Google Cloud. Notre équipe dédiée a doublé de taille, c'est maintenant le département le plus conséquent dans Google Cloud en France.

Préambule

Pour passer de la vision à la réalité, il faut en passer par les actes. La Google Cloud Platform, associée au savoir-faire et à l'accompagnement des experts de SFEIR, constitue une opportunité majeure pour votre entreprise. Entre analyses, décryptages et explications, nous avons réuni dans ce livre blanc, l'essentiel de ce que vous devez savoir pour amorcer ou accélérer votre transformation cloud.

Bonne lecture.





Cloud & Infrastructure :

Construire l'infra compatible avec le futur

3 questions à...



Franck Zerbib, Head of Google Cloud Customer Engineering, France nous livre sa vision.

Les fondations d'une infrastructure cloud peuvent reposer sur des machines virtuelles, des containers, du serverless... Quels choix devrait-on faire pour s'approcher le plus possible du "cloud natif" et bénéficier ainsi de tous les apports du cloud ?

Toutes ces technologies forment un continuum de solutions tout à fait adaptées au cloud natif. Pour moi, ce qu'on appelle cloud natif repose sur 4 grands piliers :

- Les microservices et la containerisation des applications, pour gagner en agilité, répondre plus rapidement aux enjeux métiers, monter en charge mais aussi répondre à des problématiques d'hybridation et de portabilité.
- Les services managés, pour gagner du temps sur les opérations (scalabilité, sécurité...) et se concentrer sur la valeur business créée.
- Les API et l'utilisation de standards ouverts, pour faciliter la transformation numérique de l'entreprise et son intégration au sein d'un écosystème.

→ Les pratiques DevOps et SRE (site reliability engineering), indispensables pour disposer de la vélocité attendue aujourd'hui et pour assurer le déploiement et la disponibilité des applications.

Le choix ensuite de la solution d'infrastructure dépend de plusieurs facteurs. Le serverless est le choix de prédilection pour la vélocité, le time-to-market, mais on peut opter pour d'autres solutions en fonction des contraintes liées au langage, au matériel, etc. On peut très bien exploiter les mécanismes du cloud natif avec des VM (en recourant par exemple au load balancing intégré) ou des containers (avec un orchestrateur managé, de la supervision...).

Les offres "compute" et "storage" sont-elles amenées à devenir des "commodités" dans le monde du cloud, ou bien existe-t-il des différences essentielles à prendre en compte ?

C'est vrai qu'on va plutôt chercher la valeur dans les problématiques métier, et qu'on a tendance à considérer

l'infrastructure comme une commodité, identique d'un fournisseur à l'autre. Mais il faut faire attention aux détails ! Au quotidien, dans la manipulation des infras IT, ces petites différences peuvent s'avérer cruciales. Par exemple, la rapidité de démarrage d'une VM ou d'un cluster Kubernetes. Ou la granularité très fine offerte pour la configuration des machines, avec un ajustement à la hausse ou à la baisse au plus près des besoins des workloads, qui a un impact non négligeable sur le TCO.

Il faut aussi faire attention aux outils de management, à la façon dont les environnements seront gérés de façon à assurer la continuité de service des applications. Nous proposons aussi le chiffrement par défaut des données "at rest" et nous amenons de l'intelligence au travers d'outils de recommandation, par exemple pour conseiller de redimensionner à la hausse ou à la baisse les VM en fonction de leur consommation.

Enfin il faut prendre en compte le réseau : notre VPC global permet à deux VM situées à deux bouts de la Terre de dialoguer très simplement.

Les infrastructures hybrides, multi-cloud et cloud/on-premises, se multiplient. Comment peut-on gérer cela au mieux au sein de Google Cloud ?

L'hybridation est tout à fait naturelle, et les clients veulent donc s'assurer des aspects interopérabilité et réversibilité ; c'est un sujet que nous comprenons et adressons. Certains clients ne souhaitent pas tout de suite transformer leur paysage informatique, auquel cas nous pouvons faire en sorte d'interconnecter leurs systèmes (des machines Oracle, des infras VMWare, des baies de stockage NetApp...) au plus près de nos infrastructures cloud natives.

Avec les technologies de containerisation et d'orchestration, nous créons un niveau d'abstraction supplémentaire qui donne aux clients la possibilité de s'affranchir de l'adhérence des plateformes. Anthos permet de déployer des applications sur différents types de plateformes tout en gérant la configuration, la sécurité, le catalogue d'images, ou encore la télémétrie du trafic, de façon centralisée sans avoir à agréger de multiples outils de monitoring.

BigQuery offre la possibilité de fédérer des traitements sur plusieurs plateformes, avec un point de contrôle unique.

Tendances Infrastructure

La performance, l'agilité, la flexibilité et la résilience du cloud ne font plus débat. Quel que soit le secteur d'activité ou la taille de l'entreprise, les atouts du cloud en général et de Google Cloud Platform en particulier sont admis, compris et mesurés. Selon un récent rapport de Gartner, intitulé Leadership Vision for 2022 : Top 3 Strategic Priorities for I&O Leaders, l'hyperautomatisation, l'administration en temps réel des infrastructures et les technologies composables, constituent les principales tendances avec lesquelles les responsables d'infrastructures devront composer en 2022. Gartner estime que d'ici à 2025, 60 % des équipes infras utiliseront l'automatisation augmentée par l'IA dans l'ensemble de leur organisation, contre 1 % en 2020. Les ingénieurs de Google améliorent et inventent en permanence les outils nécessaires au bon fonctionnement de leurs services. Par extension, ce sont autant de nouveautés qui sont destinées à venir renforcer l'arsenal à disposition des stratégies cloud des entreprises.



Entre gouvernance et souveraineté

« Le développement de la culture cloud dans les entreprises, commente Seifeddin Mansri, Cloud Engineering Manager pour SFEIR, passe par un travail de fond sur le plan des process, des techniques et des logiciels, pour remporter l'adhésion de l'ensemble des collaborateurs. » Pour être à la hauteur de la transformation cloud, il est urgent de sortir du mythe de l'infrastructure, chasse gardée du DSI et accepter de se recentrer sur l'essentiel.

Au rang des principaux freins à abolir, les questionnements liés à la gouvernance et à la souveraineté. Le fait que Google Cloud se dote de Datacenters en France et en Europe contribue à la réassurance mais cela ne suffit pas toujours. Pour Seifeddin Mansri, « les entreprises cloud native s'interrogent moins sur ces aspects car elles ont parfaitement intégré que la balance bénéfices/risques penche en faveur du cloud ». Le revers de la médaille : une tendance à parfois aller trop vite sur des sujets sensibles. « Les grandes entreprises peuvent être parfois plus réticentes par rapport au cloud car elles redoutent les problématiques de gouvernance et de souveraineté, et optent donc pour des architectures hybrides ou multi-cloud ».

L'optimisation de la performance des infrastructures est une promesse du cloud.

Les entreprises repensent la gouvernance du cloud sous ce prisme et « n'hésitent pas à mettre les différents cloud providers en concurrence en

fonction des besoins et des applications, déclare Seifeddin Mansri. La gestion de l'infrastructure entre alors dans une autre dimension et transforme les métiers sans les remettre en cause. »

La sécurité : l'épine dorsale de tout projet cloud

Cette dimension multi-cloud expose les organisations à de nouveaux enjeux, au premier rang desquels, celui de la sécurité. « S'engager dans la transformation cloud, c'est accepter qu'une partie des aspects de sécurité soient délégués au cloud provider. Délocaliser la gestion de la sécurité peut être vécu comme une perte de contrôle », note Seifeddin Mansri. Les couches les plus basses de la sécurité sont en effet gérées par le cloud provider. Seules quelques briques de sécurité demeurent à la charge de l'utilisateur. Dans une dimension serverless, où l'entreprise utilise des services sans se soucier de l'infrastructure sous-jacente, la gestion de la sécurité se limite même à la simple gestion des accès aux contenus. « Les enjeux réglementaires de conformité ou de données très sensibles comme les données bancaires ou de santé restent des points d'inquiétude, mais ces aspects sont davantage juridiques que techniques. L'essentiel demeure la sensation de perte de contrôle qui génère une certaine frustration. »

GreenOps : entre Green IT et FinOps

À l'instar du phénomène DevOps, qui consiste à concilier les métiers de développeur logiciel (dev) et d'administrateur de systèmes et d'architectures (ops), l'une des tendances les plus fortes en 2022 consistera à accélérer les dimensions Green IT (pour maîtriser son impact environnemental) et FinOps (pour s'assurer de créer de la valeur tout en maîtrisant les coûts). « *Les entreprises sont toujours plus sensibilisées aux enjeux de responsabilité sociale et environnementale*, précise Seifeddin Mansri. *Google Cloud Platform prend cet enjeu très au sérieux.* » Très récemment, le cloud provider a lancé la solution Active Assist. Sa mission : faciliter les prises de décision liées au cloud, grâce au machine learning et à l'intelligence artificielle.

Cette fonctionnalité a vocation à aider les entreprises à identifier les activités cloud inactives et en mesurer l'empreinte carbone. Les organisations seront ainsi en mesure de décider de les conserver ou de les supprimer. L'an dernier, ces projets inactifs ont représenté plus de 600 tonnes de CO2. En complément, Google Cloud a développé Carbon Sense Suite, qui regroupe l'ensemble de ses fonctionnalités et solutions qui permettent aux entreprises de comprendre, mesurer et, in fine, réduire leurs émissions de CO2. Pour rappel, Google est neutre en carbone depuis 2007.

