



CLOUDERA

3 Key Considerations For a Hybrid Cloud Data Platform

Unlock the Potential of Your Data

Table of Contents

The big idea: hybrid cloud	3	One dashboard for multi-function platform management	12
A balanced, enterprise-wide portfolio	4	Single console management	13
Key factors of hybrid cloud	5	Comprehensive workload insights	14
Full data lifecycle support	5	Streamlined data movement	15
Enterprise-class security and governance	6	A value-driven approach	16
Open platform	7	Take the next step	17
Managing a hybrid cloud	8		
Data access and governance	9		
Write once, apply everywhere policies	10		
Track and audit everything	11		

The Big Idea: Hybrid Cloud

The premise of a hybrid cloud is that it fuses together two data architecture form factors: the on-premises data center (private cloud) and public cloud. It is a vendor-agnostic canvas upon which IT designs and delivers fit-for-purpose solutions to their organization at the highest levels of agility, speed, and cost efficiency.

Although that sounds too good to be true, this paper will describe the key factors that make hybrid cloud possible and introduce the data management capabilities that make it a reality.

Businesses know that revenue generation and risk mitigation strategies depend on modern data platforms that include cloud-native capabilities. They've also come to realize that an integrated data architecture is the best way forward. Research shows that 96% of enterprises are actively pursuing a hybrid IT strategy.¹

This is because modern, real-time businesses require accelerated cycles of innovation that are expensive and difficult to maintain with legacy data platforms. However, cloud technologies and respective service providers have evolved solutions to address those challenges. The hybrid cloud paradigm gives companies options to leverage those solutions and

to address decision-making criteria, on a case-by-case basis, such as:

- **Economic:** Rent or buy? Hybrid cloud enables a value-driven approach to determine whether renting (public cloud) or buying (private cloud) makes the most economic sense for a particular use case.
- **Operational:** New projects are challenged with the unknown. Will it be a blockbuster or a dud? A public cloud launch affords you the flexibility and time to evaluate actual infrastructure requirements before private cloud capital investments are made.
- **Functional:** Hybrid cloud enables you to unleash the full capabilities of on-premises apps by bursting to the public cloud in order to leverage on-demand infrastructure for highly-computational workloads.

Why You Need Hybrid: Cloud Concentration Risk

Cloud Concentration Risk is a real worry for all industries with regard to ensuring uninterrupted services to their customers. But it is such a concern for the Financial Services Industry (FSI) that the issue has the attention of global and regional regulators.

That concern stems from an institution's over-reliance on one service provider to support key services.

Hybrid and multi-cloud data environments mitigate the risk of an outage of a single cloud vendor by enabling the use of a wide expanse of cloud providers and on-premises private cloud.

To mitigate cloud risk, embrace open source solutions and open architecture that ensures interoperability between clouds and portability of data and applications across any platform.

For more information, read the article, [Minimizing Cloud Concentration Risk for Financial Services Institutions, Regulators and Cloud Service Providers](#).

A Balanced, Enterprise-Wide Portfolio

The decision to place a data workload in a private or public cloud is best made by Line of Business (LOB). An enterprise benefits as workloads across LOBs are considered part of its larger portfolio, which can be balanced across private and public clouds. This provides an enterprise the best of both worlds. This also ensures that platform exposure remains within the organization's area of expertise and actual needs, including:

- **Operational efficiency²:** An effective operations team (e.g., one admin manages 1,000, not just 100 or 10 virtual machines) makes public cloud services redundant, favoring the private cloud option.
- **Platform utilization:** If on-premises infrastructure is lightly used or intermittently, it may be better to utilize public cloud services, rather than to buy and maintain infrastructure for private cloud.

Periodic rebalancing does occur. Research shows that 57% of applications are expected to move between private and public clouds³. This demonstrates that the decision of private/public cloud is not a binary one, and a hybrid approach gives the most options. The next section describes three key factors that enable private cloud and public cloud to operate as a hybrid cloud.



57% of applications are
expected to move
between private and
public clouds³

Key Factors of Hybrid Cloud

In the hybrid cloud, both private cloud and public cloud environments work in tandem with the ease of use, elasticity, and self-service that one comes to expect from cloud-native services. This enables organizations to optimize existing data, exploit new resources, and apply the latest innovations. Hybrid is only possible when three key factors are incorporated: (1) full data lifecycle support with (2) enterprise-class security and governance over (3) an open platform.

