

CLOUDERA

EBOOK

A Hybrid Data Cloud for Accelerated Insight



Table of Contents

A Letter From the Editor	3	Leveraging Data From Edge to AI	11
A Brief View on Today's CSP Landscape	4	What Does This Mean for CSPs?	11
Establishing the Essentials	5	5G Network Analytics	12
Managing Compliance Challenges	5	Cloudera 5G Network Analytics	13
Meeting Modern Requirements	6	How Can Cloudera Benefit Telcos?	14
The Architecture Challenges Facing a New World of Data-Centric CSPs	7	The Benefits of Cloudera Agile Telco Data Cloud	15
A Hybrid Cloud Data Model	8		
What Does This Mean For CSPs?	9		
Data Unification, Security, and Governance	10		
What Does This Mean For CSPs?	10		

A Letter From the Editor

Dear Reader,

Cloudera has been a part of the telecommunications journey for many years. As a key piece of the data infrastructure, Cloudera tools and platforms have, through the years, delivered significant value. And now, as the industry moves to the cloud, we are excited about a new journey. Namely, the possibilities of a hybrid data cloud strategy, and how we can accelerate the journey to cloud and faster growth in the industry.

This journey to cloud comes at a particularly critical time. Against the backdrop of the pandemic, Communication Service Providers (CSPs) have supplied the lifeblood for an increasingly digital society: helping to keep businesses functioning, remote workforces productive, and enabling all of us to access information, entertainment, and each other. CSPs have proved to be more than a utility — they are essential to our resiliency.

The issue here isn't your data capabilities — telcos have very mature architectures and use cases already in play. Rather, a bigger challenge will come from being able to manage and govern these combined data assets: dealing with duplicate and overlapping capabilities and applications; implementing process and operational efficiencies; adhering to data compliance; and developing a single view across the business.

Overcoming the aforementioned challenges, of course, connects to a crucial and overarching aim that is never far from mind: to reduce and optimize operational costs.

A key part of the solution, of course is cloud. After all, you already have various arrangements with GCP, AWS, and Azure. But the question is: do you have flexibility to be able to move between them? Can you provide consistent data governance across all environments? How do you ensure that the right cloud is managing the most appropriate workload? Are these disparate architectures hindering how you leverage 5G, IoT, and other new technologies?

I believe, as your Hybrid Cloud Data Partner, Cloudera can provide you with a platform to help centralize the control and use of data across your data architecture, accelerating the value of all enterprise data assets with a single governance layer.

This is because Cloudera can ingest and integrate any type of data, in any of your public or private clouds, to enable real-time analytics and machine learning to deliver actionable intelligence and insights.

I hope this eBook helps to highlight how I think Cloudera can help telecommunications service providers make the best success for their future, improve margins, and generate shareholder value.



Anthony Behan
Editor/Managing Director, Communications,
Media & Entertainment Industries

A Brief View on Today's CSP Landscape

While the pandemic has undoubtedly accelerated the need to change, telcos have been undergoing a fundamental transformation around who they are and what they offer for a long time.

Smartphone penetration has reached saturation point and traditional revenues have plateaued. Challenger operators have driven down prices, while telcos have lost revenue from their own core services to over-the-top (OTT) platforms. With telcos now battling for the same customers, the pressure is on to significantly reduce the costs of operations, while at the same time delivering better experiences to improve customer retention.

Indeed, while network quality can still be a differentiator, telcos need to stay relevant in this challenging environment and be able to offer more than simply end-to-end connectivity. This means embracing new opportunities. It also means recognizing the importance of the enterprise business, the wholesale business, and innovative ecosystem development.

Most obviously, these opportunities arrive with the dawn of 5G. According to a recent report, 49% of telecom leaders are already making large investments in 5G mobile technologies, with a forecasted 1.5 billion 5G subscriptions globally expected by 2024¹.

By leveraging 5G, this opens the door to a number of new, and potentially highly profitable, new business models. Creating private 5G networks with network slicing, for example, is predicted to generate \$20bn of revenue by 2024², as well as enable service providers to offer many innovative new services to enterprises. Similarly, 5G in conjunction with IoT and AI can open the door to use cases that include everything from smart homes, factories, and cities to digital healthcare and autonomous vehicles. These businesses have data and analytics at their core, whether for service automation, service assurance, or developer/client collaboration.