En effet, Google compense intégralement l'énergie utilisée par ses datacenters grâce à des énergies renouvelables et a pour objectif d'opérer sans carbone en permanence et partout dans le monde d'ici à 2030.

S'inscrire dans l'ère du Serverless

Le rapport de l'institut Gartner met en lumière une autre réalité majeure. À l'horizon 2025, les DSI seront confrontés à des difficultés de recrutement telles qu'ils seront amenés à confier près de 65 % des postes de responsable Infrastructure à des personnes n'ayant aucune expérience dans le domaine. Une réalité qui plaide en faveur de l'architecture serverless.

L'informatique serverless est un modèle qui affranchit l'entreprise de la gestion des serveurs. Le fournisseur de cloud se charge des missions d'administration et de maintenance au quotidien. Il assure la gestion de l'infrastructure de serveurs, assure son bon fonctionnement et la met à l'échelle en fonction des besoins de l'entreprise. « *Optimisation des coûts, automatisation, l'architecture serverless, loin d'être une incursion dans le pré-carré des DSI, est une perspective de recentrer les équipes sur des tâches plus exigeantes et plus proches de leur cœur de métier* », analyse Seifeddin Mansri.

Nous avons monté une équipe appelée Cloud Center of excellence. Elle accompagne chaque équipe dans leur migration sur le cloud pour rassurer, conseiller, former. Elle accomplit également cette mission auprès de la direction générale qui a besoin de réassurance.



Johan Protin
Head of Cloud Engineering -
Hipay

Nous avons migré l'intégralité de notre infrastructure sur le cloud en moins d'un an. L'adoption du cloud nous a permis d'attirer de nouveaux talents, car les nouveaux entrants dans l'entreprise ont plus d'appétence pour l'infrastructure actuelle que pour nos outils initiaux. L'autre bénéfice du cloud, c'est la maîtrise des délais dans le déploiement des applications. Une dimension qui répond clairement à notre démarche test and learn native.



Geoffrey Berard
VP of Engineering,
Deputy CTO · lesfurets.com

Anthos : écrire une fois, exécuter partout

Introduite officiellement lors de la conférence Cloud Next 2019, Anthos est une plateforme de Google Cloud qui permet d'exécuter vos applications sur votre infrastructure existante on-premises ou sur le cloud public, de manière simple, flexible et sécurisée. La promesse d'Anthos : incarner le principe « écrire une fois et exécuter partout », que ce soit sur Google Cloud Platform ou on-premises mais également sur des clouds tiers, notamment Amazon Web Services ou Microsoft Azure. Il est même possible aujourd'hui de déployer Anthos sur des serveurs physiques (bare metal). Anthos est bâtie sur les fondations de Google Kubernetes Engine (GKE), qui joue le rôle de centre de commande et de contrôle. Les utilisateurs peuvent ainsi gérer l'infrastructure distribuée exécutée sur le cloud ou sur un data center sur site. Anthos peut être considérée comme une offre de modernisation d'applications basée principalement sur Kubernetes. Toutefois, Anthos donne accès à d'autres technologies modernes qui augmentent sa puissance telles que :

- Istio : une solution de maillage de services open source qui réduit la complexité des déploiements de cloud computing ;
- Knative : une extension de Kubernetes pour gérer les environnements serverless.

Envie d'apprendre à développer vos applications sur Anthos ? Sachez que Google met à disposition la Anthos Developer Sandbox, qui peut être utilisée gratuitement. Il suffit pour en disposer de posséder un compte Google.



Seifeddin Mansri
Cloud Engineering Manager
pour SFEIR

L'avis de l'expert SFEIR

« Anthos permet de créer et de gérer des applications hybrides et multi-cloud modernes sans verrouillage. Il permet ainsi d'accélérer la productivité des développeurs et la livraison des logiciels en apportant les avantages des services cloud, des conteneurs et du serverless dans les déploiements grâce aux outils

cloud natifs et aux bonnes pratiques de Google. Anthos intègre la sécurité à chaque étape du cycle de vie des applications, du développement à l'exécution en passant par les pipelines de CI/CD. Ainsi, il automatise la gestion de la sécurité et des règles pour tous les déploiements. »

Cloud Functions :

Ne payez que pour ce que vous utilisez...

17

Avec l'offre Cloud Functions, vous pourrez assurer le développement de vos solutions plus rapidement en créant et en exécutant des petits extraits de code qui répondent à des événements. Ce produit, dévoilé en février 2016, est un service FaaS (Function as a Service). Il permet de mettre en place des services à l'échelle d'une fonction unique, loin de la complexité d'une application complète, d'un container ou d'une VM. Le but ici est d'écrire un morceau de code qui s'active en réponse à un événement (l'écriture d'un fichier sur Cloud Storage par exemple), sans serveur ni environnement d'exécution à gérer.

Avec la sortie de sa deuxième génération, Cloud Functions passe à la vitesse supérieure, en se dotant de plusieurs « super pouvoirs », y compris le support d'Eventarc qui lui permet d'écouter plus de 90 sources d'événements comme BigQuery, Cloud SQL, Cloud Storage, Firebase. Le support des langages Javascript (Node.js), Go, Python, Java et .Net Core est actuellement disponible.

Ces nanoservices, par définition éphémères, s'activent donc uniquement en fonction d'un événement spécifique, pour disparaître dès leur tâche accomplie. Le mode de facturation est particulièrement avantageux. En effet, le client ne paye que pour le temps d'exécution de la fonction – avec une précision de 100 millisecondes –, proportionnellement aux ressources (mémoire, CPU) utilisées.

Cloud Run : une

plateforme au service de votre activité

18

Google Cloud Run est un service de type CaaS (Container as a Service) qui a été dévoilé en avril 2019. Sa mission : vous permettre d'exécuter des applications packagingées dans des conteneurs. A l'instar d'offres comme Cloud Functions, Cloud Run repose sur une architecture sans serveur (serverless). Ainsi, il vous permet de vous concentrer sur l'essentiel : la création d'applications de qualité. Cloud Run est construit à partir de Knative, une surcouche open source à Kubernetes pour les applications serverless. Cela vous permet de choisir d'exécuter vos conteneurs soit entièrement gérés avec Cloud Run, soit dans votre cluster Google. Vous pouvez créer des applications à l'aide de votre langage favori, vos dépendances et vos outils préférés : Kubernetes Engine avec Cloud Run pour Anthos, ou bien sur votre propre cluster Kubernetes on-premises (sur site) sur lequel vous aurez installé Knative.

L'offre était initialement limitée à des conteneurs « sans état » (stateless) pour des applications HTTP. Il est désormais possible de disposer d'applications « avec état » en s'appuyant sur l'accès à Cloud SQL depuis les conteneurs Cloud Run. Vous pourrez également disposer d'applications capables de réagir à des événements autres que HTTP comme Pub/Sub, le service de diffusion d'événements au fil de l'eau de Google Cloud, grâce à la nouvelle solution Eventarc.

Le problème de démarrage à froid inhérent aux solutions serverless, où l'on décide d'exécuter une application inactive pour la première fois (ce qui crée une petite latence), a été résolu pour Cloud Run en donnant la possibilité de garder un premier conteneur éveillé à un tarif très préférentiel. Cloud Run fonctionne parfaitement avec l'écosystème de conteneurs : Cloud Build, Cloud Code, Artifact Registry et Docker.



Seifeddin Mansri
Cloud Engineering Manager
pour SFEIR

L'avis de l'expert SFEIR

« Cloud Functions offre une expérience développeur simple et intuitive. Les développeurs peuvent ainsi se concentrer sur leur code et laisser Google Cloud gérer l'infrastructure opérationnelle. »



Seifeddin Mansri
Cloud Engineering Manager
pour SFEIR

L'avis de l'expert SFEIR

« Cloud Run est un produit entièrement géré qui simplifie et accélère le développement et le déploiement d'applications. Il assure une très bonne expérience de développement de bout en bout grâce à ses intégrations avec d'autres produits comme Cloud Code, Cloud Build, Cloud Monitoring ou Cloud Logging. »

Private Service Connect : l'atout sécurité et conformité

19

Les entreprises souhaitent souvent accéder aux API Cloud en mode privé pour des raisons de sécurité et de conformité. C'est la fonction même que propose de remplir l'offre Private Service Connect qui peut être utilisée pour configurer et gérer ce type d'accès au sein de vos réseaux VPC (les VPC, ou virtual private cloud, sont les réseaux privés virtuels créés automatiquement par Google pour relier l'ensemble de vos ressources cloud).

Avec Private Service Connect, vous pouvez créer des endpoints privés à l'aide d'adresses IP internes globales au sein de votre réseau VPC. Vous pouvez attribuer des noms DNS plus explicites à ces adresses IP internes. Au lieu d'envoyer des demandes d'API à des endpoints de service public tels que storage.googleapis.com, vous pouvez envoyer les demandes au point de terminaison Private Service Connect, qui est privé et interne à votre réseau VPC. Ces noms et adresses IP sont internes à votre réseau VPC et à tous les réseaux sur site qui y sont connectés à l'aide de tunnels Cloud VPN ou de rattachements Cloud Interconnect (VLAN).

L'avantage ? Vous pouvez contrôler le trafic sur votre endpoint à tout moment et démontrer que le trafic reste bien dans Google Cloud.