Full Data Lifecycle Support

Data life cycles consist of collecting, enriching, reporting, and serving data on which analytics and machine learning are applied to interrogate and predict outcomes. Precision requires that those results be collected, enriched, reported, served, and analyzed further in an iterative cycle of action and insight.

However, any point of the data lifecycle can present a bottleneck. For example, efforts made by enterprise teams to enforce global infrastructure standards by controlling, migrating, replicating, or balancing workload placement may disrupt one or more LOBs from running any number of data analytic workloads.

To ensure that LOB and enterprise priorities don't conflict, hybrid cloud separates data management from infrastructure strategy, providing flexibility that enables organizations to move data and apps from one environment to another without rewriting applications and retraining personnel. This ensures interoperability, so that multi-function analytics can be run across private and public clouds without disruption.

Flexibility and interoperability assumes uniform policies across all environments, which brings us to the next key factor of hybrid cloud.



Enterprise-class Security and Governance

An organization, its partner, and customers, all benefit from enterprise-wide multi-function analytics only if data is protected, secured, and governed throughout. Analytics and machine learning can become a risk if data security, governance, lineage, metadata management, and automation is not holistically applied across the entire data lifecycle and all environments.

If data access policies and lineage aren't consistent across an organization's private cloud and public clouds, gaps will exist in audit logs. This is an InfoSec compliance nightmare because each gap is a potential lawsuit and/or regulatory fine, not to mention the damage to the brand of the company. Gaps also lead to inconsistent insight and, with that, decisions that impact the business' ability to innovate and differentiate.

A hybrid cloud has unified metadata, data access, governance, and lineage across all environments through one common user interface, regardless of where the data is sourced, migrated, or replicated.

A large orange '2X' graphic with a textured, striped pattern inside the numbers, indicating a doubling of speed.

Faster workload processing speed

Two for One with Hybrid: New Insights and Faster Service

One of the largest providers of finance, marketing, and credit solutions in South America selected Cloudera CDP for Public Cloud to create a fully automated data pipeline that leverages the best of both public and private cloud form factors.

The hybrid architecture enables this company to uncover customer insights that are often hidden within vast quantities of publicly available data. Through this, they have been able to generate new revenue opportunities, including up-leveling product offerings.

Furthermore, workload processing speed doubled, allowing them to consistently meet SLAs more efficiently.

A number of factors went into selecting Cloudera over the competition. Differentiators include low TCO and the streamlined process by which to lift and shift workloads and to leverage burst to cloud capabilities.

Open Platform

The third factor that makes hybrid cloud possible is having an open architecture, open source software, open APIs, and an open storage format.

Open source technologies enable extensibility, flexibility, and avoidance of vendor lock-in, regardless of where the data is stored and workloads are run. This is one of the most important factors of a hybrid cloud because it enables novel solutions in the following ways:

- 1. Access to innovation:** The open source community drives change, innovation, and feature functionality at a higher rate than any one organization, even with heavy R&D investment.
- 2. Community of expertise:** Mature and active open source communities ensure that your developers are self-sufficient and productive with easy access to expertise and examples.
- 3. Flexibility and choice:** Organizations always have access to their data, and don't get locked-in by any vendor.

Open source enables interoperability because the same services that run in private cloud are run in public cloud, so companies aren't beholden to any one cloud provider's business model and priorities.

Enterprises are thus free to establish patterns of infrastructure that are the most economical, resilient, performant, green, and/or ethical for them. Furthermore, each LOB technology team has the flexibility to adapt with shifting business priorities.

The three factors described above: (1) full data lifecycle support; (2) enterprise-class security and governance; and (3) open platform are what makes a hybrid cloud possible. But the real challenge is in how to manage it and make it a reality. That is addressed next.

Extend Data Streaming to the Public Cloud

Enterprises struggle with finding a streaming data platform that can seamlessly span across hybrid cloud.

Cloudera Data Platform (CDP) addresses that challenge with Cloudera DataFlow (CDF).

CDF supports the entire set of streaming data capabilities. From data capture and flow management at the edge, to provisioning that data directly to/from your messaging backbone and/or stream processing and analytics engines.

With CDP DataFlow, you extend that streaming experience to the cloud.

With a "one-click" approach, the CDP interface accelerates deployment of flow management, streams messaging, and streaming analytics clusters to the cloud in just minutes, without the heavy burden of generating infrastructure requirements to develop and configure them.