Telecom operators can reduce costs by up to

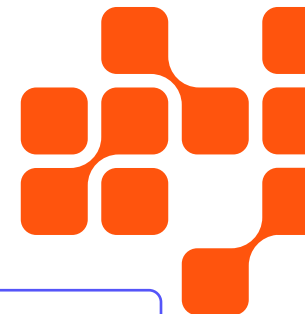
40%

using machine learning and applied analytics³

³ McKinsey, A Future for Mobile Operators: The Keys to Successful Reinvention

¹ Telecom's 5G future, IBM

² 5G Slicing Set to Propel New Growth in the Industry and Generate Revenue of US\$20 Billion by 2024, ABI Research



Establishing the Essentials

Before this untapped revenue can be captured, and before new business models can be introduced, there is a new operational foundation of data governance that CSPs need to establish. This first — and biggest — challenge is one that we're all familiar with: bridging problematic data and business silos.

With data often spread across countless numbers of fragmented legacy technologies, or created through mergers and acquisitions, the issues associated with silos are well documented. For example, disparate data sources impacting data integrity and accuracy, costly and inefficient resource duplication, incompatibilities between departments and systems of record hindering collaboration, and of course, that essential consolidated view of company data.

This holistic view is crucial for the introduction of advanced analytics, reporting, and data management capabilities. It also helps enable increased levels of operational agility and efficiency, employee productivity and engagement, and to improve customer experiences.

⁵ [17 Major GDPR Fines of 2021 & 2021, Cooke Law Info](#)

Managing Compliance Challenges

CSPs also must negotiate a continually changing regulatory environment. With the speed of innovation, the introduction of 5G, and as the industry evolves from traditional telecommunications to internet-based networks comes complex, and often ill-defined legislation. And, as telcos migrate to the public cloud they find themselves in an environment that brings both security and privacy issues. At the same time, the scale of the data challenge becomes larger, while demands on speed — and real-time capacity — grow.

Regulators can be slow to update policies or make contradictory decisions, which can make it increasingly difficult for CSPs to ensure they are meeting their compliance requirements. And, with breaches of GDPR leading to punitive costs up to the equivalent of 4% of global turnover, fines often run into tens of millions. It's worth noting, over the course of 2020–2021, that three separate telecoms operators featured in the top ten of major fines, for a combined total of €56.75m⁵.

CSPs with a comprehensive, analytics-based approach can reduce churn by as much as

15%⁴

⁴ McKinsey, Reducing Churn in Telecom Through Advanced Analytics

Meeting Modern Requirements

CSP architectures need to support the delivery of new products and technologies to consumers, while also leveraging many of these same technologies for their own business. At a top level, this requires an architectural foundation to address challenges including:

- **A compatible and modular design:**
Architectures to leverage new technologies, with open interfaces to aid implementation, but also designed with future use cases in mind.
- **Scalability and flexibility:**
With demand for data intensive applications such as AI, VR, and AR growing, architectures need to be able to scale according to different usage scenarios.
- **User performance demands:**
Users are quick to change suppliers if their service expectations are not met, meaning network infrastructures need to deliver the experience they demand, every time.

- **Round-the-clock security:**
CSPs need to know they can prevent, mitigate, and respond to cyber issues — not just for themselves and their users, but also to meet the needs of regulators.
- **Sustainability:**
The need to take action around ESG is increasingly important. This means architectures have to consider performance, security, reliability, and scalability, but also efficiency and carbon output.

In this eBook, we'll look at approaches to address some of these architectural and data governance issues and approaches that telcos can take as you migrate to the cloud, embrace the opportunities of 5G, and address the challenges and opportunities in the aftermath of the global pandemic.



The Architecture Challenges Facing a New World of Data-Centric CSPs



CSPs are increasingly reliant on data and analytics to help them enhance the end-to-end customer experience, optimize their network and operations, and enable new revenue streams and business models. Moreover, with a direct connection to tens of millions of customers, they are also one of the largest aggregators of data, from a wide variety of data sources. With the advent of 5G, and the increasing adoption of IoT — where there are expected to be 24.6bn new connections by 2025⁶ — this vast pool of data will continue to grow.