Workflows : gérez les flux de travail

20

Vous souhaitez orchestrer n'importe quelles API, services Google Cloud ou services personnalisés avec une API ou bien automatiser les processus de façon fiable, ou encore implémenter un traitement en temps réel avec des exécutions à faible latence basées sur des événements ? Alors l'offre Workflows de Google Platform constitue un point de passage incontournable ! Les flux de travail qui peuvent être traités par Workflows permettent de combiner des services tels que des services personnalisés hébergés sur Cloud Run ou Cloud Functions, des services Google Cloud tels que Cloud Vision AI et BigQuery et n'importe quelle API HTTP.

Les workflows s'exécutent sans serveur et s'adaptent en fonction des besoins. Aucuns frais n'est facturé en cas d'inactivité. Étant donné qu'un workflow ne contient aucune dépendance de code ou de bibliothèque, il ne nécessite pas de correctifs de sécurité. Une fois que vous avez déployé un workflow, vous pouvez vous assurer qu'il s'exécutera de manière fiable sans maintenance. Un workflow peut conserver l'état, réessayer, interroger ou attendre jusqu'à un an.

Un workflow se compose d'une série d'étapes décrites à l'aide de la syntaxe Workflows et peut être écrit en YAML ou JSON. Il s'agit de la définition du workflow. Une fois le workflow créé, il est déployé, ce qui le rend prêt à être exécuté. Toutes les exécutions de workflows sont indépendantes, et le scaling rapide du produit permet un grand nombre d'exécutions simultanées.

L'avis de l'expert SFEIR



Seifeddin Mansri
Cloud Engineering Manager
pour SFEIR

« En permettant aux utilisateurs de publier et consommer des services à l'aide d'adresses IP internes définies au sein de leur réseau VPC, Private Service Connect peut être utilisé pour accéder aux API et services de Google, ou aux services gérés d'un autre réseau VPC. »

L'avis de l'expert SFEIR



Seifeddin Mansri
Cloud Engineering Manager
pour SFEIR

« Workflows est une plateforme d'orchestration qui exécute des services dans un ordre bien défini. Ces flux de travail peuvent combiner une variété de services, y compris des services personnalisés hébergés sur Cloud Run ou Cloud Functions, des services Google Cloud tels que Cloud Vision AI et BigQuery, et plus généralement n'importe quelle API basée sur HTTP. »

Compute

Compute Engine	Compute Engine est le produit IaaS (Infrastructure as a service) par excellence. Il permet le provisioning de machines virtuelles ayant des configurations différentes en termes de processeur, mémoire, stockage et réseau.
Kubernetes Engine	Kubernetes Engine (GKE) comme son nom l'indique se base sur le système d'orchestration de conteneurs open source Kubernetes.
VMware Engine	Google Cloud VMware Engine fournit un service pour provisionner votre infrastructure VMware sur Google Cloud en quelques dizaines de minutes sans modifications majeures de votre infrastructure VMware existante.
Anthos	Anthos est une offre de modernisation d'applications de Google Cloud basée principalement sur Kubernetes mais aussi sur d'autres technologies modernes qui augmentent sa puissance telles que Istio ou Knative. Anthos vous permet d'exécuter vos applications n'importe où, sur votre infrastructure existante on-premises ou sur le cloud public, de manière simple, flexible et sécurisée.

Serverless

Cloud Run	Google Cloud Run est un service de type CaaS (Container as a Service) qui vous permet d'exécuter des applications packagées dans des conteneurs en mode serverless puisqu'il se base sur Knative, une surcouche open source à Kubernetes pour les applications serverless.
Cloud Functions	Cloud Functions est un service de type FaaS (Function as a Service). Il permet d'écrire un morceau de code qui s'active en réponse à un évènement (l'écriture d'un fichier sur Cloud Storage par exemple), sans serveur ni environnement d'exécution à gérer.
App Engine	Google App Engine (GAE) est né en 2008 constituant l'un des premiers produits de la GCP. Il a fallu du temps pour faire comprendre son principe de PaaS (Platform as a Service) et voir les différents frameworks du marché s'adapter à ses spécificités. Son rôle est de permettre de créer des applications Web et des systèmes back-end pour applications mobiles évolutives, en faisant cette fois totalement abstraction de l'aspect serveur. On se concentre sur le code, et rien que le code. Si jamais l'application devient très populaire, App Engine se chargera de trouver les ressources nécessaires à son fonctionnement.

Network Services

Private Service Connect	Private Service Connect permet la consommation privée de services sur des réseaux VPC appartenant à différents groupes, équipes, projets ou organisations. Il est ainsi possible de publier et consommer des services à l'aide d'adresses IP internes à votre réseau VPC.
-------------------------	---

Application Integration

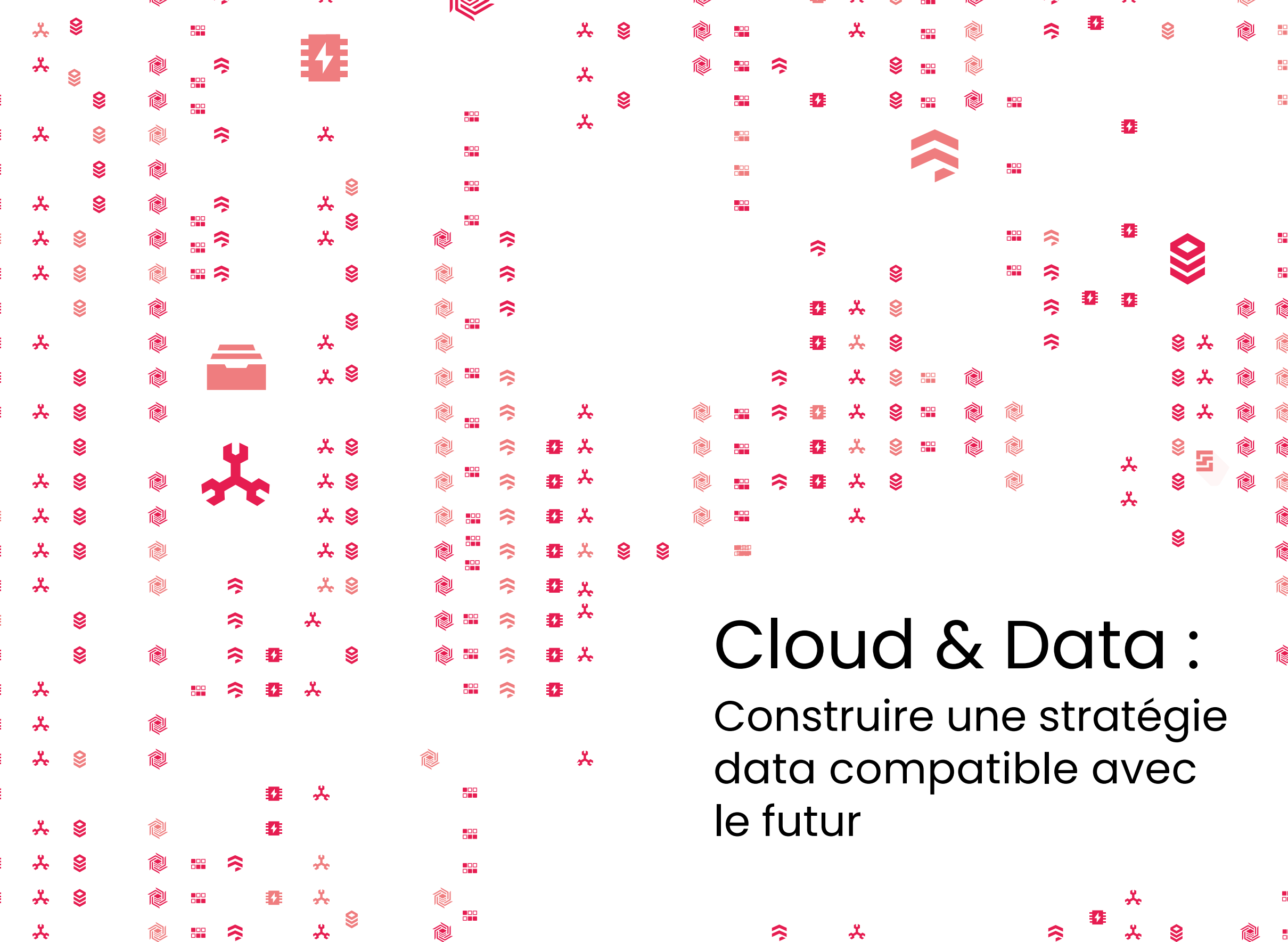
API Gateway	Avec API Gateway, vous pouvez créer, sécuriser et surveiller des API pour les back-ends serverless de Google Cloud, y compris Cloud Functions, Cloud Run et App Engine. A base de proxy Envoy, API Gateway vous offre des performances élevées, une évolutivité et la liberté de vous concentrer sur la création d'applications de qualité.
Workflows	Workflows permet de combiner les services et les API Google Cloud pour créer facilement des applications fiables, automatiser des processus et des pipelines de données et de machine learning.
Eventarc	Eventarc vous permet de créer des architectures pilotées par les événements sans avoir à implémenter, personnaliser ou maintenir l'infrastructure sous-jacente.

