

---

# Cloudera in the Cloud: The Future of Cloud-based Analytics

Cloudera Enterprise

CLOUD

## The Challenges of Cloud-Based Analytics Today

Today, public cloud is a compelling proposition for businesses and government organizations seeking to be more agile. Increasingly, self-service is seen as the most effective way to scale user access to data for analytics and operations. Cloud elasticity, combined with the right user applications, can reduce the friction of waiting for IT to fulfill requests and provision resources and data. As such, we're seeing cloud-based big data growing exponentially for Cloudera customers and across the market as a whole.

For these organizations, it's essential that they are also managing the associated risks and costs of going to a cloud environment. One of the most common fears is vendor lock-in and becoming too dependent on a single cloud service provider, especially with the ever-evolving nature of the market's competitive landscape. Many cloud providers have their own "house brand" offerings, but often these were cobbled together through a mish-mash of OEM deals, acquisitions, or open source code, and thus heavily modified as to have been permanently forked. The result is sub-optimal, fragmented, and interoperability is often very limited. The vast volumes of big data in these environments imposes some challenges, by magnifying the long-term costs associated with renting cloud storage and compute resources. Understanding the technical and financial implications is crucial before becoming locked into a particular cloud environment.

Further, businesses and governments need to ensure proper control of sensitive data. Any particular cloud provider may offer a range of data management services, often composed of a miscellany of acquired and home-grown products— some ported to run in cloud, some native. This creates complications, as fragmented offerings lead to data silos and having to manage multiple distinct tools for security and governance. Simply finding what data lives where, and who has access, becomes hard if there is no common control plane for the data.

Hybrid- and multi-cloud environments are also becoming the norm for many companies. These strategies carry the potential to help find the right mix of qualities, while balancing costs and preserving negotiation power. Some data sets and applications can be kept within the walls of the data center or country borders, while other applications can live in whichever cloud is most optimal for its own requirements. The solution is maintaining a standard data management approach across the variant data platforms.

## How Cloudera Delivers a Better Cloud Experience

A noteworthy point is that Cloudera complements popular cloud services, such as Amazon Web Services (AWS) and Microsoft Azure. While cloud services do provide useful resources — such as compute instances and object storage on demand — Cloudera offers the unified platform to organize, process, analyze, and store data at large scale... anywhere. The same enterprise capabilities delivered on premises are optimized for cloud environments, taking advantage of the elasticity and delivering the self-service desired. Therefore, having a common, open standards-based data platform offers flexibility, avoiding lock-in to any proprietary service. As cloud vendors continue to compete with ever lower costs, data and analytics remain portable. This allows businesses to arbitrage the costs, and run workloads where they make the most sense.

There are two common ways to approach cloud services with Cloudera. The first is to "bring your own" Cloudera licenses to the cloud of your choice. Popular products like Cloudera Data Science Workbench, Analytic DB, Operational DB, and Essentials can be run natively in cloud environments, deploying the cloud service's resources much the same way they would on-premises. Cloudera Director facilitates provisioning and management of clusters in the cloud. Simply choose the AWS or Azure compute instances, class of storage, and go.

The newer, second option is to leverage Cloudera Altus “data engineering-as-a-service,” which simplifies the experience by eliminating even the initial steps of installing software or configuring clusters. Instead, users can immediately begin to build data pipelines and run jobs for their applications. Either way, the Cloudera platform promotes a focus on driving business results, not troubleshooting the inevitable challenges that come with ensuring compatibility amongst pieced-together components. Cloudera products cooperate universally and smoothly across your choice of environments and applications.

A number of notable advantages come with a unified, all-encompassing data platform, which Cloudera calls the Shared Data Experience (SDX.) A shared catalog defines what data is available and how it should be used, independent of where it lives. Security policies are set once and implemented uniformly upon data and users anywhere. Governance is handled evenly across all environments and applications, both for technical and business metadata. Self-service becomes easy with a unified view of who has rights to what types of data. While these goals sound simple, they are all but impossible to deliver if the analytics services aren’t designed holistically from the beginning, a common problem amongst other cloud data and analytics services. Retrofitting functionality across products adds complexity and risk, and often has limited to no capability to ensure enterprise standards are met.

### Three Essential Goals for Cloud Analytics

As the market moves toward cloud-based big data and analytics, three qualities emerge as vital for success. While many services will get some traction without meeting all three goals, they will also disappoint users and cause perpetual headaches for IT. At Cloudera, we see these undisputable attributes to be:

**Easy** – Certainly no one goes out looking for a harder way to do their job. Cloud IaaS, powered by Cloudera Director, facilitates resource self-service provisioning, eliminating the hassles of procurement and deployment on-premises. Cloud PaaS, on the Cloudera Altus platform, takes this a step further and allows users to focus directly on building data pipelines, training machine learning models, developing analytics applications — all the value creation efforts, vs the infrastructure operations. End-user focused tools accelerate daily tasks like job submission, performance tuning, and workload analytics. Intelligent defaults and built-in logic eliminate much of the guess work. The net result is much improved productivity for data engineers, data scientists, and analysts.

**Unified** – Conceptually, cloud sounds like a single place to host diverse, data-intensive functions. In practice, many services end up reproducing the silos that existed on-premises. A far superior approach is to truly consolidate data in one, persistent object store, and then bring different applications and workloads to bear against that set. Fragmented services lead to fragmented controls, when in actuality, what people really want is a common platform and control plane to manage everything, even across hybrid- and multi-cloud deployments. An advantageous side benefit of a unified approach is lower total cost of ownership, stemming from eliminating redundant data storage, leveraging transient compute, and simplifying management overhead. Cloudera Enterprise is a unified platform serving a diverse range of analytics workloads, all in one.

**Enterprise-grade** – Perhaps this goes without saying, but enterprises need cloud to be every bit as robust as their traditional approaches. To be acceptable, cloud analytics platforms must meet or exceed corporate requirements around security, governance, and management. Central control for role-based access, authentication, authorization, encryption, keys – these are all must haves to pass audits and show compliance. The ability to discover and define

metadata definitions for the business context of data is a critical enabler for self-service functions, especially across multi-tenant, multi-cluster, and even hybrid- and multi-cloud environments. Not least, businesses will want a data platform that has been proven out in the market by their most demanding peers. Cloudera Enterprise is that platform, curated and ready for your mission critical analytics applications.

What we repetitively see is that these goals are harder to meet than vendor marketing might suggest. Make sure any cloud-based analytics service meets these criteria. Ultimately, Cloudera's goal is to efficiently deliver machine learning and advanced analytics capabilities that leverage the power of big data. These qualities are at the heart of our differentiation. Strategize on how the right approach and right technology can make data in the cloud your most important asset. We're here to help.