CSPs need to be able to make the most of this increased volume, variety, and complexity of data in order to gain the enterprise data insights they need to drive business value. This means they need the ability to handle any data, from anywhere, from edge to AI.

When transforming a CSP to become more data-driven, three main challenges arise:

- Architecture data silos cause bottlenecks and floods of data that occasionally overwhelm systems during surges of activity
- Enterprise data policy is fragmented and inconsistent across multiple point solutions and data sources
- Integrating data functions from point of collection at the network edge all the way through to AI applications requires detailed planning and testing

To manage these many and varied data sources, and some of the challenges above, different parts of the business have introduced multiple cloud services and applications that are distributed across on prem, private, and public clouds. Some sources are visible at an infrastructure level. Some are buried in an OEM.

Added to that are strategies that differ by department and the continuing challenge of shadow IT, and IT ownership itself being taken increasingly out of the hands of the CIO.

Although moving to cloud has solved issues of flexibility, scalability, agility, and cost, working with hyper-scaler cloud vendors has unfortunately created others. This can include:

- Dependence on a single cloud vendor creating a vendor lock-in situation
- Regulatory, security, and privacy concerns about moving certain workloads and data sets beyond the firewall (in some cases out of the country)
- A network cloud strategy from cloud service providers that could encroach on the opportunities of the CSP (including network slicing, virtual network functions, supporting edge workloads)

So, while CSPs need to work with hyper-scalers to meet their cost, usability, and time-to-value requirements, it also means they need to develop strategies to ensure they have relationship off-ramps, with the options to bring more sensitive workloads on prem. The answer points toward a hybrid cloud data model.

⁶ [New Internet of Things \(IoT\) connections in 2025 compared to 2019, Statista](#)

A Hybrid Cloud Data Model

In recent years the public cloud has been seen as a solution for many digital transformation strategies. However, a recent study has shown that while the cloud is a viable solution for start-up, expanding, and emerging use cases, its true cost on market capitalization is vastly underestimated⁷.

On a surface level, the flexibility and agility of a public cloud infrastructure, along with the ability to shift costs from CAPEX to OPEX makes it an attractive proposition. Particularly when you add to this the ability to reduce IT management, maintenance, depreciation, and upgrade costs. However, when it comes to enterprise data workloads, CSPs are finding that scaling down isn't quite as achievable as scaling up, meaning more mature workloads are spiraling in cost.

This is leading to questions about the best approach for their data architecture. Do they need a cheap cloud for high volume/low value data? A more feature-rich cloud for their low volume/high value data? Does personal data need to be stored on a private cloud? Does specific client data need something even more secure? Furthermore, can this data be processed or aggregated at the edge?

As a result, many CSPs are opting for a hybrid cloud architecture. This approach enables CSPs to unleash the full capabilities of on premises apps by bursting to the public cloud to leverage on-demand infrastructure for highly computational workloads. Choosing enterprise data partners that can deliver hybrid workload solutions reduces risk by leveraging data where it lies, mitigating data transfer risk, minimizing redundant data, and optimizing architecture use and spend.