CI/CD

Artifact Registry	Artifact Registry est un produit qui permet une gestion centralisée des images de conteneurs et des packages de langage de programmation (tels que Maven et npm). Il est entièrement intégré aux outils et aux environnements d'exécution de Google Cloud et prend en charge les protocoles d'artefacts natifs. Cela facilite son intégration aux outils CI/CD pour configurer des pipelines automatisés.
Cloud Deploy	Google Cloud Deploy rend la livraison continue d'application vers GKE simple et puissante. Les utilisateurs auront la possibilité de définir les versions et les faire progresser dans les différents environnements tels que dev, staging et production. Google Cloud Deploy permet de promouvoir et de restaurer facilement et en une seule étape les versions via la console web, l'interface de ligne de commande ou l'API. Les métriques intégrées permettent d'avoir un aperçu de la fréquence et du succès des déploiements.

Tools

Carbon Footprint	Carbon Footprint est un service de calcul et de surveillance des émissions brutes dans le cloud au fil du temps par projet, produit et région, offrant ainsi une transparence sur les émissions associées aux applications cloud et fournissant aux équipes informatiques et aux développeurs des mesures qui peuvent les aider à améliorer leur empreinte carbone.
------------------	---

The background of the slide is a light gray with a repeating pattern of small, red, stylized icons. These icons include various symbols related to cloud computing and data, such as server racks, data centers, clouds, and abstract geometric shapes. The icons are arranged in a grid-like pattern, with some icons appearing more frequently than others.

Cloud & Data :

Construire une stratégie data compatible avec le futur

3 questions à...



Sébastien Colombo, Head of Data & Analytics France, Google Cloud, s'est prêté aux trois questions.

Existe-t-il pour vous un schéma idéal pour assembler les différentes technologies data de Google et parvenir à mettre en place une plateforme de données à même de supporter la stratégie data-driven des entreprises ?

D'un point de vue technologique, l'assemblage est assez simple : un pipeline de données pour ingérer la data et la mettre dans une base analytique, puis un outil self-service pour que les utilisateurs puissent les exploiter. Le plus compliqué est de déterminer où on met le curseur, entre le rôle de l'IT et celui des métiers.

La BI traditionnelle était dans un mode producteur-consommateur figé, où les équipes IT montaient des systèmes complexes qui fournissaient des rapports statiques. La vague des outils self-service a bousculé ce schéma, en permettant aux utilisateurs de constituer leurs propres modèles, mais au prix de la gouvernance : les métiers se sont retrouvés avec des données souvent incorrectes, ce qui n'a pas amélioré leur productivité.

Une base massivement parallèle comme Big Query rebat les cartes. Peu chère, scalable à l'infini, elle permet d'agréger toute la donnée et de la mettre à disposition des métiers, au travers de "produits data", des solutions créées avec Looker par exemple, qui ne leur demandent que quelques clics pour accéder à l'information et prendre des décisions. C'est une inversion du schéma : on traite vraiment la donnée comme un produit, qu'on met entre les mains des consommateurs.

Google Cloud multiplie les possibilités d'utiliser de l'intelligence artificielle sans avoir besoin de créer soi-même des algorithmes ; est-ce que l'avenir du machine learning est d'être caché, intégré au sein des autres solutions ?

Idéalement, les utilisateurs doivent pouvoir bénéficier de cette intelligence dans leurs processus et applications de façon transparente. On parle chez Google d'AI comme "applied" et "invisible" : l'intelligence artificielle et le

machine learning sont présents dans tout ce qu'on fait. Nous travaillons par exemple à la possibilité d'interroger les données dans Looker en langage naturel.

Il faut que l'IA soit facile à utiliser, facile à appliquer, transparente pour les utilisateurs, les "decision makers". Pour y parvenir, nous devons aussi donner aux data scientists, aux data engineers et à tous les experts de la donnée les outils nécessaires pour créer les modèles de machine learning. Aussi bien avec des outils experts qu'avec des solutions qui démocratisent l'accès au machine learning, comme Big Query, qui permet de créer des modèles juste avec du SQL - ce qui accélère aussi la mise en production, puisque tout est déjà dans la même solution.

Au-delà de la mise en œuvre, comment s'assurer justement que les algorithmes passent bien en production, supportent la mise à l'échelle et que les résultats soient de qualité ?

On aborde là le sujet de l'industrialisation : passer un modèle en production, c'est une étape importante. Souvent, les data scientists se focalisent sur la

production du modèle, en mode pilote, mais se préoccupent peu de la suite. Or, les modèles peuvent s'avérer mal adaptés aux données de production, et on sait qu'ils se dégradent avec le temps, lorsque les conditions changent. Si on considère la donnée comme un produit, cela change la perspective. On va commencer à monitorer son utilisation, analyser les gains de performance que ce produit octroie, mesurer sa valeur : il faut faire de l'analytique sur son analytique.

Et puisqu'on raisonne en termes de produits, l'idéal sera de développer une politique de mises à jour régulières, qui viendront prendre progressivement le relais, sur une partie des données, dans une optique d'amélioration continue.

Tendances data

Si l'on se réfère aux prévisions relayées par Statista, le volume de données générées ou répliquées à l'échelle mondiale a été multiplié par plus de trente au cours de la décennie écoulée. On a passé le cap des 64 zettaoctets en 2020, et la courbe de croissance est exponentielle (un zettaoctet représente un milliard de teraoctets !). Certaines prévisions laissent entrevoir, à l'horizon 2025, un volume de données de 180 zettaoctets, soit une croissance de près de 40 % en quelques années. Ce constat global ne s'explique pas seulement par l'acculturation des entreprises par rapport à l'enjeu stratégique que constitue la data. « *Les solutions cloud contribuent à la généralisation des usages de la donnée, observe Florent Legras, Data Engineering Manager chez SFEIR. Plus faciles à manipuler, plus simples à exploiter, la difficulté ne porte plus sur la capacité à traiter de très importants volumes de données. De 100 teraoctets à plusieurs petaoctets, les modules de la Google Cloud Platform ont permis cet élargissement de la base des utilisateurs de solutions data dans les entreprises.* »

Simplicité ne rime pas avec instantanéité

Puissance de calcul, ergonomie des modules fonctionnels, acculturation croissante des équipes, la place de la donnée dans les stratégies des entreprises, mais aussi dans les processus décisionnels du quotidien, ne cesse de croître. Mais, si elles sont plus accessibles et plus largement utilisées, les solutions cloud dédiées à la data « *nécessitent malgré tout une réelle technicité* », confie Florent Legras.

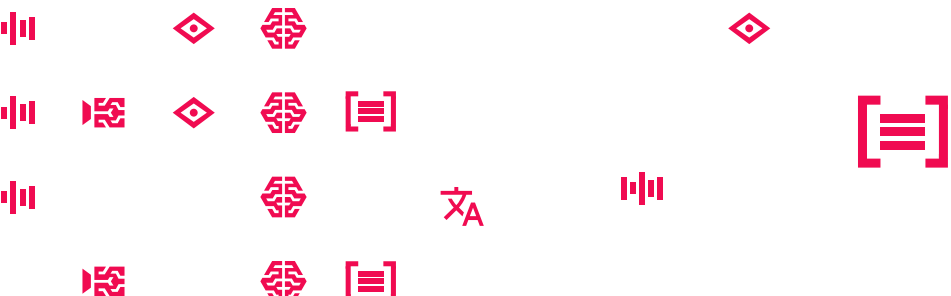
Par ailleurs, si la simplicité est plus grande, il ne faut pas croire pour autant au miracle technologique. « *Gérer et exploiter des volumes de données toujours plus colossaux nécessite du temps et la simplicité ne rime pas toujours avec instantanéité* », continue Florent Legras. L'une des tendances les plus structurantes en matière de gestion de la data sur le cloud, c'est le déport de cette technicité au profit de solutions low code / no code, qui permettent à des utilisateurs de créer des tableaux de bord voire des applications manipulant de la donnée sans passer par le service IT.

« *Dans ce contexte, le recours à des solutions comme Cloud Dataprep ou Data Fusion va sans aucun doute s'accélérer considérablement dans les mois et les années à venir* », analyse Florent Legras. Parmi les conséquences les plus évidentes liées à cette transformation des usages, qui repose sur un double phénomène d'ouverture et de densification de la donnée, « les enjeux de gouvernance de la data, et

la place centrale du Data lineage qui est indispensable à la traçabilité de la donnée sur l'ensemble de son cycle de vie ». Ces enjeux de gouvernance ont une autre incidence sur les stratégies des entreprises qui s'interrogent également sur la place que doit occuper un Data Catalog dans l'arsenal de la gestion optimale des données.

Une exigence de transparence dans les coûts...

Parce que les solutions data disponibles au sein de la Google Cloud Platform sont susceptibles d'être utilisées par des non-spécialistes, il est nécessaire d'accompagner les entreprises. « *La richesse de la plateforme est telle que le risque de se perdre dans une offre pléthorique est réel, observe Florent Legras. C'est aussi dans cette dimension que l'accompagnement délivré par les experts de SFEIR contribue à faire les bons choix.* » Des choix d'autant plus cruciaux que derrière les enjeux data, plane en permanence l'ombre du retour sur investissement. « *L'agilité et la facilité de mise en œuvre intrinsèque des solutions cloud est, en la matière, un précieux atout* », continue l'expert. En effet, les organisations ont la possibilité d'explorer le champ des possibles, d'élargir leurs horizons en lançant des POC (Proof Of Concept) pour évaluer la pertinence d'une solution par rapport à un usage donné.