To understand how Cloudera DataFlow provides the complete streaming platform for all clouds, download the eBook, [Extend Your On-premises Streaming to the Public Cloud](#).

Managing a Hybrid Cloud

A hybrid cloud is only as good as the platform and tooling used to manage it. Considering all the moving parts, having a holistic solution is critical. This section describes how to make hybrid cloud a reality through data management capabilities that align to the three hybrid cloud factors described above.

The diagram to the right shows how one platform with a single data architecture is able to bridge two different form factors across multiple public clouds and an on-premises private cloud. In this example, we illustrate Cloudera Data Platform (CDP).

As it relates to the hybrid cloud factors described earlier, CDP supports the full data lifecycle for analytics and machine learning from your on-premises private cloud and multiple public clouds (AWS and Azure, with GCP coming soon). It also provides comprehensive data security and governance (through Cloudera Shared Data Experience (SDX)), while the entire platform is on a foundation of open source software.

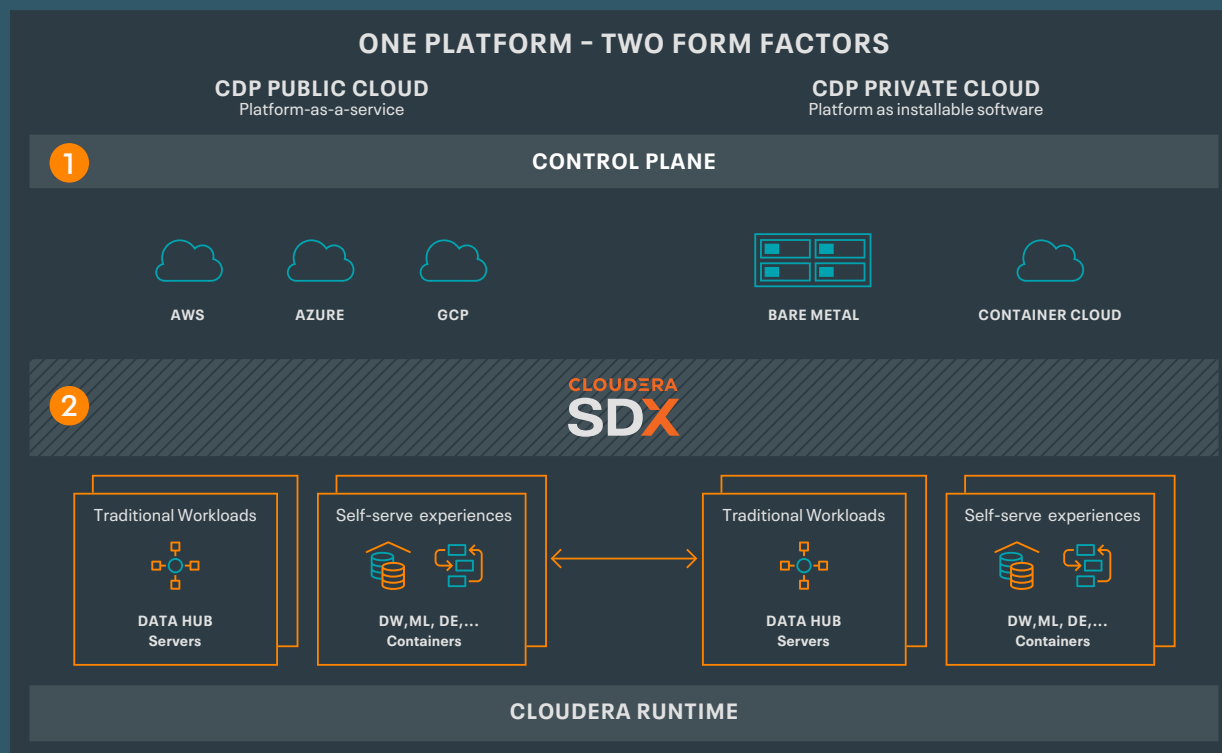


Figure 1: Concept diagram of how (1) the Control Plane and (2) Shared Data Experience (SDX) bridge two data architecture form factors: private cloud and public cloud to create hybrid cloud.

Data Access and Governance

A key factor for hybrid cloud success is the ability to not just establish but also maintain consistent data security and governance policies between the different analytic functions across the data lifecycle and their deployment to public and private cloud.

