



Association Nationale des Dirigeants en Sciences de l'Information
www.andsi.fr

Un architecte dans les nuages....

Compte rendu rédigé par ANDSI & Pierre Delort

En bref...

Tarik BOUKHERISSA proposera un exposé dédié au *Cloud* et à la data, et ce, en vue de traiter de l'émergence du nouveau métier d'architecte solution. Pour ce faire, il convoquera notamment son expérience professionnelle et expliquera les raisons pour lesquelles la data constitue un élément central dans son parcours.

L'Association Nationale des Directeurs des Systèmes d'Information organise des débats et en diffuse des comptes-rendus, les idées restant de la seule responsabilité de leurs auteurs. Elle peut également diffuser les commentaires que suscitent ces documents.

Le Cloud : du hardware au software

En préambule, Tarik BOUKHERISSA aborde les chiffres clés du *Cloud*. Les investissements ont dépassé depuis 2020 ceux de l'*on-premise*. En 2020, les dépenses sur le *Cloud* représentaient 125 milliards de dollars tandis qu'elles ressortent à 500 milliards en 2022. Les trois acteurs principaux sont AWS, Azure et GCP.

Un écosystème s'est développé autour du *Cloud*, d'abord avec les *Cloud providers*. Ensuite, et avec les *Cloud native services*, il devient possible d'interconnecter les services du *Cloud* et de déployer des solutions interopérables rapidement. L'interopérabilité et la rapidité correspondent à des éléments essentiels. Aujourd'hui de nombreuses *start-ups* débutent directement sur le *Cloud*. Des entreprises comme Netflix ou Uber ont développé des solutions pour leurs propres besoins et celles-ci ont été souvent en open-source.

Désormais, il est aisé de déployer une solution sur le *Cloud* sans investir sur le long terme. Tout le paiement repose sur « de la facturation à la seconde des services ».

Ensuite, des concepts technologiques ont permis de faire en sorte que le *Cloud* fonctionne et un des principaux, dans le domaine de la data, s'appuie sur le découplage entre la capacité de calcul et la capacité de stockage. Il est possible d'accroître la capacité de stockage indépendamment de la capacité de calcul.

Un parcours autour de la data

Tarik BOUKHERISSA explique avoir étudié à Centrale et réalisé un double diplôme en mathématiques à Cambridge. Il indique ensuite avoir travaillé chez Schlumberger, dans les services pétroliers, en tant qu'ingénieur mécanique au Texas puis dans les domaines de l'électronique et de la data. Il a également occupé un rôle à l'international en tant que chef de projet R&D. D'une manière générale, la data est à la base de son parcours.

En 2015, au moment de la survenue de la crise pétrolière, il a ensuite travaillé dans le domaine de la donnée et plus précisément sur l'IA.

A cette époque, des accords ont été passés entre Google et Schlumberger afin d'accéder au *Cloud* tandis que des *data scientists* ont été recrutés. Il indique ensuite s'être formé au code pour comprendre les nouveaux métiers de la data science.

En 2018, la crise pétrolière constituant toujours une réalité, Tarik s'est concentré sur les métiers de la data et de l'IA. Il précise avoir interviewé une centaine de personnes afin de mieux comprendre ces différents métiers. Il est apparu qu'il était nécessaire de disposer d'une expérience technique, ce qui l'a conduit à réaliser des certifications, notamment sur le *Cloud*, et à réaliser des missions de consultant. Il précise avoir été contacté par des acteurs du *Cloud*, dont l'entreprise pour laquelle il travaille depuis deux ans.

En conclusion, la data est son « nouveau pétrole ».

Du hardware au software : du Capex à l'Opex

Tarik BOUKHERISSA explique avoir travaillé sur des projets dont la durée moyenne en développement s'établissait à une dizaine d'années. Il importait de correctement choisir, au moment du démarrage du projet, les éléments qui devaient être développés. Des tests étaient ensuite conduits avant un passage en production. Ces projets demandaient des investissements lourds en Capex et les compétences techniques des équipes nécessitait une longue montée en compétence. Chez Schlumberger l'*onboarding* s'élevait à trois ans tandis qu'il s'établit à quelques mois typiquement dans les acteurs du *Cloud*. Une telle différence s'explique en raison des technologies open-sources et interopérables sur lesquelles s'appuient les acteurs du cloud.

