CLOUDERA

Managing Data and Analytics in the Public Cloud

Do Cloud Better with an Enterprise Data Cloud

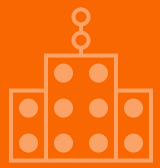
Table of Contents

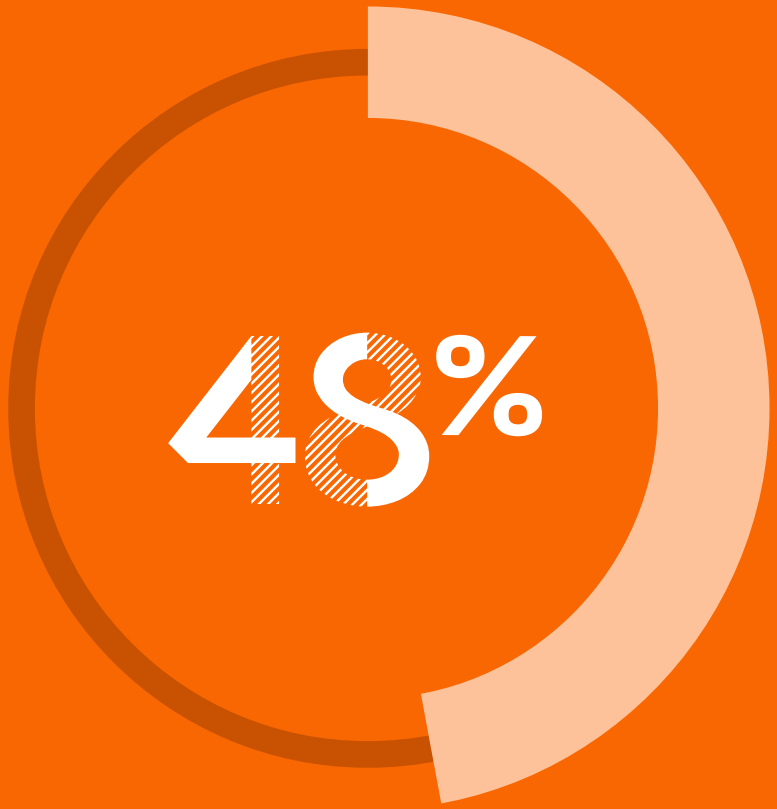
Introduction

Organizations today recognize the importance of unlocking insights from their data to help them run their business better in the present and also shape it more strategically in the future, and have significantly increased their use of analytics to do so.[1]

Most companies find themselves managing data and analytics in one or more public clouds. Cloud may have been a conscious choice, the result of mergers and acquisitions or arrived at through business units' ad hoc analytics projects. Many organizations have been disappointed by unmet expectations of flexibility, cost reduction and performance. They have instead encountered cloud charges spiraling out of control... vendor lock-in... difficulties getting their data back on-premises... and unexpected security and governance risks.

A recent Harvard Business Review Pulse Survey of 185 global executives across multiple industries shows data is managed across a wide variety of deployments, ranging from on-premises to private, hybrid, single and multi-cloud.





Few organizations manage a sizable chunk of their data in public cloud: 48% manage no more than 25% of their data.

A recent Harvard Business Review Pulse Survey of 185 global executives across multiple industries shows data is managed across a wide variety of deployments, ranging from on-premises to private, hybrid, single and multi-cloud. Few organizations manage a sizable chunk of their data in public cloud: only 48% manage no more than 25% of their data. Problems with legacy applications (49%), security (42%), and cost (36%) remain the biggest issues with cloud service providers are the reasons that organizations are not making greater use of the public cloud today.[2]

All the while, shadow IT grows because enterprise IT can't keep up with the speed of business in providing access to data and analytics. Business users then find their own creative ways to get the answers they need via cloud applications. SaaS now makes it quick and easy for users to deploy their own analytics—introducing further cost and risk.

However, when it comes to data and analytics, is it possible to do cloud the right way so that it provides the analytics business users crave while being cost-effective, flexible, and secure? Organizations often believe they can only optimize two of those three, so they compromise, but is it possible to have it all?

Fortunately, today, the answer is *absolutely* yes.