Of its many powerful features, below we briefly describe two of SDX's key capabilities that, as part of the wider platform, provide unified data access and governance.

Public Cloud Security

Ensuring consistent data security and governance while having the flexibility needed to make data available to end users is a key challenge that organizations face today.

Download the white paper [CDP Public Cloud Security](#) to learn about key technologies that enable organizations to maintain security and governance while taking full advantage of the agility and elasticity of the public cloud.

Write Once, Apply Everywhere Policies

Hybrid and multi-cloud environments bring new challenges of keeping data security and governance policies consistent between different deployments. Implementing and synchronizing policies between private and (multiple) public clouds can result in a tremendous amount of operational effort.

SDX addresses those challenges for all CDP deployments. Provided as a seamless integrated data context layer, SDX delivers transparent data security and governance policy management as well as enforcement. Administrators set policies once and have them consistently applied everywhere without additional effort, enabling safe, secure, and compliant end user access to data and analytics.

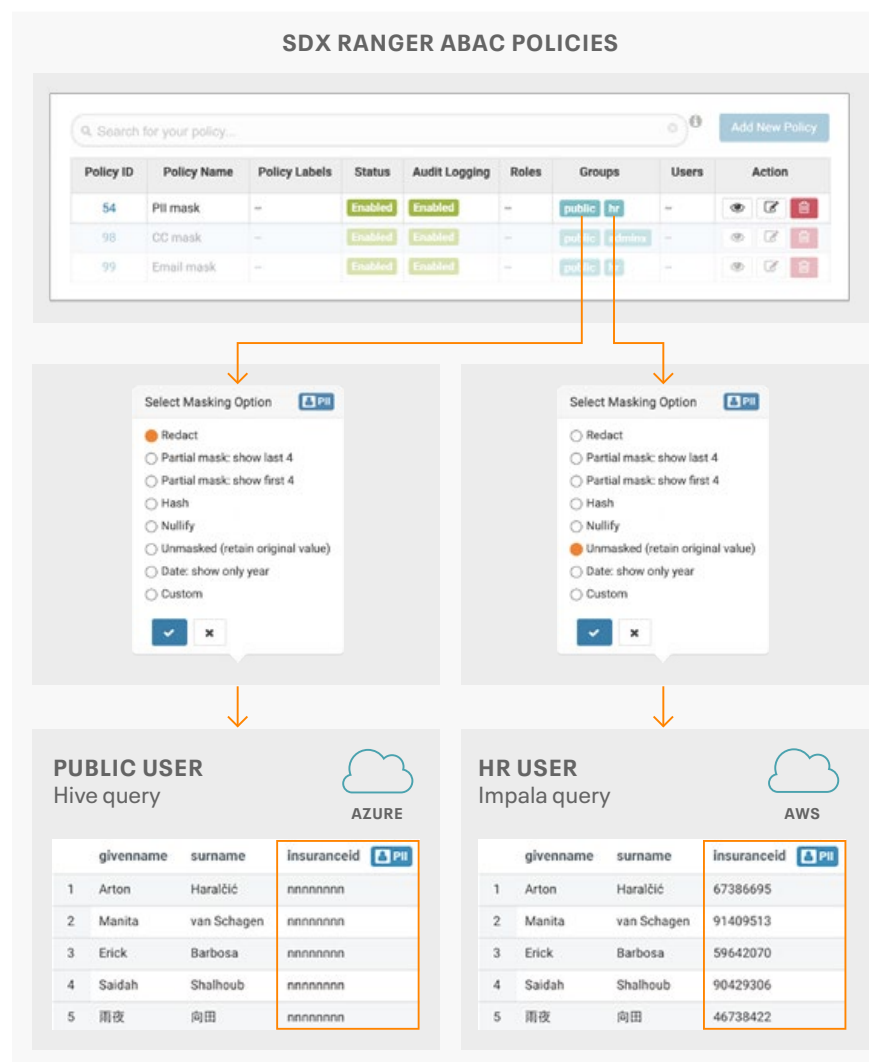


Figure 2: Example of Attribute Based Access Control policy applied across clouds. SDX's data access capabilities work hand in hand with its data governance, described on the next page.

Track and Audit Everything

Enterprise data lineage and full end-to-end audit capabilities across the entire ecosystem of CDP deployments is critical to hybrid data governance, all of which are provided by SDX.