Int : Spark est un des outils de traitement data principaux aujourd'hui.

TB : Les solutions de data que j'utilise reposent sur des capacités qui se déploient au gré des besoins. Auparavant, la data dans l'entreprise était gérée par des entrepôts de données qui étaient déployés en avance et taillés sur des besoins prédits. Dans ce cadre, Oracle représentait une solution idéale. Désormais et s'agissant de la manière de gérer des solutions data sur le *Cloud*, il n'y a pas d'infrastructure pré-provisionnées, elles le sont au moment de la réalisation des calculs. Ensuite, les infrastructures sont détruites. Spark permet de coordonner des calculs sur ces infrastructures distribuées.

Int : Je m'interroge à propos de ce que fait spécifiquement votre entreprise.

TB : Nos activités s'articulent autour de la donnée. Nous permettons l'ingestion des données, par exemple, des applications – ERP, manufacturing, web app, etc. – afin que celles-ci soient corrélées, et ce, en vue d'une prise de décision (analyse d'une base client, construction de modèles de *Machine Learning*, etc.).

La limite n'est pas toujours la technologie mais parfois l'organisation doit s'adapter. Tarik BOUKHERISSA explique avoir déployé, avec un *data scientist*, un outil permettant de réaliser des calculs relevant du domaine de la physique. Ce travail qui a nécessité quelques mois a consisté à récupérer les données existantes et automatiser la mise en valeur avec un algorithme de *Machine Learning*. Toutefois, le point de blocage a été le déploiement commercial de cet outil car il ne suivait pas les processus standardisés des équipes marketing. Le *product champion* n'a pas pu commercialiser cette solution auprès du client, car il n'avait pas confiance dans les capacités du *Machine Learning*. Il apparaît nécessaire de faire monter en compétence les acteurs sur ces nouveaux outils.

Int : L'innovation de Spark a consisté à faire faire des calculs en mémoire...

TB : Il n'est pas possible de faire de la data science sur de gros volumes de données sans être capable de les transformer rapidement et de façon itérative. C'est dans ce cadre que Spark a été inventé car Hadoop ne permettait pas aux Data Scientists d'explorer de larges volumes de données de façon suffisamment rapide.

Le rôle de l'architecte solution

Tarik BOUKHERISSA explique que son rôle d'architecte solution consiste à travailler avec les clients afin de comprendre leurs besoins pour bâtir une solution. Une partie du travail s'effectue grâce à des cas d'usage spécifiques. Chaque cas d'usage donne lieu au déploiement d'une architecture particulière.

En outre, les interlocuteurs sont moins dans les services IT que dans les métiers. Il s'agit de CDO, de *project owners*, etc. qui sont directement liés au business. L'architecte solution se retrouve dans de nombreux domaines, dont chez les *Cloud providers*. Des certifications sont proposées à propos de cette fonction, ce qui permet à chacun de se former.

Enfin, l'architecte solution couvre les rôles de data engineer, de data scientist et de business analyst.

Débat

Int : Les data scientists sont-ils chez vous ou sont-ils des interlocuteurs chez vos clients ?

TB : Je réponds par l'affirmative à ces deux assertions. En tant qu'architecte solution, je travaille sur une architecture générale d'un cas d'usage. Pour autant, il peut s'avérer parfois nécessaire de réaliser un travail d'approfondissement sur une solution spécifique. Par exemple, je travaille avec un client à propos de la manière de faire de la prédiction sur des ventes de magasin, avec un modèle unique. Dans ce cadre, je m'appuie sur un expert data scientist en interne.

Int : Pouvez-vous préciser qui sont les interlocuteurs mobilisés s'agissant du travail de prévision pour des magasins ?