« Le recours aux POC permet d'optimiser le ROI. Une fois le ROI confirmé, il est possible d'entrer dans une phase d'industrialisation ». Si les technologies cloud ont la réputation (justifiée !) de réaliser des économies d'échelles importantes, des freins persistent dans les entreprises. En cause : le manque de visibilité sur les coûts du fait d'une facturation à l'usage. « Potentiellement, les cas d'usages sont si nombreux et les volumes de données si importants que cela peut inquiéter, c'est tout l'intérêt de s'engager sur le chemin des POC pour éviter toute surprise. »

L'entreprise devra également évaluer sa stratégie data à l'aune de son approche FinOps, en comparant la valeur prévue et les coûts prévisionnels de la plateforme.

Un indispensable accompagnement

Parmi les grands enjeux liés à la place de la data, la notion de culture de la donnée est centrale. Ce défi, déjà souligné par Florent Legras, est naturellement lié à la vulgarisation des usages de la data. Les collaborateurs d'une entreprise n'ont souvent qu'une idée très vague, et datée, de ce que la donnée permet de mettre en œuvre.

« Les solutions cloud permettent à des profils plus profanes de travailler la donnée. Mais la complexité reste réelle et les faux semblants ou les idées reçues à combattre sont réels. » Sur le plan de la sécurité bien sûr, mais aussi par rapport aux questions de confidentialité ou encore de localisation de la donnée, des questionnements demeurent. « Imaginer que Google va exploiter les données de l'entreprise pour son propre compte, les fantasmes liés au Patriot Act, nous amènent à faire de réels efforts de pédagogie auprès des entreprises que nous accompagnons. » L'enjeu : réaliser un travail de démythification qui peut constituer un frein à l'exploitation du plein potentiel de la data. « C'est ce genre de missions que nous accomplissons au quotidien, confirme Florent Legras. Former, sensibiliser, inspirer, accompagner, rassurer... Une présence qui permet aux entreprises d'accélérer la valorisation de leurs actifs data ! »

Nous avons migré notre Datalake sur GCP ce qui nous épargne le recours aux solutions classiques de datawarehouse, en nous appuyant sur du BigQuery car nous avons d'importants besoins en self-service analytics au service des équipes business.



Guillaume Pinot
Technology Manager pour
Renault Digital

Nous voulions renforcer l'autonomie de nos équipes de data scientists. Le cloud nous a permis de les recentrer sur le cœur de métier sans qu'ils aient à subir ou gérer des contraintes de gestion d'infrastructures.



Paul Marcombes

Head of Data pour Nickel

BigQuery : le compagnon idéal des Data Analysts

Disponible depuis 2010, BigQuery est un service serverless totalement managé qui repose notamment sur la technologie Dremel de Google. La vocation de cette technologie, c'est d'offrir la possibilité d'extraire facilement des informations depuis de grands ensembles de données. De nombreux cas d'usage sont envisageables, à commencer par l'utilisation qu'en fait Google, pour savoir combien d'affichages Google Ads (publicités de la régie de Google) ont eu lieu à un moment T, sur une zone géographique ciblée. Dremel est en mesure d'effectuer ce genre de recherche sur des ensembles de données immenses en quelques secondes, grâce à des requêtes de type SQL très simples. Ce système de requêtes massivement parallèle peut effectuer un scan de 35 milliards de lignes en quelques dixièmes de secondes, et ce, sans index ! BigQuery permet le même type d'opérations pour des clients externes. Par exemple, associé à Google Analytics Premium et à un outil de CRM, il permet de réaliser des reportings très précis et très rapidement. BigQuery peut digérer un téraoctet de données en quelques secondes et prend en charge des pétaoctets à traiter sans problème. Ces données peuvent être chargées par exemple depuis Cloud Storage, Cloud Firestore ou encore envoyées en flux continu vers BigQuery. BigQuery est le compagnon idéal des Data Analysts. Le service intègre de nombreuses applications. Il simplifie et accélère les usages dédiés aux datawarehouses, notamment en permettant de s'inscrire dans une dimension cross data, d'analyser et explorer des données, de remplacer des batchs de traitement de données et de générer des reportings. BigQuery permet de requêter de gros volumes de données très rapidement. Il est désormais possible d'entraîner des modèles de Machine Learning (régression linéaire, régression logistique, kmeans) directement dans cette plateforme avec une simple requête SQL. Enfin, BigQuery se charge d'optimiser votre modèle en fonction de vos données et de le mettre à disposition pour l'inférence (l'application du modèle sur les données à traiter).

L'avis de l'expert SFEIR



Florent Legras

Data Engineering Manager
pour SFEIR

« Nous avons opté pour BigQuery parce que c'est l'outil central pour les données chez GCP. Nous avons dans un seul produit un entrepôt de données permettant d'interroger des téraoctets de données et de se connecter nativement à la majorité des outils de stockage de données sur GCP. En plus de cela, BigQuery permet de faire des algorithmes d'apprentissage automatique en SQL. »

Dataflow : le traitement de données par lot ³²

Dataflow est un service de traitement de données entièrement géré, compatible avec l'exécution de pipelines par flux (stream) et par lots (batches). Les travaux pour donner naissance à cet outil sont également à l'origine du modèle de programmation Apache Beam.

Concrètement, en tant que développeur, il suffit d'écrire le code de traitement des données. Le service prend ensuite en charge la gestion de l'infrastructure nécessaire pour réaliser l'opération. Google Cloud Platform rassemble les données, optimise les performances et gère les ressources. Ce service, qui s'intègre parfaitement avec d'autres modules de GCP comme Cloud Storage, Cloud Pub/Sub, Cloud Firestore, Cloud BigTable et BigQuery, rassemble des applications dans tous les domaines et peut gérer les cas d'usage complexes comme la détection de fraude en finances ou encore l'analyse de données en provenance de millions d'objets connectés... en faisant des appels d'API de modèles d'intelligence artificielle.

Dataflow est entièrement managé, il propose une seule API pour les traitements batch et les traitements streaming. Il délivre de hautes performances et offre une scalabilité automatique. En associant l'autoscaling des ressources à un traitement par lot (promesse de maîtrise des coûts !), Dataflow offre des capacités presque illimitées pour gérer les pics d'activité saisonniers sans dépasser le budget.

Vertex AI : le couteau suisse du Machine Learning ³³

Vertex AI regroupe les services Google Cloud pour créer des modèles de machine learning dans une interface utilisateur et une API unifiées. Avec Vertex AI, vous pouvez comparer facilement des modèles à l'aide d'AutoML. Tous les modèles sont stockés dans un seul dépôt central. Une centralisation qui facilite le déploiement de ces mêmes modèles.

Avec Vertex AI, Google cible principalement les néophytes en leur permettant un accès simplifié au machine learning. Mais elle apporte également aux profils plus experts une ergonomie intuitive qui leur permet de gagner un temps précieux pour assurer la gestion des tâches de base.

Vertex AI est intégré de manière native à BigQuery, Dataproc et Spark. Vous pouvez utiliser BigQuery ML pour créer et exécuter des modèles de machine learning dans BigQuery à l'aide de requêtes en SQL standard sur des outils et feuilles de calcul d'informatique décisionnelle existants, ou exporter des ensembles de données depuis BigQuery directement vers Vertex AI Workbench et y exécuter vos modèles. Vous pouvez utiliser des outils tels qu'AutoML, Explainable, Vizier, Edge Manager, etc. pour travailler sur des modèles d'apprentissage automatique. Parmi les principales promesses de Vertex AI, on retiendra la perspective de réduire le coût de mise en place de votre propre infrastructure (via Kubernetes, par exemple) mais aussi la réduction de l'effort de configuration ou de gestion de votre propre infrastructure pour former des modèles d'apprentissage automatique. Les outils de modèle personnalisés de Vertex AI sont compatibles avec le codage ML avancé. La promesse, selon Google : une réduction de près de 80 % de lignes de code grâce aux bibliothèques personnalisées.

L'avis de l'expert SFEIR



Florent Legras

Data Engineering Manager
pour SFEIR

« Dataflow, est un outil ETL (Extraction, Transformation, Chargement) désormais totalement serverless avec Dataflow Prime. Il permet de développer et de déployer des pipelines de transformation de données, en streaming ou en batch ».

L'avis de l'expert SFEIR



Florent Legras

Data Engineering Manager
pour SFEIR

« Vertex AI est une plateforme qui regroupe l'ensemble des services d'IA de Google. Agissant comme un tout-en-un prêt à l'emploi, Vertex AI permet d'avoir en un seul endroit tous les outils qui permettent de déployer et d'optimiser les projets de data science. »

Cloud BigTable : indispensable pour les grands ensembles de données

Exploitée depuis 2005 en interne par Google, la version publique de BigTable n'est apparue qu'en 2015. C'est une des nombreuses technologies indispensables à Google et elle est vitale pour une large gamme de services : Google Search, Maps, YouTube ou encore Gmail, pour ne citer que quelques exemples servant chacun plus d'un milliard d'utilisateurs...