For instance, with SDX, you have access to the lineage, metadata, and metrics associated with data utilization across environments. The propagation of data classifications—automatically gleaned through profiling—along the lineage ensures data access policies are consistently and demonstrably enforced, even as data is moved or derived.

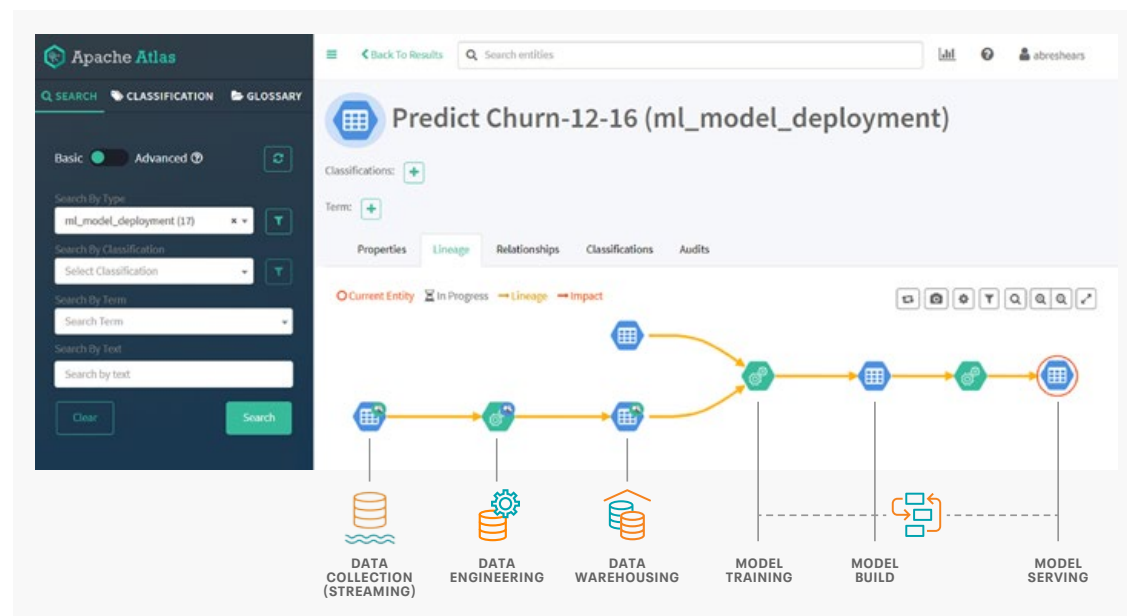


Figure 3: SDX leverages Apache Atlas to follow data lineage across multiple environments.

Commitment to the Open Source Community

Cloudera has a long and proven track record of identifying, curating, and supporting open standards that provide the mainstream, long-term architecture upon which new customer use cases are built.

The foundation of CDP is open source. Deep engineering relationships with a variety of open source projects have led to critical innovations and product improvements in the following areas.

- Data Warehousing with Apache Impala, Apache Hive, and Apache Druid
- Data Engineering with Apache Spark
- Operational Database with Apache Hbase and Apache Kudu

Additionally, Cloudera provides the best real-time streaming and data analytics ecosystem through:

- Apache Nifi for edge and flow management
- Apache Kafka for streams messaging
- Apache Flink and other open source stream processing and analytics engines

Read the white paper to learn how [Cloudera's Kafka ecosystem](#) ensures a sustainable and adaptable end-to-end streaming architecture.

One Dashboard for Multi-Function Platform Management

Modernizing a data platform is a complex process run by a diverse set of people, most of whom might be unaware of who is doing what. Transparency helps to reduce those complexities.

A 10,000 foot view of the entire workspace helps to alleviate the struggles that come with multiple teams managing data across multiple environments with multiple tools. In the CDP architecture, this problem is addressed through the CDP Control Plane.

The Control Plane in CDP has a common set of tools for management, workload analysis, and data movement across multiple environments. LOB DevOps teams can easily spin up and manage clusters based on their specific use case while the enterprise teams are provided a comprehensive view of everything that is going on across the clusters.

The next few pages describe some of those tools, including: CDP Management Console, CDP Workload Manager (WXM), and CDP Replication Manager.

Successfully Migrate to the Public Cloud

Organizations continue to optimize how they gain value and insight from their data in order to make better business decisions. More often than not, they've done this by getting smarter about the public cloud.

