WHO CARES HOW BIG THE DATA IS? IT DOESN'T REALLY MATTER

Only a couple of years ago, while everyone was talking about big data, the impression was that very few people knew what it meant. Different people had different interpretations of the term big data. Yes, most everyone in the industry could name the Volume – Variety – Velocity trilogy but the market was still immature, and the majority of big data projects were in their infancy.

Only a few years have passed and while big data is now considered mainstream, and the big data market is flourishing, we also hear more and more that "Big data is dead". A bit confusing, I admit.

Let's try and clear this confusion by understanding what has changed in the big data market and key trends driving those changes including cloud, Machine Learning and Artificial Intelligence (AI).

The cloud. Who doesn't have something "moving to the cloud"? With that in mind, many of the new big data applications are being developed to run in the cloud. Given some of the clear advantages that cloud can offer for managing and "crunching" a lot of data, it has become the platform of choice for many big data projects. The major cloud vendors (AWS, Azure and Google Cloud) are offering a comprehensive, rich list of big data related services; they'll move your data, store your data, process your data and, of course, analyze your data.

Machine Learning and AI. Expecting systems to self-improve means learning from experience and that requires being able to use large amounts of data. Machine Learning algorithms rely on data – on its quantity but nonetheless, also on its quality.

Nobody is questioning that there is an enormous amount of data available and more being collected every day, hour, minute, even by the microsecond. And there are [reports](#) out there that estimate how much data is produced today and how much will be produced by some date years down the road. How big is big then? Does it really matter?

Think about it this way. Organizations are sitting on goldmines of data and they care only about one thing – how to make the most out of this data and how to provide insights to the business, not only to remain competitive and increase profits, but also to thrive, not just survive, in the market.

Organizations implementing big data need to adopt new technologies and new platforms such as Hadoop, Spark, Mesos and of course the multiple cloud vendor provided solutions. They will be ingesting high volumes of data from multiple sources and processing this data, before making it available for analytics.

And finally, organizations use many, many tools, lots of time and much of their valuable talent to develop scripts to integrate it all. Yet, integrations are not easy and manual scripting doesn't always easily deliver scalable results. This is where many organizations are struggling:

- How do I successfully orchestrate my big data workflows?
- How do I ensure my SLAs are met?
- How do I have my data engineers focused on actionable data rather than spending precious time on operational plumbing?

Here are a few tips for orchestrating your big data workflows:

- **Start early** – Orchestration of the data pipelines is critical to the success of the projects and delaying it until the application is ready to move to production, may result in unnecessary errors and delivery delays.
- **Think big** - You will be using multiple technologies for the various steps of your big data implementation and the technologies will change often. Consider having an application workflow orchestration solution that can cope with this diverse data.
- **Avoid automation silos** – you want to be able to get end-to-end visibility of the data pipeline across disparate sources of data
- **Developers are expensive** – get them to focus on the data itself rather than building scripts for operational plumbing

Want to learn more? Check out [Control-M for Big Data](#)