TB : Il s'agit de data scientists et de data engineers. Ensuite, plusieurs modes d'organisation peuvent être envisagés. Il est possible de positionner les data scientists avec les data engineers dans l'équipe IT. Il est également possible de positionner les data engineers dans l'équipe IT et les data scientists dans l'équipe métier. Dans l'hypothèse où le data scientist est avec le métier dans les BU, les discussions porteront plutôt sur des cas d'usage métiers et quand ils sont dans les équipes IT, plutôt sur la façon de les industrialiser.

Int : Votre solution s'interface-t-elle à la solution du client ?

TB : Oui. J'ajoute que nous sommes 100 % *Cloud*, à l'instar des grands acteurs. Dans le passé, des solutions technologiques étaient « containerisées » et déployées *on-premise*.

Int : Disposez-vous de suffisamment de ressources afin d'assurer le déploiement sur les différents *Cloud* ?

TB : Tout à fait.

Int : Nous avons identifié un autre métier, à savoir le data policy manager. Des règles de rapprochement d'agrégats doivent être déterminées entre ceux qui développent les flux et ceux qui traitent les données.

TB : En effet. Un aspect qui se rapporte au développement du DevOps.

Int : Vous avez évoqué des certifications. Que pouvez-vous nous dire concernant la qualité de la formation ?

TB : De nombreux sites permettent d'apprendre. Néanmoins, les tutoriels ne m'ont pas semblé pertinents. Les certifications et le travail sur des cas usages réels m'ont permis de développer mes compétences.

Int : Vous mentionnez Amazon, Google et Azure. Or, leurs offres ne sont pas complètement comparables.

TB : Nous ne sommes pas en compétition avec ces acteurs. Nous sommes dépendants des solutions que nos clients choisissent de déployer. Les approches sont différentes pour répondre aux besoins clients, AWS

s'appuie sur plus de 400 services pour répondre aux besoins de ses clients tandis que Google propose une solution plus intégrée qui peut demander plus d'adaptation pour un acteur qui ne maîtrise pas le *Cloud*.

Int : Les certifications expirent-elles passé un certain délai ?

TB : Elles expirent généralement au bout de deux ans.

Int : Avez-vous passé la certification Azure de Microsoft ?

TB : Non, et ce, pour des raisons de disponibilité. J'ai passé la certification data chez Google et architecture cloud chez AWS.

Int : Le métier d'architecte solution risque-t-il d'entraîner la disparition de celui d'architecte d'entreprise ?

TB : Non. En revanche, la migration d'architecture monolithique vers des architectures microservices, nécessitent d'avoir un déploiement d'architecture par cas d'usage.

Int : Êtes-vous confrontés à des entreprises qui font du SaaS pour répondre à un besoin du client et qui se demandent pour quelles raisons elles construiraient leur solution sur AWS avec plusieurs services au lieu de recourir à une société ?

TB : Des réflexions sont conduites par le CDO ou l'architecte solution à propos du *make or buy* en fonction du niveau de personnalisation requise, les compétences internes et la criticité pour le business. Le socle transverse sera souvent une architecture en PaaS avec des solutions SaaS sur les couches supérieures.

Int : Je ne perçois pas la différence entre ce que propose un des trois acteurs du *Cloud* et des acteurs spécifiques.

TB : Tous proposent des solutions qui répondent à des besoins clients. Un client a le choix de faire de l'Open Source complet mais devra utiliser ses ressources pour construire une plateforme (IaaS), ou peut s'appuyer sur les services des Cloud Providers et construire le liant entre les différents services (PaaS) ou s'appuyer sur des plateformes plus spécialisées qui ont des niveaux d'optimisation inclus ou finalement des solutions faites sur étagères (SaaS).

Présentation des orateurs

Tarik BOUKHERISSA, Centrale Paris et Cambridge Université, qui a travaillé une vingtaine d'années chez Schlumberger en R&D, production puis transformation digitale et a évolué en Lead Architecte chez un acteur du Cloud.