Biggest Issues with Cloud Service Providers for Organizations using Public Cloud

The Truth About Cloud for Analytics

Organizations have sky-high expectations of public cloud. These expectations—that cloud will be cheaper and better than data center deployments—are rarely met.

Do not be deceived; you have to be incredibly careful and critical in your decisions around the cloud to harness its power and turn that into value and advantage for your organization.

One of the biggest stumbling blocks for organizations in terms of cloud is that they often let their cloud strategies impact—that is, inhibit—their data strategy, effectively putting the cart before the horse. Many organizations have landed on a hybrid cloud strategy by default, with their infrastructure, data and applications housed on a relatively unplanned mix of data centers and public clouds, creating new data silos and tremendous challenges managing, accessing, securing and governing—much less leveraging—their data.

Leading with an Enterprise Data Strategy

Doing cloud right means making sure your organization has a strong enterprise data strategy that informs and drives your hybrid cloud strategy, instead of the other way around, as discussed in the Cloudera white paper Why a Successful Hybrid Cloud Strategy Requires an Enterprise Data Strategy.[3]

Why? Because data is a strategic asset, and cloud is a delivery model—a robust and scalable one, to be sure, but just a delivery model at the end of the day. Your cloud strategy could hamper your efforts to manage, access, secure and derive good insights from your data.

A well-planned data strategy helps you make the most of your data, making it known, discoverable, available, trusted and compliant. It supports business objectives, like increasing revenues, improving customer experience and driving profitability, by giving your business units and users access to relevant data to quickly gain the insight they need. It also helps you control costs and reduce risks, enforcing consistent security and governance across all your enterprise data assets.

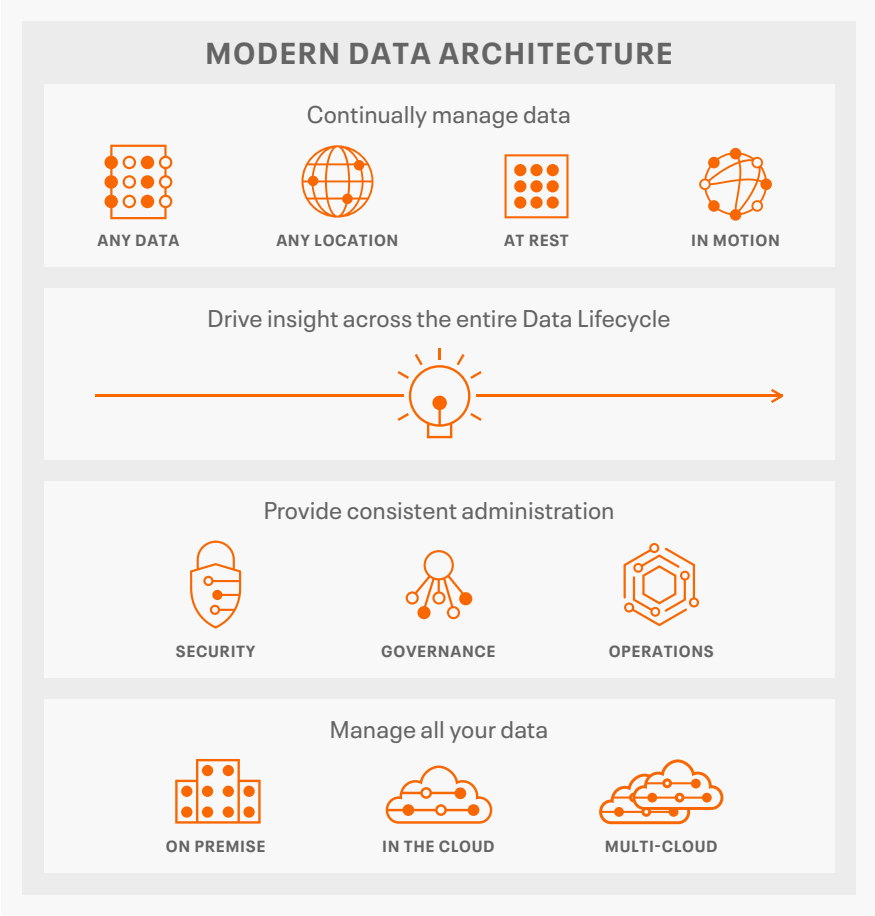
Make sure your organization has a strong enterprise data strategy that informs and drives your hybrid cloud strategy.

