

Challenges working with big data

Unified Data Analytics emerged as a way to help organizations struggling with working with big data. In this video, we'll review some of those common challenges.

Challenge number one - big data is inherently complex to work with. This is because big data differs from the traditional data that many of us are used to working with - it is coming in, in massive volumes, faster than ever before, and in a variety of new formats.

As data practitioners work to design their organization's big data infrastructure, they often ask and need to answer questions like:

- Where/how will we store our big data?
- How can we process batch and stream data?
- How can we use different types of data together in our analyses (unstructured vs. structured data)?
- How can we keep track of all of the work we're doing on our big data?

As you can imagine, there are many ways that an organization can set up big data infrastructure -- getting it right is no easy task.

Siloed roles lead to organizational inefficiencies

Even once a big data infrastructure is set in place, many organizations suffer from the challenges of having siloed functional roles for individuals on their data science teams. As we mentioned, working with big data is complicated, and without team collaboration and transparency on big data workflows, inefficiencies can ripple through an organization. For example, it is not uncommon for a data scientist to build and train a machine learning model in a vacuum on their own computer, with little to no visibility to related work being done by, for example, the data engineer preparing that data for them, or the data analysts who might be using results from their experiments to produce dashboards.

Protecting customers and their data is difficult

According to Gartner, 80% of organizations will fail to develop a consolidated data security policy. This leaves them and their data vulnerable to security breaches.

Think about the ramifications of a security breach. Beyond just the immediate monetary cost, there is a long-lasting loss in customer trust and company reputation. If you've ever been a customer of a company that has suffered a security breach, you know first-hand how long it can take to rebuild trust.

In addition to protecting data from leaking out, organizations must also make sure they're compliant with data protection regulations like GDPR (European Union's General Data Protection Regulation) and HIPAA (Health Insurance Portability and Accountability Act), or that they have required certifications to run their businesses. And, there can be hefty penalties involved if they are not compliant.

Traditional architectures for working with big data need improvement

Not all architectural patterns work well for big data management and analytics. For example, older architectural patterns might struggle to simultaneously process batch and streaming data. This means that anytime a data engineer needs to validate, reprocess, or update batch and streaming data, they might deal with:

- Complexities from having to manage separate code bases and workflows
- Difficulties merging/reconciling data for one single source of truth

Aside from this, using older architectural patterns can make it difficult to guarantee data availability for everyone (who can access it and when), implement security controls or know which data can be trusted.

In summary, it means that data teams end up spending more time processing and managing data than actually working with it to derive insights.

The emergence of unified data analytics stemmed from helping organizations overcome these challenges.