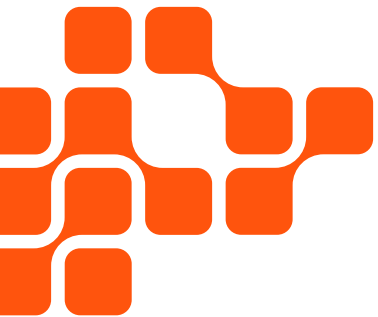
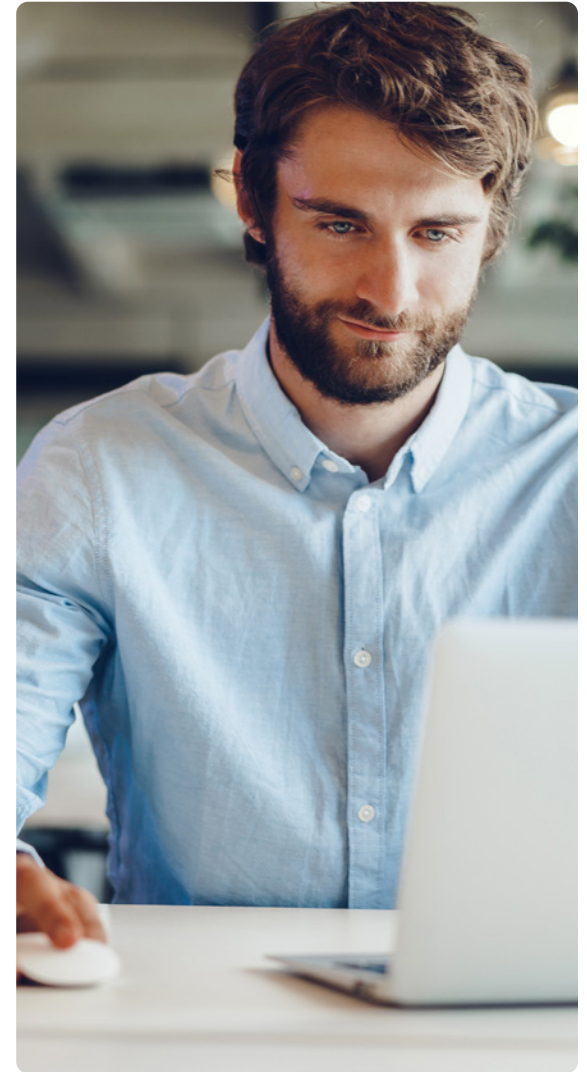
⁷ [The Cost of Cloud, a Trillion Dollar Paradox, Andreessen Horowitz](#)

What Does This Mean For CSPs?

Hybrid data architecture allows cost optimization and the ability to manage different data sources in ways appropriate to their sensitivity (privacy), or scale (volume).

Highly transactional data-in-motion workloads that control and run essential business operations stemming from log or time series data (for example online transactions, purchase order management, store operations, supply chain management) can be run in one optimized environment, while analytical based workloads (sales analysis and forecasting, market research, budgeting) that plan, solve problems, support decisions, and discover hidden insight can be run in another.

The benefits of the public cloud are clear, in terms of cost controls, speed to deploy, and agility. However, public cloud resources themselves can be expensive, while security, data residency, compliance and other control issues make the argument for an on prem or private cloud capacity. At Cloudera, the hybrid data cloud approach offers the best of both worlds, providing CSPs a choice where and when data is stored, used, and extracted opens options that lower cost and improve responsiveness.



Data Unification, Security, and Governance

With data protection legislation differing around the world, the location of Personally Identifiable Information (PII) is of critical importance. If it is moved outside of its jurisdiction this could lead to compliance issues, and it's why some service providers are moving this data away from cloud, and back to on premise.

Moreover, analytics and machine learning can become a risk if data security, governance, lineage, and metadata management is not holistically applied across the data lifecycle and all environments.

Inconsistent data access policies and lineage cause inaccuracies in audit logs. This can result in a compliance nightmare, a loss of customer trust, lawsuits and/or regulatory fines, not to mention brand damage. Gaps in data also lead to inconsistent or inaccurate insight, affecting decisions that impact the business' ability to innovate and differentiate.

A hybrid data cloud strategy, however, 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.

What Does This Mean For CSPs?

Cloudera's solution, Shared Data Experience (SDX), delivers an integrated set of security and governance technologies built on metadata, delivering persistent context across all analytics as well as public and private clouds. Consistent data context simplifies the delivery of data and analytics with a multi-tenant data access model that is defined once and seamlessly applied everywhere.

SDX reduces risk and operational costs by delivering consistent data context across deployments. IT can deploy fully secured and governed data lakes faster, providing more users of customer, partner, or enterprise data access their appropriate data, without compromise.

A true hybrid platform enables CSPs to discover

250%+

and

20–30

times fewer false positives

Leveraging Data From Edge to AI

By now we have established that CSPs need the ability to ingest, process, store, and analyze many different types of data.

However, while many speak about leveraging enterprise data, few do it well. They need to be able to manage this data regardless of where it lands: at the edge, in the data center, in any public cloud, or in a hybrid cloud. They need to be able to drive insights and use cases not only from data at rest, but also from data-in-motion and streaming data sources in real time.