And finally, it establishes a modern data architecture that serves the needs of your organization by finally putting the horse before the cart. This hybrid architecture is designed to provide consistent data services and functionality that enables you to share data, metadata and workloads across data centers and public clouds. It allows you to:

- Continually manage data from any location, at rest or in motion
- Drive insight across the entire data lifecycle
- Provide consistent administration
- Manage all your data, no matter where it resides

Combining different infrastructure form factors with diverse analytic capabilities with consistent security and governance on a single, open platform is known as an enterprise data cloud, which represents a new data management architecture that includes:

- Analytics across the complete data lifecycle, including data processing and analysis at the edge to data warehousing, real-time operations and machine learning
- Unified data security, governance and insight to drive your data enterprise strategy
- Open source and frameworks for community innovation and integration with third-party tools
- The flexibility to deploy use cases on any cloud and data center to leverage any type of data wherever it lives



What It Means to Do Cloud Better

In an ideal world, organizations could leverage the most appropriate public clouds for data and analytics in a secure, cost-effective and scalable manner, meeting the needs of all users.

That would mean:

- Data users get the experience they need and want to use
- Central IT maintains deployment flexibility, as well as centralized administrative control—including consistent data security, governance and control

In short, IT could deliver integrated, multi-function analytics on centrally-managed and secured business data, deployed anywhere and with a consistent experience—on-premises or in hybrid and multi-cloud.

And finally, not only can organizations simultaneously enjoy cost savings, flexibility and agility, and security—again, without having to choose or compromise—but they can also control the degree to which each is necessary to meet their objectives at any given point. Instead of choosing once (at the time of deployment), which of those is their primary goal, they would have the ability to continually adjust their focus from day to day and hour to hour, based on business needs.

Deliver integrated, multi-function analytics on centrally-managed and secured business data, deployed anywhere and with a consistent experience—on-premises or in hybrid and multi-cloud.

How Do You Get There from Here?

The journey to doing cloud better begins from where you are today. Most organizations have one of three departure points.

Leveraging Cloud as IaaS

Organizations running CDH or HDP workloads on IaaS are often surprised and frustrated they do not experience the cloud benefits they expect. These expectations can range from elasticity and agility to scalability and a pay-as-you-go approach. IT/Infrastructure Managers are concerned about infrastructure and operational costs as spinning up new workloads requires manpower and additional infrastructure. In this context, when business users ask IT for more data or analytics capabilities, the answer is often either a flat-out "no" or involves a multi-month project to provision the new capabilities. Additionally, on-going concerns about securing and governing the data hamper the speed at which IT moves.

These organizations can increase cost-effectiveness and control with very little effort. Cloudera Data Platform Public Cloud (CDP-PC) is a cloud-native hybrid data architecture that is easy to deploy, manage, and use. It delivers powerful, self-service analytics across hybrid and multi-cloud environments with the granular security and governance policies that IT leaders require to keep their organizations' data safe. CDP-PC's cloud-native characteristics allow it to make more efficient use of cloud infrastructure, leading to significant savings on that front, reducing overall TCO.

For companies deploying CDH or HDP on IaaS, moving to CDP-PC could result in a cost savings of 25% or more with the least amount of effort.

For companies deploying CDH or HDP on IaaS, moving to CDP-PC could result in a cost savings of 25% or more with the least amount of effort. For starters, shifting from IaaS to PaaS saves money, as true cloud agility means deployments can dynamically scale up as well as down so that you always have the optimal performance to meet SLAs yet don't pay for infrastructure you don't use. The move to CDP-PC also enables more complex, multi-stage and multi-tenant use cases and data pipelines. Furthermore, it increases security and governance through a future proof architecture for data and analytics.

