

## EN MARCHÉ VERS LA BIG DATAVIZ

## AVIS D'EXPERT



Sakil MAMODE ALLY, Directeur Hub Data | Keyrus



Mor Lubranski, Directeur Produit quilliuip | Keyrus

*Le déluge de données annoncé au début des années 2000 est aujourd'hui une réalité pour toutes les organisations. Généralisée dans les environnements de Business Intelligence traditionnels, la visualisation de données (Dataviz) est un maillon décisif pour passer à une large exploitation du potentiel du Big Data à tous les niveaux de l'entreprise. Mais la transition vers la Big Dataviz est plus exigeante qu'on ne l'imagine...*

Le discours sur le Big Data a été longtemps empreint de magie : la digitalisation de l'entreprise et, plus largement, de toute la société avec l'essor des réseaux sociaux et de l'Internet des objets, met chaque entreprise à la tête d'un inépuisable trésor de données où elle pourra puiser en permanence pour trouver de nouveaux leviers d'efficacité et de nouvelles sources de création de valeur. La minimisation des exigences techniques et des expertises nécessaires pour concrétiser cette promesse a laissé plus d'une entreprise dans l'incapacité de faire face à la réalité diluvienne du Big Data et, *a fortiori*, d'exploiter cette masse de données pour gagner en performance et mieux piloter son activité en ayant de celle-ci une vision à 360° en temps réel.

## LA BI ÉVOLUE... LE BIG DATA AUSSI !

Pour extraire l'intelligence de leurs données, les entreprises s'appuient depuis des décennies sur des systèmes de Business Intelligence permettant aux utilisateurs – directions générales et métiers – d'interroger les données et de visualiser les résultats sous forme de rapports, d'indicateurs et de graphiques, sans être confrontés à la complexité sous-jacente des données. Historiquement, ces systèmes ne prenaient en compte que les données structurées générées par les applications de l'entreprise, un Data warehouse géré par la DSI assurant généralement la centralisation et la normalisation des données pour qu'elles soient « requêttables ». Aujourd'hui, il s'agit de prendre en compte des données de provenance et de formats extrêmement divers – structurées, non structurées, images, vidéos, sons, e-mails, publications et commentaires sur les réseaux sociaux, données générées par les capteurs des objets connectés, etc. – dont, de plus, la vitesse de production et les volumétries sont sans commune mesure avec ceux pour lesquels les environnements de BI traditionnels ont été conçus.

C'est l'objet même du Big Data de rendre exploitable l'ensemble des données que l'entreprise génère et peut collecter. Si un réceptacle tel qu'une base Hadoop – typiquement un lac de données (Data Lake) – peut centraliser les données de tous types sans limite quantitative, la difficulté est d'étendre la Dataviz à ce type d'environnement. La visualisation est en effet un maillon clé de la chaîne puisque c'est elle qui permet aux utilisateurs d'accéder à l'intelligence des données et, de ce fait, de se concentrer sur l'analyse et les actions à valeur ajoutée plutôt que sur la collecte, la consolidation et la vérification des données.

## NE PAS SOUS-ESTIMER LES PRÉ-REQUIS TECHNIQUES

Il est illusoire de croire qu'il suffit de « pluguer » un outil de Data Visualisation sur un Data Lake pour étendre les bénéfices

du Big Data à toute l'entreprise et à toutes les catégories d'utilisateurs. Si certains éditeurs proposent des solutions de Dataviz « Big Data ready » avec des connecteurs pour Hadoop et d'autres types de bases, les performances – notamment les vitesses d'exploration et de restitution – sont rarement au rendez-vous pour des raisons de volumétrie et de structure des données. Pour que les performances soient acceptables, un travail de structuration et d'optimisation est indispensable, même avec les outils de Dataviz utilisant la technique du In-Memory<sup>1</sup>.