When developing a data strategy that uses all data several constraints arise, the most significant being that CSP data sources are vast and varied, both in terms of content and form. They can include:

- Data generators: Network, billing and payments, care, third party, call center, survey, customer sentiment, and marketing
- Data types: Structured, unstructured, semi-structured video, and audio time series
- Data speeds: Batch, stream, real-time, push/pull, and sub/pub

- Data sensitivity: Private or public, PII, consented, unconsented, third-party client

In addition to this data complexity, CSPs are often attached to specific point solutions for data warehousing or business intelligence reporting. This will require a platform like the Cloudera, with open architecture that allows easy integration with existing systems.

It is also worth mentioning that retaining point solutions can come with risk. Many point solutions promote themselves as specialists in specific business process areas and the preferred solution — data warehousing, machine learning for example — can lead to costly integration issues, as can dealing with the multitude of unsynchronized and uncoordinated upgrades.

A hybrid cloud, on the other hand, brings open architecture, open-source software, open APIs, and an open storage formation. This means extensibility and flexibility, regardless of where the data is stored and workloads are run. Furthermore, a hybrid cloud also enables:

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.

- 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.
- 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.

What Does This Mean For CSPs?

Now, they can access data anytime, from any source and use it to power valuable use cases — whether that's improving customer experience, improving the effectiveness of marketing and sales, network optimization, or customer care optimization.

5G Network Analytics

Back at the dawn of CSP big data, over a decade ago, network monitoring consisted of reactive, passive monitoring tools — only alerting to problems after they happened, and never keeping the data. As compute and storage got more affordable, this opened the door to advanced, predictive, and streaming analytics.

Today, CSPs already apply network analytics for several business benefits, including customer retention, understanding usage behaviors, detecting fraud, and network service performance optimization. However, with the transition to 5G, this will bring with it a more distributed architecture, providing compute, storage, and analytical services at the data edge, and the ability to handle the business demands introduced by 5G.

Network analytics will play an important role in the design, build, and operation of complex 5G networks, as well as provide opportunities for new revenue streams. Designed to facilitate connected devices and automated systems, 5G will enable intelligent network and application services with connectivity to remote sensors, massive amounts of IoT data, and low latency data transmissions for all manner of applications, from industrial automation to healthcare to connected cars.

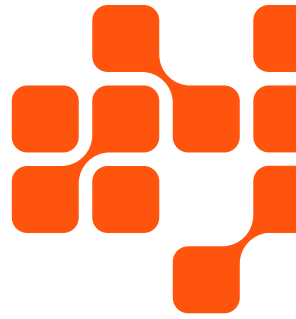
With 5G's ability to support massive connectivity across diverse devices backed by the distributed compute architectures, this creates the ability to translate the big data into real-time insights with actionable intelligence. However, network analytics will also play a role in the roll-out of 5G and network operations. Network data analytics capabilities will continually be collecting data from network functions that provide 5G connectivity as well as provide insights back to network function and business and operational support systems.