BigTable s'adresse en particulier aux applications qui nécessitent un débit et une évolutivité élevés. Ce service est également un excellent moteur de stockage pour les opérations MapReduce par lot (pour les analyses Big Data), le traitement et l'analyse par flux et les applications de machine learning (apprentissage automatique). BigTable stocke les données dans des tables extrêmement évolutives, chacune constituant une carte de clés/valeurs triées. Une table est composée de lignes, chacune décrivant généralement une seule entité, et de colonnes contenant des valeurs individuelles pour chaque ligne. Chaque ligne est indexée par une clé de ligne unique et les colonnes liées les unes aux autres sont généralement regroupées dans une famille de colonnes. Chaque colonne est identifiée par la combinaison d'une famille de colonnes et d'un qualificatif de colonne, c'est-à-dire un nom unique dans la famille de colonnes.

BigTable est un service de base de données NoSQL, un concept qui en s'éloignant des bases de données relationnelles classiques permet de s'adapter aux besoins du web moderne. BigTable est en effet capable de fonctionner sur plusieurs machines différentes simultanément, ce qui permet de monter en charge et de gérer des volumes de données colossaux. Les systèmes de gestion de base de données non relationnelles étaient déjà utilisés sur les systèmes mainframes aux débuts de l'informatique, pour des usages où les écritures étaient plus nombreuses que les lectures. NoSQL modernise ce principe et l'adapte aux technologies actuelles. En basant son application sur BigTable, un client peut envisager de traiter des centaines de pétaoctets sans avoir à se soucier des besoins matériels qu'une telle opération nécessite. Une opération totalement impensable il n'y a pas si longtemps !



Florent Legras

Data Engineering Manager
pour SFEIR

L'avis de l'expert SFEIR

« Cloud BigTable est une base de données analytique NoSQL qui permet de traiter des pétaoctets de données avec de faibles latences. Cette base de données est idéale lorsque vous avez besoin de réactivité sur de grands ensembles de données. »

Data Fusion : décloisonnez les données

Cloud Data Fusion a été conçu pour permettre la création de datalakes évolutifs et distribués sur Google Cloud en intégrant des données provenant de plateformes cloisonnées. Les utilisateurs peuvent profiter de la scalabilité du cloud pour centraliser leurs données et en dégager ainsi davantage de valeur.

Parce que le recours à un large éventail d'outils déconnectés et de solutions hétéroclites pose de véritables problèmes en termes de qualité et de sécurité des données, Data Fusion se charge de toutes les réconcilier. Grâce à Cloud Data Fusion et à sa large gamme de connecteurs, d'interfaces visuelles et d'abstractions basées sur la logique métier, vous pourrez réduire vos coûts et améliorer la standardisation de vos actifs data.

Utilisable en libre-service, Cloud Data Fusion augmente la visibilité sur les données et aide les entreprises à mieux comprendre leurs clients. Comment ?

En découplant les données et en permettant le développement de solutions agiles d'entrepôt de données dans BigQuery. Avec une vue unifiée et fiable sur l'engagement et le comportement des clients, vous pouvez améliorer leur expérience, et ainsi accroître la rétention et les revenus associés à chaque client. Doté d'une interface intuitive (fonctionnant sur la base de simples glisser-déposer), de connecteurs pré-intégrés et d'un modèle en libre-service d'intégration de données sans code, Data Fusion élimine les goulots d'étranglement liés aux expertises techniques, et accélère l'obtention d'insights. Grâce à des fonctionnalités intégrées telles que la traçabilité des données de bout en bout, les métadonnées d'intégration et les services de sécurité et de protection des données cloud natives, Data Fusion contribue également à mieux respecter les exigences en matière de conformité.



Florent Legras

Data Engineering Manager
pour SFEIR

L'avis de l'expert SFEIR

« Data Fusion est une plateforme de données globale avec une approche no code / low code. Elle permet de créer des pipelines de données via une interface graphique et les outils de surveillance d'une plateforme de données tels que le lignage et le catalogage des données. »

Machine Learning

Auto ML	Cloud AutoML est constitué d'un ensemble de technologies de machine learning qui permet aux développeurs d'entraîner des modèles sans aucune connaissance en Data Science.
Cloud AI API	Google met à disposition des API destinées à aider à la mise en place d'applications "intelligentes". Ce sont en fait des modèles pré-entraînés grâce aux quantités de données impressionnantes que manipule Google, ce qui permet d'obtenir d'excellents résultats.

Traitement de données

Data Fusion	Basé sur le framework CASK Cloud Data Fusion est un ETL natif cloud avec une large bibliothèque de composants de transformation open-source et plus d'une centaine de connecteurs prêts à l'emploi pour une large gamme de systèmes et de formats de données.
Dataproc	Dataproc est basé sur un écosystème open source connu (Apache Hadoop, Apache Spark, Apache Pig et Apache Hive), qui ne nécessite pas d'apprendre de nouveaux outils ou API ni de modifier des projets existants. Comme de nombreux services de la GCP, son rôle est d'affranchir l'utilisateur de la gestion des ressources nécessaires à l'exécution d'une tâche donnée en créant un cluster totalement managé par Google.

Bus Messaging

Cloud Pub/Sub	Private Service Connect permet la consommation privée de services sur des réseaux VPC appartenant à différents groupes, équipes, projets ou organisations. Il est ainsi possible de publier et consommer des services à l'aide d'adresses IP internes à votre réseau VPC.
---------------	---

Stockage de données

Cloud Firestore	Firestore associe une base de données sans schéma capable de facilement s'adapter à l'évolution des applications, à un langage de requête basique, qui donne la possibilité de rechercher des données en fonction de plusieurs propriétés.
Cloud Spanner	Avec ce produit récent, disponible pour le public depuis 2017, on entre dans le monde des bases de données relationnelles et transactionnelles pouvant peser plusieurs pétaoctets, capables de fonctionner grâce à des bases de données relationnelles (SGBDR) regroupées sous la dénomination NewSQL.

Cloud SQL	Cloud SQL est un service de base de données entièrement géré, destiné à héberger des bases de données relationnelles MySQL, PostgreSQL et Microsoft SQL Server. 100% compatible avec les versions on-premises.
Cloud Bigtable	Bigtable est un service de base de données NoSQL, un concept qui en s'éloignant des bases de données relationnelles classiques permet de s'adapter aux besoin du web moderne. Bigtable est en effet capable de fonctionner sur plusieurs machines différentes simultanément, ce qui permet de monter en charge et de gérer des volumes de données colossaux.
Cloud Storage	Cloud Storage (GCS) est une solution de stockage d'objets (object store), destinée aux professionnels (développeurs, entreprises). Surcouché de Colossus, il s'agit, encore une fois, d'une technologie dérivée des efforts de Google pour mieux répondre aux besoins de ses applications.

Data management

Data Catalog	Data Catalog est un service de gestion des métadonnées entièrement géré et évolutif avec une interface de recherche pour la découverte de données.
Data Plex	Data Plex est un service qui vous aide à unifier les données distribuées et à automatiser la gestion et la gouvernance de ces données afin d'optimiser les analyses à grande échelle. Cet outil permet d'accélérer une approche DataMesh pour une plateforme data.

Ordonnancer

Composer	Apache Airflow managé par Google permettant de disposer d'un outil d'ordonnancement et schedulling.
----------	---

Méthodologie/ accompagnement

À vos côtés pour construire le futur

SFEIR accompagne les entreprises dans leur transformation numérique en mettant à leur disposition un accompagnement et des ressources spécialisées dans le développement d'applications de pointe. Pour aider ses clients à concevoir avec succès la dernière génération de solutions numériques et d'applications métier, pour contribuer à accélérer leur migration vers le cloud et dans le développement des usages de la data, **SFEIR s'appuie sur l'expertise de ses 700 développeur·euse·s spécialisé·e·s dans le développement d'applications de dernière génération.** SFEIR, c'est aussi un écosystème de partenaires technologiques mobilisés à vos côtés pour capitaliser sur les meilleures solutions afin de dynamiser votre stratégie.

Votre expérience, notre expertise...

Dans sa dimension infrastructure, comme dans son acception data, le cloud est, par nature, toujours en mouvement.

Les ingénieurs de Google améliorent et inventent en permanence les outils nécessaires au bon fonctionnement de leurs services. Par extension, ce sont autant de nouveautés qui sont destinées à venir renforcer l'offre de la Google Cloud Platform. Un contexte d'évolution continue, d'innovation permanente qui nécessite une veille de tous les instants. Une veille qui a vocation à tenir une promesse capitale : ne manquer aucune opportunité !

Qu'il s'agisse d'améliorer vos produits, votre recherche et développement, la connaissance de vos clients, de vos écosystèmes, de consolider vos stratégies, le cloud ouvre des perspectives quasiment infinies. Mais, pour bénéficier de ce formidable potentiel, deux éléments clés s'imposent.

D'un côté, l'expertise de votre métier, que vous êtes seul à maîtriser. De l'autre, une connaissance approfondie des solutions disponibles sur la Google Cloud Platform que nos experts possèdent et cultivent jour après jour. Concilier ces deux éléments essentiels constitue votre principale promesse de succès sur le chemin du cloud.

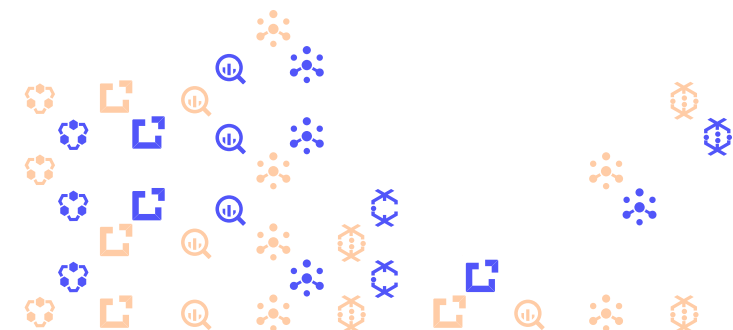
Un accompagnement à la carte

Lorsqu'une entreprise engage ou accélère sa transformation cloud, elle doit le faire parce que le projet correspond à une vision stratégique.