Cloudera can help you simplify your move to CDP on public cloud and go from pilot to production quickly, painlessly, at lower cost, and with peak performance that is underpinned with consistent data security, governance, and management.

For more information, download this eBook, [3 Steps to Successfully Migrate to the Public Cloud](#).

Single Console Management

To make hybrid a reality, data platform admins need one convenient user interface from which to administer, manage, and provision users, environments, and services across all clusters and analytic workspaces.

In the CDP architecture, this is done by the Management Console, which enables data platform admins to:

- Register all public cloud accounts (AWS and Azure, with GCP coming soon) as well as private cloud and launch clusters and analytic workspaces within each environment as needed.
- Track the details of all clusters ranging from function and performance to location, status, metrics, and more.
- Easily onboard existing users and assign them CDP roles, groups, and resources to tightly control access to platform components and services.
- Integrate the full range of CDP services into your applications via your favorite interface (a web GUI or the Java SDK) and automate routine tasks with terminal windows through a Command Line Interface.

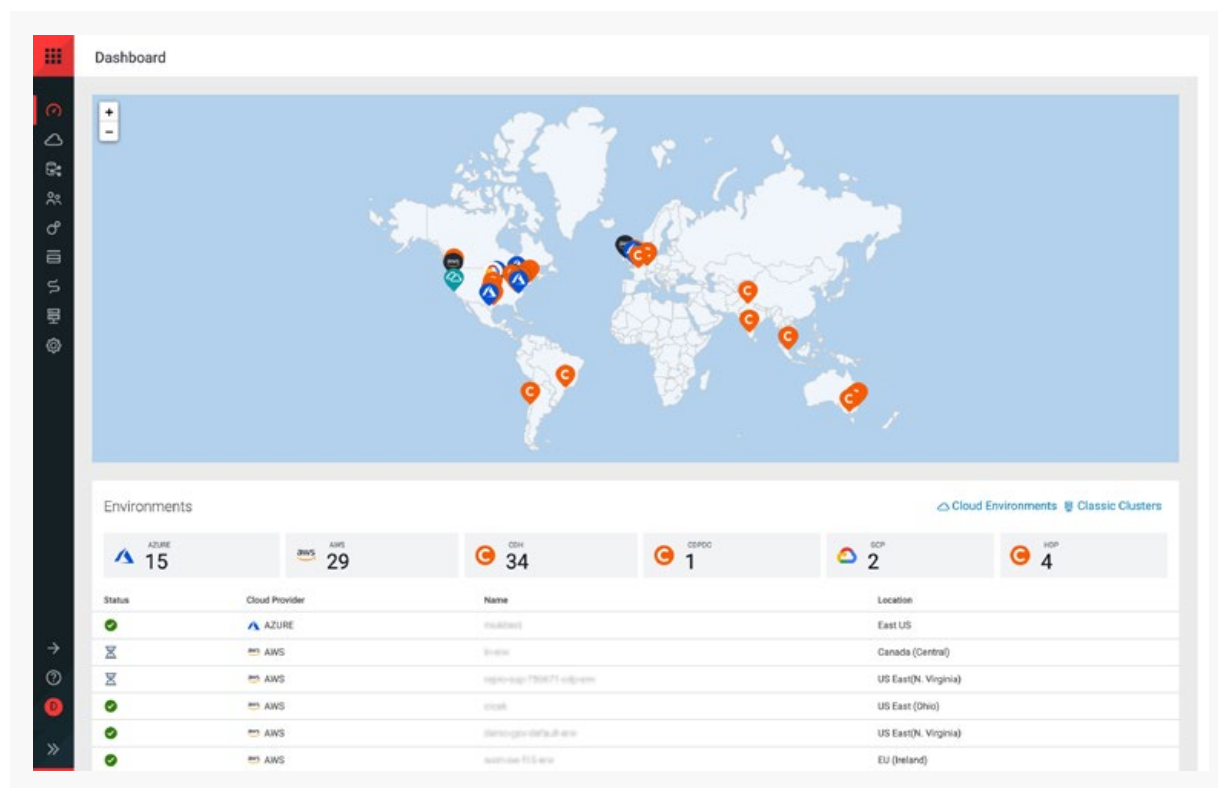


Figure 4: The Control Plane in CDP is one pane of glass from which to monitor and work across all environments, users, and analytic work spaces.