Du fait de la diversité des données, les Data Lakes ne sont jamais organisés sous forme de bases de données relationnelles. Ce sont généralement des bases de type « NoSQL » n'obéissant pas au schéma en étoile/flocon<sup>2</sup> qui prévaut dans les environnements de BI traditionnels. Il existe cependant des stratégies pour interroger les données qui se trouvent dans un Data Lake avec des outils de type SQL, même si les données ne sont pas stockées sous forme relationnelle. C'est typiquement ce que permettent de faire les outils on-Hadoop, au-dessus desquels on peut facilement brancher un outil de Dataviz classique. L'autre possibilité est d'exposer les données non pas en SQL, mais par le biais de web services. Les données appelées sont alors restituées sous une forme graphique en JavaScript dans une page web ou via un portail.

Quelle que soit l'option, la préparation des données est indispensable – sachant que, dans un environnement Big Data, le caractère non figé du schéma des données rend la constitution d'une couche sémantique beaucoup moins aisée que dans un environnement de BI classique. Pour pallier l'impossible recours à un schéma en étoile dans un Data Lake, on dénormalise les données. Cette opération consiste à créer une table unique contenant toutes les données. Cette table est souvent extrêmement volumineuse parce que les dimensions qui étaient auparavant mutualisées dans un modèle en étoile sont obligatoirement dupliquées, ce qui pose inévitablement deux types de problèmes :

- **Des problèmes de cohérence des données**, obligeant à avoir une politique de mise à jour et d'alimentation du Data Lake extrêmement rigoureuse.
- **Des problèmes de performance** au niveau des requêtes et de la restitution, obligeant à pré-définir les périmètres et la granularité des requêtes de façon à constituer des sous-ensembles de données plus facilement exploitables par la Dataviz.

Des optimisations côté serveur sont également nécessaires pour rendre les requêtes SQL on-Hadoop extrêmement efficaces.

<sup>1</sup> Technique consistant à monter et garder les informations d'une base de données en mémoire vive afin d'accélérer les temps d'accès et de réponse.

<sup>2</sup> Base structurée sous forme de tables reliées entre elles et où chaque table est reliée à des tables de dimension correspondant aux axes selon lesquels les faits peuvent être explorés et analysés.

Quels que soient le modèle des données et la volumétrie, c'est un travail d'expert qui fait intervenir des outils intermédiaires, des mises en cache et des indexations. L'objectif de ce travail de structuration et d'optimisation est que la navigation reste agréable dans l'outil de Dataviz, ce qui est une condition *sine qua non* pour les utilisateurs.

Conscient des enjeux autour de la performance et de la cohérence des données, Keyrus et sa filiale Vision.bi ont créé la plate-forme quillip afin d'aider les entreprises à contrôler efficacement la qualité de leurs données et les aider dans leurs prises de décisions.

quillip permet également d'améliorer la gouvernance de toutes les sources de données pour garantir la cohérence et pertinence de l'ensemble des tableaux de bord permettant de piloter l'entreprise.

## PARTIR DES USAGES & DE LA MATURITÉ DES UTILISATEURS

Le but de la visualisation des données est de permettre aux utilisateurs de détecter facilement, et surtout plus rapidement, une anomalie ou une problématique dans leur activité, de façon à mettre en œuvre des actions pour y remédier. Des orientations d'action peuvent d'ailleurs être suggérées par le biais d'une couche d'analyse intelligente. Mais, avec ou sans ce type de recommandation, c'est à l'expérience utilisateur proposée dans l'interface de restitution et à la pertinence de représentation graphique qu'il faut veiller si l'on veut vraiment capitaliser sur la dimension visuelle.

Il faut pour cela connaître un certain nombre de règles et de bonnes pratiques. On sait par exemple que, sur un écran, le regard va de haut en bas et de gauche à droite, ce qui est

déterminant pour positionner les différentes informations. On sait aussi qu'au-delà de 4 ou 5 indicateurs sur une même page, l'utilisateur est obligé de faire un effort d'attention supplémentaire. Enfin, tous les utilisateurs n'étant pas des experts en sémiologie graphique et les outils offrant de plus en plus de possibilités, il est souvent nécessaire de les guider dans le choix des représentations visuelles pour qu'elles soient vraiment adaptées à leurs indicateurs. Par exemple, un graphe à bulles sera la forme la plus efficace pour analyser la répartition de la masse salariale par genre et par niveau de salaire. En revanche, ce type de graphe ne sera pas approprié pour présenter les résultats d'un benchmark. Seule l'expérience permet de le savoir.