Une réalité qui vaut autant pour l'infrastructure cloud que pour un projet data. L'action des experts de SFEIR s'articule autour de trois axes. WENVISION vous permet de créer une stratégie gagnante basée sur l'innovation grâce à notre offre de conseil et d'accompagnement technique et technologique. Avec MAKE REAL, nous concevons, créons et dimensionnons à vos côtés les infrastructures qui assurent le développement optimal de vos projets.

Et parce que l'incarnation de votre stratégie cloud intègre nécessairement une dimension humaine forte, LEARN & SHARE, portée par les formations SFEIR Institute assure la formation de vos équipes.

En tant que "Premier Partner Google Cloud", SFEIR a démontré son expertise et son expérience dans l'accompagnement de ses clients dans leurs projets sur Google Cloud. Revendiquant les spécialisations Google Cloud liées à l'infrastructure et la Data Analytics comme sur le volet formation, SFEIR vous offre un accompagnement d'excellence dans toutes les dimensions de votre projet.



Le culte du sur-mesure

Un projet cloud comprend plusieurs strates et, par conséquent, les experts de SFEIR peuvent intervenir dans différents domaines. Des domaines techniques, bien sûr, mais pas seulement. L'acculturation des équipes de nos clients est une tâche prioritaire car il ne suffit pas d'avoir la meilleure infrastructure cloud. Il faut qu'elle soit parfaitement exploitée, maintenue, pilotée. Parce qu'en matière de transformation cloud, le pire ennemi demeure l'inaction, chaque collaborateur doit pouvoir se saisir de tous les enjeux et contribuer à l'effort de transformation.

Pour y parvenir, nos experts viennent en immersion au sein des équipes clients pour observer, comprendre analyser la situation et identifier les besoins. Une « Core Team », fédérant tous les métiers de l'entreprise peut

ensuite être constituée. Elle joue alors le rôle de référent autant que d'ambassadeur. Limiter un projet cloud à sa seule dimension technique est une erreur. Les enjeux politiques, stratégiques, humains doivent être embrassés pour s'engager sur le chemin du succès. C'est pourquoi nos experts agissent comme un trait d'union entre les équipes métiers et les C-Level ou les dirigeants. Des comités réguliers, de pilotage et/ou stratégiques sont organisés pour ajuster en permanence, le projet, la méthode et la stratégie. Le cloud, c'est toujours du sur-mesure !



Seifeddin Mansri
Cloud Engineering Manager
pour SFEIR

Concilier & réconcilier vision et application

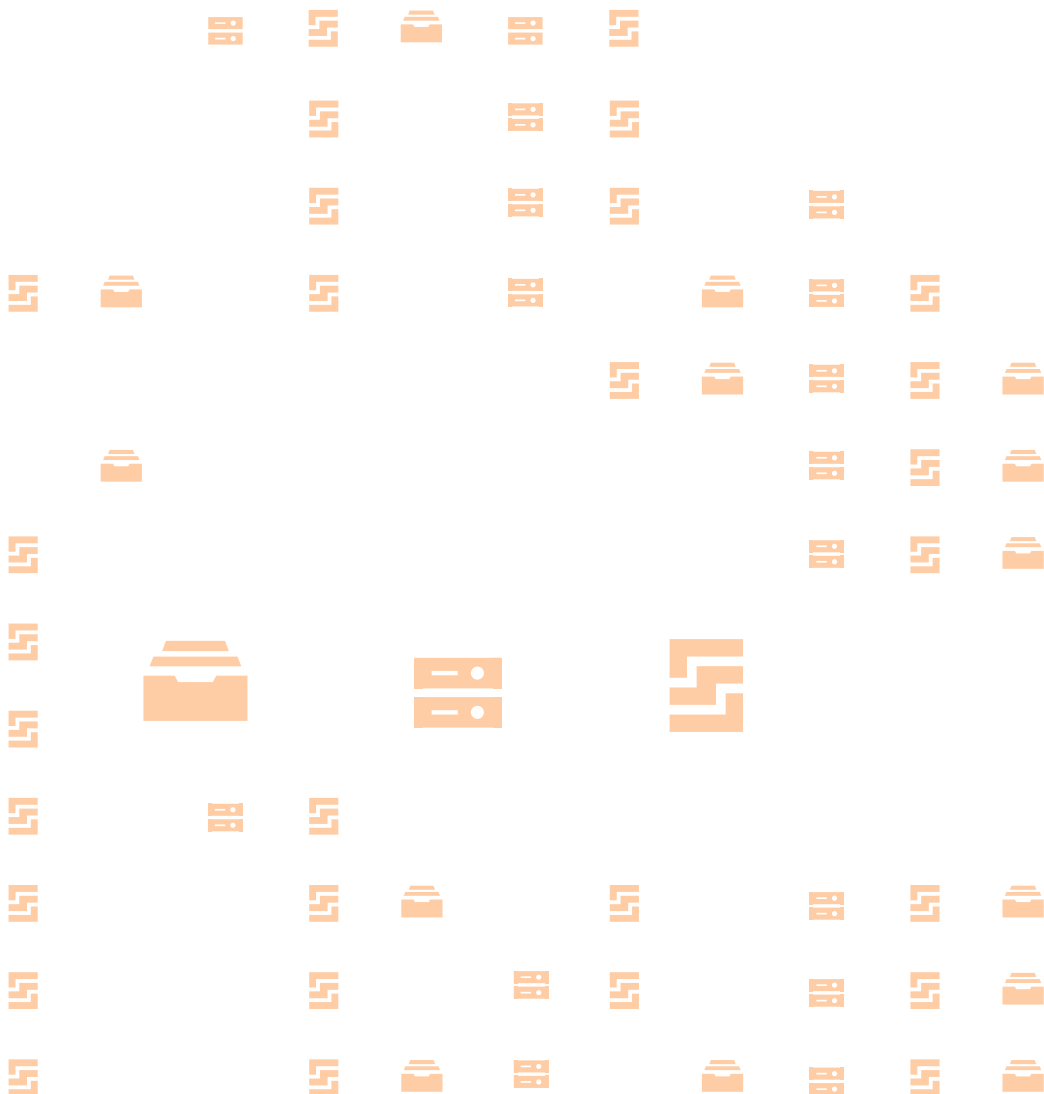
La phase amont est essentielle dans le cadre d'un projet data. C'est elle qui permet de partir sur de bonnes bases. Connaître les besoins de nos clients, identifier et cartographier leur patrimoine data, évaluer la maturité par rapport aux enjeux et aux possibilités techniques ou stratégiques, impose une vraie compréhension du contexte de l'entreprise. Ensuite, et ensuite seulement, nos experts identifient les solutions technologiques qui peuvent être employées pour servir le projet. Il n'existe aucun projet identique. La mise en place d'architecture data peut prendre deux mois, comme elle peut s'étendre sur plusieurs années. Quelle que soit l'ampleur ou la durée du projet data, notre méthode repose sur l'agilité. Tout ou presque peut et doit être

challenge. C'est ainsi que l'on parvient à créer cette alchimie qui permet aux entreprises de s'inscrire dans une vraie culture data. Le secret du succès : une approche par Quick Wins, animée par une démarche ROIste. Chaque succès nourrit le suivant. Aux côtés des CTO et des DSI, auprès du Business, jusqu'aux développeurs, notre mission consiste à concilier et réconcilier en permanence la vision et l'application. Grâce à cette réelle flexibilité technique et intellectuelle qui fait parfaitement écho à la flexibilité native du cloud, nous mettons tout en œuvre pour ouvrir le champ des possibles.



Florent Legras
Data Engineering Manager
pour SFEIR

Conclusion



SFEIR : votre partenaire cloud

Dans une société globalement digitalisée, marquée, quel que soit le secteur d'activité par une accélération du time-to-market et de la concurrence, frappée du sceau de la personnalisation et de l'instantanéité, **passer à la puissance cloud n'est plus un choix. C'est une nécessité.**

Les transformations apportées par près de deux années de crise sanitaire ont encore accentué l'impérieux besoin des entreprises de disposer de toujours **plus de flexibilité dans leur organisation, dans le fonctionnement dans l'édification et l'application de leurs stratégies.**

Fondée autant sur la place centrale de la donnée que sur la robustesse des infrastructures, **cette accélération de l'adoption des usages cloud implique d'être maîtrisée, ajustée, pilotée, accompagnée.**

En associant la puissance, l'adaptabilité et la richesse fonctionnelle de la Google Cloud Platform et le savoir-faire, le goût du sur-mesure des équipes d'experts de SFEIR, les entreprises peuvent envisager le cloud autrement. **Un cloud fonctionnel, pragmatique qui dépasse le cadre des promesses et des incantations pour s'incarner dans une perspective d'efficacité et de performance pour aujourd'hui mais aussi pour demain.**

Gouvernance, souveraineté, sécurité, résilience, orchestration, le cloud inquiète parfois, il intrigue souvent, mais il tient toujours sa promesse de **transformation**. Machines virtuelles, containers, dimension serverless, Google Cloud multiplie également les possibilités d'utiliser de l'intelligence artificielle sans avoir besoin de créer soi-même des algorithmes. Ce potentiel, cette force constituent **une opportunité majeure pour les entreprises de consolider leur croissance, d'asseoir leurs process, de renforcer leur efficacité.**

Chez SFEIR, nous incitons et aidons les entreprises à se mettre en mouvement.