Comprehensive Workload Insights

Accelerated product development cycles require a streamlined step-by-step process by which to migrate workloads and burst to cloud. It also requires transparency into the health of your clusters and other valuable workload insights.

CDP Workload Manager (WXM) provides cross-platform visibility while also automating the process, including:

- Evaluating the cloud readiness of each workload.
- Creating a tailor-made capacity plan that ensures workload optimization.
- Generating a replication plan based on deep analysis of each workload (integrated with Replication Manager, next page)

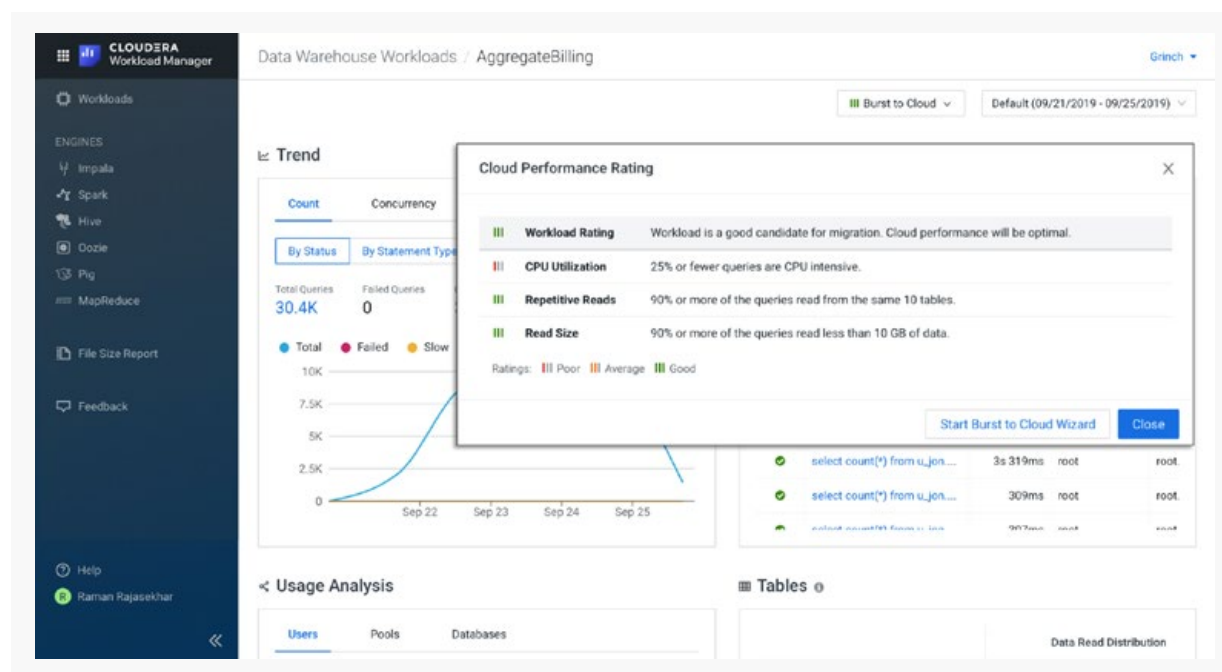


Figure 5: CDP WXM showing replication plan results and one click burst to cloud capability.

Streamlined Data Movement

Hybrid cloud often has clusters distributed across wide geographic areas, so ensuring that any set of clusters are synchronized and knowing the status of each is critical.

CDP provides a centralized management tool called Replication Manager that replicates and migrates Hive, Impala, and HDFS data. It respects and retains the data context set forth by CDP SDX as it moves data for any purpose and between environments. It is a simple, easy-to-use, feature-rich data movement tool providing:

- Integration with WXM to simplify workload migration.
- Controlled data and metadata replication that adheres to standard policies.
- Hybrid cloud flexibility through continuous synchronization.
- Complete backup and disaster recovery.
- Convenient creation of development and test systems.