Data Center Inflexibility

Other organizations run legacy Cloudera or Hortonworks distributions in their own data center. These deployments can be much more difficult to set up and upgrade. Though they provide all building blocks for complete data lifecycle analytics, organizations still need to create their own services. Upgrading these deployments also pose their own challenges. Without containerization and separation of compute and storage, platform upgrades are complex, all-or-nothing affairs that require extensive planning and testing. As a result, organizations simply remain on older versions and do not get the full benefit from their investment.

A move to the public cloud results in vastly improved ease of use. With CDP-PC, not only is deployment of the platform as a whole greatly simplified, organizations now also have a choice between either creating their own customized services using the capabilities of the platform or leveraging one of the ready made Experiences to deliver common analytics like data warehousing, machine learning, data engineering or more. Finally, the switch to PaaS means these organizations will never again have to upgrade their Cloudera distribution—it is always on the latest and greatest version.

In short, when these organizations leave the data center behind for CDP-PC, they will be able to:

- Reduce operational overheads and maintenance costs
- Deliver efficient scalability for more users, use cases and workloads
- Improve security and governance with a future-proof architecture for data and analytics

Moving from legacy Cloudera platforms to CDP-PC

A US-based telecommunications company had deployed workloads using a legacy Cloudera platform on IaaS. Leveraging CDP-PC's cloud-native hybrid cloud data architecture, the organization was able to make significant savings in cloud storage and compute resource costs, amounting to 40% on an annual basis.[4]

To sum up, the shift from IaaS to PaaS with CDP-PC:

- Improves cost-effectiveness and control
- Enables more complex multi-stage and multi-tenant use cases and pipelines
- Increases security and governance through a future proof architecture for data and analytics

Patchworked Cloud Point Solutions

Central IT is faced with two major challenges with cloud point solutions.

First, and for those cloud point solutions that resulted from shadow IT, they introduce tremendous security and compliance risks, accounting for around 40% of all IT budgets and reduce the opportunity for better data exploitation.

Secondly, and regardless of whether the choice for cloud point solutions was made consciously or through adopted shadow IT, organizations need to realize that integrating these systems, including the establishment of consistent security and governance, becomes their responsibility and that they have to incur the "integration tax."

Both of these issues are addressed in CDP-PC. An integrated platform, it offers the complete range of analytical capabilities across the data lifecycle that end users want to use and can use in a self-service manner. Enterprise IT is confident in the knowledge that access to data and analytics is carefully controlled with the shared data experience (SDX) layer, providing consistency across analytics and deployments without additional effort. There is no more integration tax to contend with. Compared to point solutions like Snowflake, CDP is non-proprietary, both in implementation (open source) and data formats.

For these organizations, CDP:

- Delivers a single platform for full data lifecycle use cases without data or vendor lock-in
- Provides more flexibility, security and governance across form factors
- Improves TCO with high performance, security and multi-tenancy out of the box

40%

Cloud point solutions introduce tremendous cost with shadow IT accounting for around 40% of all IT budgets.

The Beauty of the Enterprise Data Cloud

The truth is, a big data platform is no longer what modern enterprises need these days. Instead, they need an enterprise data cloud that includes the following within a single platform:

- Analytics for the complete data lifecycle
- Support for any cloud and data center
- Consistent security and governance
- Openness: open source and open frameworks

Cloudera is the very first—and so far the only—enterprise data cloud. To learn more, discover Cloudera Data Platform.

Cloudera Data Platform (CDP): The industry's first enterprise data cloud.

Learn More

Cloudera Data Platform powers data-driven decision making by easily, quickly, and safely connecting and securing the entire data lifecycle. Learn more at cloudera.com

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com | US: +1 888 789 1488 | Outside the US: +1 650 362 0488

Sources

[1] Critical Success Factors for an Enterprise Data Strategy, Cloudera, 2020

[2] Critical Success Factors for an Enterprise Data Strategy, Cloudera, 2020

[3] Why a Successful Hybrid Cloud Strategy Requires and Enterprise Data Strategy, Cloudera, 2020

[4] CLOUDSMART: CDP Public Cloud Adoption Service, Cloudera, 2019

Cloudera, Inc. 5470 Great America Pkwy, Santa Clara, CA 95054 USA cloudera.com

CLOUDERA