Cela implique des changements organisationnels et technologiques ; il faut devenir plus agile, être en mesure de s'adapter et d'innover en permanence.

Notre implication au quotidien, nos méthodologies d'accompagnement, notre sens des valeurs comme notre engagement de proximité, sont autant d'atouts que nous tenons **à la disposition des entreprises dont nous nous pensons**

avant tout et pour toujours, leurs partenaires !



Remerciements

Nous remercions chaleureusement toutes les personnes qui ont accepté de témoigner dans ce livre blanc : Johan Protin de Hipay, Geoffrey Berard de lesfurets.com, Guillaume Pinot de Renault Digital et Paul Marcombes de Nickel.

Un grand merci aussi aux équipes de Google, en particulier Anthony Cirot, Franck Zerbib et Sébastien Colombo, pour leur éclairage.

Et enfin tous nos remerciements à Seifeddin Mansri, Florent Legras, pour leurs analyses ainsi qu'à l'ensemble des Sfeiriennes et des Sfeiriens et des équipes de WENVISION, qui vivent au quotidien avec ces technologies, les mettent en œuvre et accroissent ainsi jour après jour le niveau global de notre expertise.

Auteurs

Olivier RAFAL

Olivier Rafal est consultant, directeur du conseil de WENVISION. Il réalise des missions de conseil auprès des DSI et dirigeants d'entreprise sur l'optimisation conjointe des stratégies numériques et business.

Olivier Rafal était auparavant VP de teknowlogy Group. Il a été journaliste dans le domaine IT pendant 15 ans, puis analyste et consultant pendant 10 ans, avant de rejoindre SFEIR en 2020.

Seifeddin MANSRI

Seifeddin Mansri est Cloud Engineering Manager chez SFEIR, où il dirige l'activité liée aux infrastructures cloud et DevOps. Il pilote une équipe de 60 experts cloud. Il a obtenu de nombreuses certifications cloud (GCP, AWS, Kubernetes). Il est également Google Cloud Authorized Trainer et Linux Foundation Authorized Instructor expérimenté pour SFEIR Institute.

Florent LEGRAS

Florent Legras est Data Engineering Manager chez SFEIR où il aide nos clients à valoriser leurs données pour améliorer leur prise de décision. Florent est également Google Authorized Trainer pour SFEIR Institute.

Bibliographie

Compute Engine - IaaS → Google Cloud Platform

→ <https://cloud.google.com/compute/>

Load Balancing → Compute Engine Documentation → Google Cloud Platform

→ <https://cloud.google.com/compute/docs/load-balancing/>

Load Balancing and Scaling → Compute Engine Documentation → Google Cloud Platform

→ <https://cloud.google.com/compute/docs/load-balancing-and-autoscaling>

Google Kubernetes Engine → Google Cloud Platform

→ <https://cloud.google.com/kubernetes-engine/>

What are Containers and their benefits → Google Cloud Platform

→ <https://cloud.google.com/containers/>

Kubernetes - Wikiwand

→ <http://www.wikiwand.com/en/Kubernetes>

Accueil → Pokémon GO

→ <https://www.pokemongo.com/fr-fr/>

Google Cloud Platform Blog: Bringing Pokémon GO to life on Google Cloud

→ <https://cloudplatform.googleblog.com/2016/09/bringing-Pokemon-GO-to-life-on-Google-Cloud.html>

App Engine - Platform as a Service → Google Cloud Platform

→ <https://cloud.google.com/appengine/>

The App Engine Standard Environment → App Engine Documentation

→ Google Cloud Platform

→ <https://cloud.google.com/appengine/docs/standard/>

App Engine Flexible Environment → Google App Engine flexible environment docs → Google Cloud Platform

→ <https://cloud.google.com/appengine/docs/flexible/>

Cloud Functions - Environnement sans serveur pour créer et connecter des services Cloud → Google Cloud Platform

→ <https://cloud.google.com/functions/>

Anthos Developer Sandbox - Environnement pour apprendre à développer gratuitement sur Anthos

→ <https://cloud.google.com/blog/topics/anthos/introducing-the-anthos-developer-sandbox>

Bigtable - Service de base de données NoSQL évolutif → Google Cloud Platform

→ <https://cloud.google.com/bigtable/>

NoSQL - Wikiwand

→ <http://www.wikiwand.com/en/NoSQL>

Cloud Storage - Stockage de données en ligne → Google Cloud Platform

→ <https://cloud.google.com/storage/>

Archival Cloud Storage - Data Backup & Disaster Recovery

→ Google Cloud Platform

→ <https://cloud.google.com/storage/archival/>

Cloud SQL - Service de base de données relationnelle MySQL

→ Google Cloud Platform

→ <https://cloud.google.com/sql/>

51

Cloud Spanner → Automatic Sharding with Transactional Consistency at Scale → Google Cloud Platform

→ <https://cloud.google.com/spanner/>

Système de gestion de base de données - Wikiwand

→ http://www.wikiwand.com/fr/Syst%C3%A8me_de_gestion_de_base_de_donn%C3%A9es

NewSQL - Wikiwand

→ <http://www.wikiwand.com/en/NewSQL>

Google Cloud Platform Blog: Inside Cloud Spanner and the CAP Theorem

→ <https://cloudplatform.googleblog.com/2017/02/inside-Cloud-Spanner-and-the-CAP-Theorem.html>

Google Cloud Platform Blog: Behind the scenes with the Dragon Ball Legends GCP backend

→ <https://cloudplatform.googleblog.com/2018/06/Behind-the-scenes-with-the-Dragon-Ball-Legends-GCP-backend.html>

Firestore - Base de données NoSQL sans schéma → Google Cloud Platform

→ <https://cloud.google.com/firestore/>

MapReduce - Wikiwand

→ <http://www.wikiwand.com/fr/MapReduce>

BigQuery - Entrepôt de données pour l'analyse → Google Cloud Platform

→ <https://cloud.google.com/BigQuery/>

Dremel (software) - Wikiwand

→ [http://www.wikiwand.com/en/Dremel_\(software\)](http://www.wikiwand.com/en/Dremel_(software))

52

Cloud Pub/Sub - Intergiciel de messagerie → Google Cloud Platform

→ <https://cloud.google.com/pubsub/>

Cloud Dataflow - Traitement des données par lots et par flux

→ Google Cloud Platform

→ <https://cloud.google.com/dataflow/>

Apache Beam - Wikiwand

→ http://www.wikiwand.com/en/Apache_Beam

Cloud Dataproc - Service Spark et Hadoop géré → Google Cloud Platform

→ Google Cloud Platform

→ <https://cloud.google.com/dataproc/>

Apache Hadoop - Wikiwand

→ http://www.wikiwand.com/en/Apache_Hadoop

Apache Spark - Wikiwand

→ http://www.wikiwand.com/en/Apache_Spark

Pig (programming tool) - Wikiwand

→ [http://www.wikiwand.com/en/Pig_\(programming_tool\)](http://www.wikiwand.com/en/Pig_(programming_tool))

Apache Hive - Wikiwand

→ http://www.wikiwand.com/en/Apache_Hive

Cloud Datalab - Outil interactif d'analyse de données → Google Cloud Platform

→ Google Cloud Platform

→ <https://cloud.google.com/datalab/>

Project Jupyter Home

→ <https://jupyter.org/>

Jupyter and the future of IPython - IPython

→ <https://ipython.org/>

53

Analyses prédictives - Cloud Machine Learning Engine → Google Cloud Platform

→ <https://cloud.google.com/ml-engine/>

TensorFlow

→ <https://www.tensorflow.org/>

API Vision - Analyse du contenu des images → Google Cloud Platform

→ <https://cloud.google.com/vision/>

API Video Intelligence - Analyse du contenu des vidéos → Google Cloud Platform

→ <https://cloud.google.com/video-intelligence/>

API Speech - Reconnaissance vocale → Google Cloud Platform

→ <https://cloud.google.com/speech/>

API Cloud Natural Language → Google Cloud Platform

→ <https://cloud.google.com/natural-language/>

API Cloud Translation - Traduction instantanée → Google Cloud Platform

→ <https://cloud.google.com/translate/>

Produits et services → Google Cloud Platform

→ <https://cloud.google.com/products/>

54

Paris

48, rue Jacques Dulud
92200 Neuilly-sur-Seine
+33 1 41 38 52 00

Luxembourg

5 Place de la Gare
1616 Luxembourg
+352 26 54 471

Belgique

Avenue des Arts 6
1210 Bruxelles
+32(0)2 899 83 70

Lille

74, rue des Arts
59800 Lille
+33 3 66 72 61 32

Bordeaux

Centre les Grands
Hommes
Place des Grands
Hommes
33300 Bordeaux

Strasbourg

Crystal Park
1, avenue de l'Europe
67300 Schiltigheim
+33 3 88 47 04 38

Nantes

Halle 6 Est
40 rue de la Tour
d'Auvergne
44200 Nantes
+33 2 55 59 07 00

[sf≡ir]

Sfeir.com