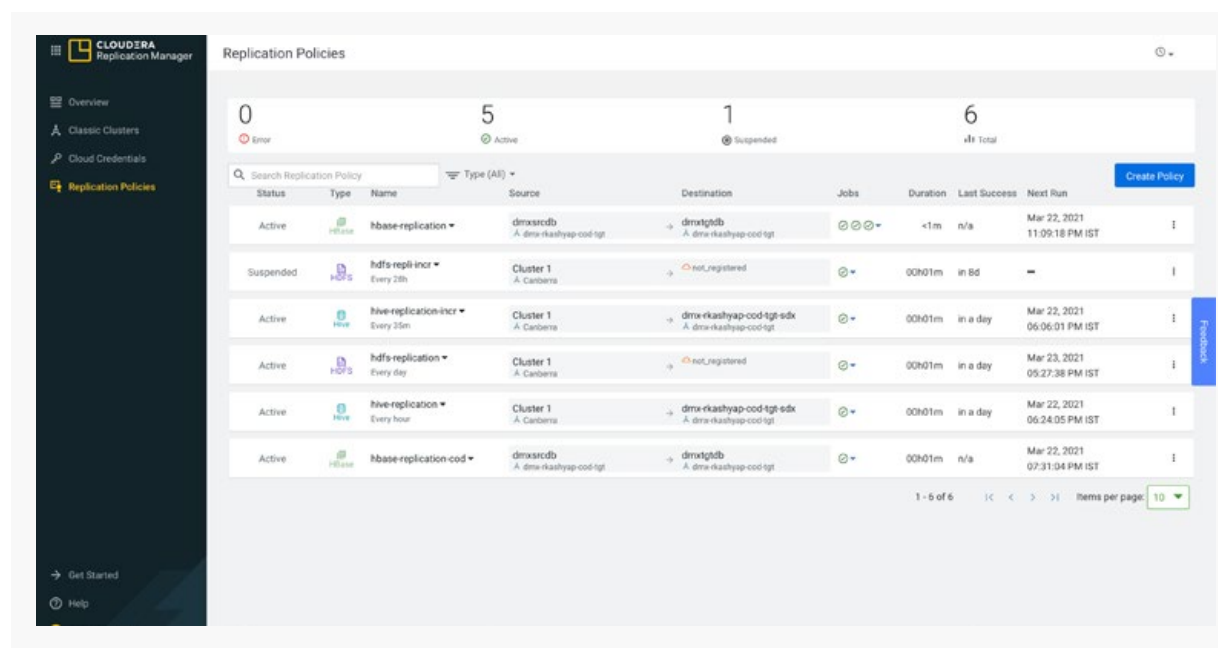


Figure 6: CDP Replication Manager dashboard summarizing critical replication information.

A Value-Driven Approach

A hybrid cloud strategy is a pragmatic one because it enables business and technology organizations to adopt a value-driven approach for a data platform that best serves each LOB and the enterprise as a whole. This paper outlined the key characteristics that make hybrid cloud possible and described the technological innovations and data management capabilities that make it a reality.

CDP makes hybrid cloud a reality by addressing the key characteristics described below.

- **Any cloud:** Operates across all major public clouds (AWS and Azure, with GCP coming soon) and the private cloud with a cloud-native experience.
- **Multi-function analytics:** Integrates data management and analytic experiences across the entire data lifecycle for data anywhere.
- **Secure & governed:** Delivers data security, governance, and metadata management across all environments.
- **Open platform:** Open source, open integrations, extensible, and open to multiple data stores and compute architectures.

Take the Next Step

CDP has been built to meet the needs of companies on their hybrid cloud journeys. With extensive experience implementing data-driven use cases, Cloudera has the expertise to help you simplify your move to a hybrid cloud and go from pilot to production quickly, painlessly, at lower cost, and with peak performance that is underpinned with consistent data security, governance, and management. To move forward, consider these next steps:

- [Read more](#) about how CDP unlocks the potential of private *and* public clouds with a hybrid cloud platform designed for any cloud, any analytics, and any data.
- Get hands-on with the platform and services with a [test drive](#).

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

¹ 451 Research – Voice of the Enterprise – Cloud, Hosting & Managed Services Vendor Evaluations 2019

² Engagement with Cloudability, Cost Management in the Cloud Age, 2019

³ 451 Research – Voice of the Enterprise – Cloud, Hosting & Managed Services Vendor Evaluations 2019

© 2021 Cloudera, Inc. All rights reserved. Cloudera and the Cloudera logo are trademarks or registered trademarks of Cloudera Inc. in the USA and other countries. All other trademarks are the property of their respective companies. Information is subject to change without notice. 4779-001 2021

[Privacy Policy](#) | [Terms of Service](#)

CLOUDERA