

MOVING TO BIG DATAVIZ

EXPERT OPINION



Sakil MAMODE ALLY, Data Hub Director | Keyrus



Mor Lubranski, Director of Product quillip | Keyrus

The onslaught of data that was predicted in the early 2000s has now become a reality for every organisation. Data visualisation (or dataviz for short), a typical feature of traditional business intelligence environments, is the key to getting the most out of big data at every level of a company. But transitioning to "big dataviz" is more complicated than you might imagine.

There has been a lot of magical thinking surrounding big data: as businesses and, more broadly, society as a whole have gone digital with the rise of social media and the internet of things, every single company is now sitting on an inexhaustible trove of data that it can draw on continuously to discover new ways to drive efficiency and create value. Reductions in the technical requirements and expertise needed to make this promise a reality have left more than one company unable to deal with the flood of big data and, as a result, unable to leverage this trove of data to establish a 360° vision of their business in real time, a vision that would help them improve performance and better manage their business lines.

BI IS EVOLVING...AND SO IS BIG DATA!

To extract intelligence from their data, companies have for decades now been relying on business intelligence systems that help users – executives and department management teams – analyse their data and view the results as reports, indicators and graphs, without having to deal with the underlying complexity of the data. In the past, these systems only accommodated the structured data generated by the company's applications, while a data warehouse managed by the information systems department centralised and standardised the data to make it searchable. These days, systems must take in data from many different sources and in a wide range of formats – structured, unstructured, images, videos, audio, emails, social media posts and comments, data generated by sensors on smart objects, etc. – and all produced at a rate and in quantities that traditional BI environments were simply not built to handle.

The whole point of big data is to make it possible to use all of the data that a company generates and collects. While a container like a Hadoop framework – typically a data lake – can centralise all kinds of data, the challenge is how to extend dataviz to this type of environment. Visualisation is a key step in the process, as it enables users to access data intelligence and, as a result, concentrate on analysis and value-added actions rather than collecting, consolidating and verifying the data.

DO NOT UNDERESTIMATE TECHNICAL REQUIREMENTS

It takes more than just plugging a data visualisation tool into a data lake if you want the entire company and all its different user categories to reap the benefits of big data. While some publishers offer "big data ready" dataviz solutions with connectors for Hadoop and other types of frameworks, these rarely perform well – especially when it comes to mining

and output – due to the volume and structure of the data. To achieve the kind of performance you need, the data must be structured and optimised, even with dataviz tools that use In-Memory¹ techniques.

Due to the diverse nature of the data involved, data lakes are never organised into relational databases. These are usually "NoSQL" databases that do not utilise the star/snowflake² schemas commonly used in traditional BI environments. There are strategies for using SQL tools to search the data in a data lake, despite the fact that said data is not stored in a relational format. On-Hadoop tools typically allow you to do this, and you could plug a traditional dataviz tool into one of these. The other possibility is to present data using web services rather than SQL. The collected data is presented in the form of a JavaScript graphic on a web page or a portal.

No matter which option you choose, you must prepare the data beforehand, keeping in mind that in the big data environment, the shifting nature of the data schema makes it much more difficult to create a semantic layer than in a traditional BI environment. To make up for the fact that star schemas cannot be used with data lakes, the data is denormalised. This means creating a single table containing all the data. This is an extremely large table because the dimensions that were previously pooled in a star model have now been duplicated, which inevitably leads to two types of problems:

- **Data consistency problems**, requiring a strict policy for updating and adding to the data lake.
- **Performance problems** when it comes to searching and output, necessitating predefined scopes and granular searches in order to create data subsets that more readily lend themselves to dataviz.

Optimisations on the server side are also needed for highly efficient on-Hadoop SQL searches. No matter what data models and volumes are involved, this work requires a great deal of expertise in the use of intermediary tools, caching and indexing. The goal of this structuring and optimising is to make browsing easier in the dataviz tool, which is a must for users.