S'appuyer sur les règles et les bonnes pratiques des spécialistes de l'UX et de la visualisation des données permet de construire pour chaque catégorie d'utilisateurs des visualisations simples, épurées et faciles d'usage parce qu'elles mettent en valeur les indicateurs importants pour son activité. De même qu'il doit être protégé de la complexité et de l'hétérogénéité des données sous-jacentes par le travail de structuration des données et d'optimisation des requêtes réalisé par les ingénieurs de données, l'utilisateur final ne doit pas avoir d'effort à faire pour comprendre ce qu'il a devant les yeux et en tirer parti pour améliorer son activité. Une approche couplant ces deux dimensions – ingénierie des données et expertise UX – permet d'élever le niveau de data-alphabétisation des collaborateurs de l'entreprise et de démocratiser les usages du Big Data. Négliger l'une ou l'autre, c'est condamner le Data Lake, dont se dotent de plus en plus d'entreprises, à rester longtemps le « bac à sable » des seuls Data Scientists.

S.M.A.

## À PROPOS DES AUTEURS

### Sakil MAMODE ALLY

Avec plus de 17 ans d'expérience dans les domaines de la Data Intelligence où il a notamment exercé des responsabilités de développement d'activités BI Data Reporting mais également de business development, Sakil a construit son style de management autour des valeurs humaines et technologiques qu'il met au service des clients et des collaborateurs.

Il rejoint **Keyrus** comme responsable de la practice Smart Data à Paris pour répondre aux enjeux/besoins des clients de **Keyrus** autour des problématiques de Dataviz et de Dataprep.

Avec son appétence au commerce et son background opérationnel, il a ensuite accompagné au développement du secteur Retail & CPG, tant sur les plans business et opérationnel que stratégique, avant de prendre la direction du Hub Data. Il a notamment pour mission d'accompagner **Keyrus**, autant les Sales, les Opérationnelles que l'avant-vente, sur des sujets autour de la Data.

### Mor LUBRANSKI

Depuis plus de 8 ans, Mor dirige les équipes R&D et produits dans divers secteurs.

Il a débuté sa carrière en tant qu'ingénieur de données et a été impliqué dans la conception et le développement pratique de solutions complexes Big Data. Sa capacité à traduire les besoins des clients en exigences techniques lui confère une grande capacité à mener des projets complexes et réussis.

Sa vaste expérience dans les domaines de la technologie et des affaires, ainsi que ses compétences en présentation et sa compréhension du marché, lui permettent de diriger le produit quillip de **Keyrus**.

## À PROPOS DE KEYRUS

### Keyrus, créateur de valeur à l'ère de la Data et du Digital

Acteur international du conseil et des technologies, spécialiste de la Data et du Digital, **Keyrus** a pour mission d'aider les entreprises à tirer profit du paradigme de la Donnée et du Numérique pour accroître leur performance, faciliter et accélérer leur transformation et générer de nouveaux leviers de croissance, et de compétitivité.

Plaçant l'innovation au cœur de sa stratégie, **Keyrus** développe une proposition de valeur unique sur le marché autour d'une offre novatrice qui s'appuie sur la combinaison de trois expertises majeures et convergentes :

- **Data Intelligence**

Data Science – Intelligence Artificielle – Big Data & Cloud Analytics – Business Intelligence – EIM – CPM/EPM

- **Digital Experience**

Innovation & Stratégie Digitale – Marketing Digital – DMP & CRM – Commerce Digital – Performance Digitale – User Experience

- **Conseil en Management & Transformation**

Stratégie & Innovation – Transformation Digitale – Pilotage de la Performance – Accompagnement des Projets

Présent dans une quinzaine de pays et sur 4 continents, le Groupe **Keyrus** emploie plus de 3 000 collaborateurs.

**Keyrus** est coté sur le compartiment C de l'Eurolist d'Euronext Paris

(Compartiment C/Small caps - Code ISIN: FR0004029411 – Reuters KEYR.PA – Bloomberg : KEY:FP)

Plus d'informations sur : [www.keyrus.fr](http://www.keyrus.fr)