Well-versed in the challenges involved in data performance and consistency, Keyrus and its subsidiary Vision.bi have created the quillip platform to help companies effectively control the quality of their data and assist them in their decision-making.

quillip also improves the governance of all data sources to ensure the consistency and relevance of all business dashboards to drive the company.

¹ This technique consists of retrieving information from a database and storing it in active memory to improve access and response times.
² A database organised into interconnected tables, where each table is connected to dimension tables that correspond to the ways in which facts can be explored and analysed.

WORKING WITH USER HABITS AND EXPERIENCE

The goal of data visualisation is to help users quickly and easily identify irregularities or problems in their business lines so that they can take action to correct them. An intelligent analysis layer may even be able to point to potential plans of action. With or without these kinds of recommendations, the user experience offered by the output interface and the relevance of the visuals are the keys to truly capitalising on the visual aspect.

For this, you must be familiar with some rules and best practices. For instance, we know that when looking at a screen, people's eyes move from top to bottom and left to right, which determines where different information should be placed. We also know that putting more than four or five indicators on a page means that users will have to work harder to pay attention to them. Users are not all proficient in graphic semiology, and as tools come out with more and more features, it is often necessary to offer guidance in choosing the visual representations that are best suited to the indicators in question. For instance, a bubble chart is the most effective way to analyse the breakdown of salaries by gender and wage level. On the other hand, this kind of graph would not be appropriate for presenting results for a benchmark. This kind of knowledge is gained from experience.

By leveraging the rules and best practices from UX and data visualisation experts, you can build simple, streamlined visuals for each category of users, who will find them easy to use because they highlight the important indicators for their business line. End users must be protected from the complexity and heterogeneity of the underlying data through the data structuring and search optimisation work performed by data engineers, and it should be easy for these users to understand what they are seeing and use it to improve their business lines. By combining these two aspects – data engineering and UX expertise – you can improve the data literacy of company employees and help more of them become proficient in big data. By neglecting one or the other, you are condemning the data lake, which more and more companies are acquiring, to remain a sandbox that only data scientists can play in.

S.M.A.

M.L.

ABOUT THE AUTHORS

Sakil MAMODE ALLY

In his 17+ years in the field of data intelligence, working in both BI data reporting development and business development, Sakil developed a management style that incorporates interpersonal and technological expertise, which he leverages to help his clients and colleagues.

He started at **Keyrus** in 2016 as a smart data manager in Paris, meeting the needs and overcoming the challenges of Keyrus clients in the areas of dataviz and dataprep.

Making good use of his business sense and his operational background, he went on to help develop the Retail & CPG sector, working on business and operational plans as well as strategy, until he was promoted to Data Hub director. His main duties include helping **Keyrus France** manage, maintain and develop data skills.

Mor LUBRANSKI

For more than 8 years now, Mor has been led both R&D and product teams in various industries.

He started his career as a Data Engineer, and was involved in the design and hands-on developments of complex end-to-end Big-Data solutions. His ability to translate customer needs into tech requirements grant him great ability to lead successful and complex projects.

His wide experience in both tech and business fields, along with strong presentation skills and market understanding, qualify him to lead the quillip product by **Keyrus**.

ABOUT KEYRUS

Keyrus, creator of value in the era of Data and Digital

An international player in consulting and technologies and a specialist in Data and Digital, **Keyrus** is dedicated to helping enterprises take advantage of the Data and Digital paradigm to enhance their performance, facilitate and accelerate their transformation, and generate new drivers of growth and competitiveness.

Placing innovation at the heart of its strategy, **Keyrus** is developing a value proposition that is unique in the market and centred around an innovative offering founded upon a combination of three major and convergent areas of expertise:

- **Data Intelligence**

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

- **Digital Experience**

Innovation & Digital Strategy – Digital Marketing – DMP & CRM – Digital Commerce – Digital Performance – User Experience

- **Management & Transformation Consulting**

Strategy & Innovation – Digital Transformation – Performance Management – Project Support

Present in some fifteen countries on four continents, the **Keyrus** Group has more than 3,000 employees.

Keyrus is quoted in compartment C of the Eurolist of Euronext Paris

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

Further information at: www.keyrus.com