by 2035
5G will contribute

\$12.3T

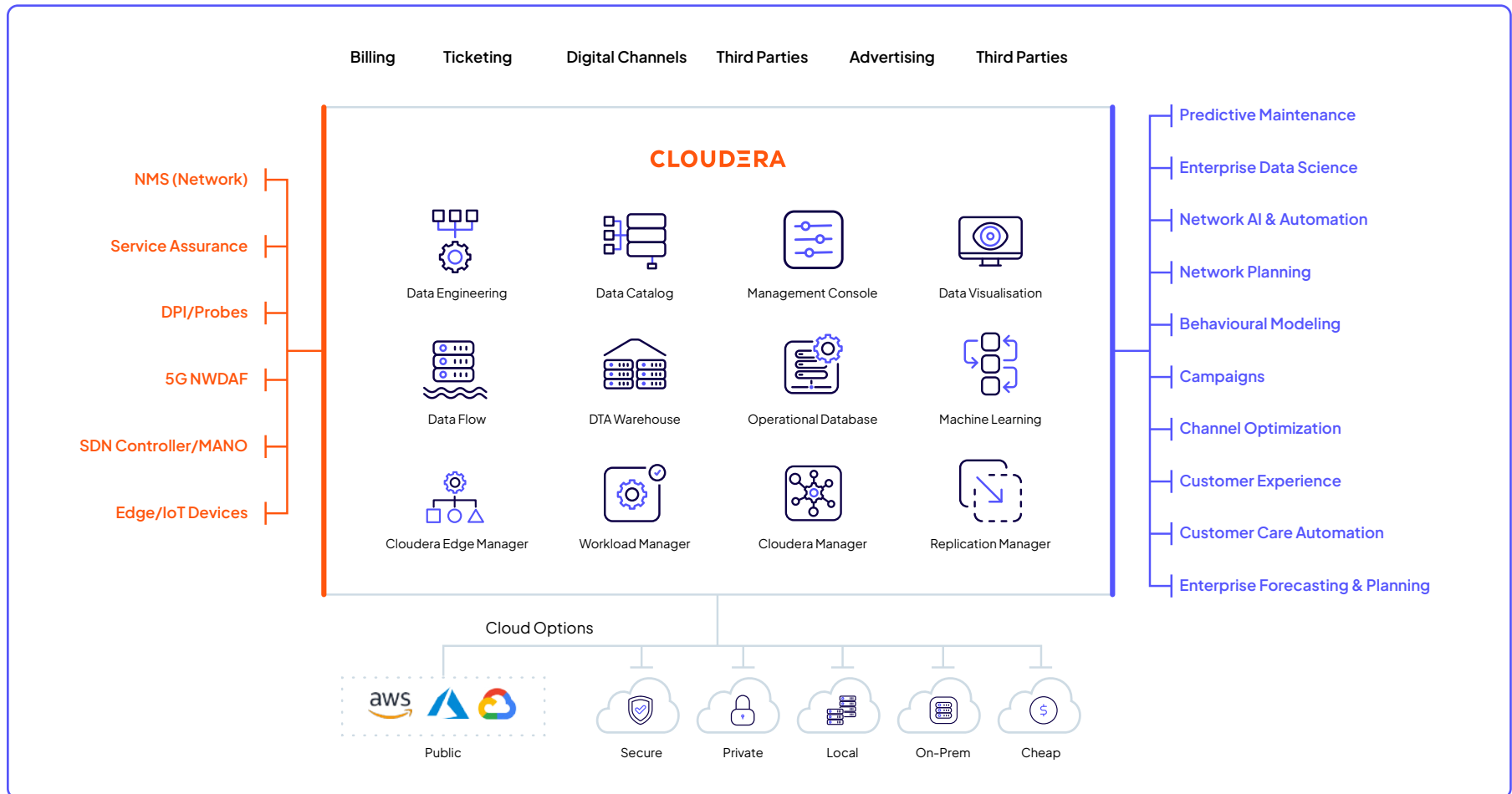
in global economic value⁸

⁸ IHS Matrix, The 5G Economy



Cloudera 5G Network Analytics

Network data sources combine with other application sources across multiple clouds to deliver myriad analytics and AI applications.



How Can Cloudera Benefit Telcos?

With Cloudera and a central, unified view of data, we can enable telcos to ingest, analyze, and make better use of streaming data and process large datasets in near real-time, as well as optimize their operations and empower you to deploy new machine learning use cases.

This Agile Telco Data Cloud brings your disparate data sources and systems together, ensuring all of your technology benefits from integrated security and governance — no matter the capability, or where the workload is run.

As your Hybrid Cloud Data Partner, we can provide you with a toolkit of capabilities that encompasses the end-to-end data lifecycle. This can include, but is not limited to, tools that can:

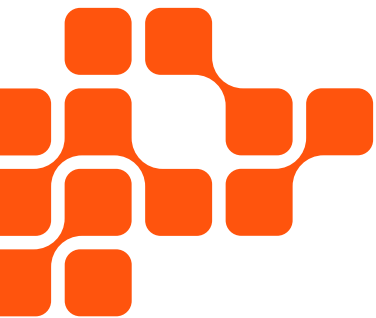
- standardize data as it moves around the business
- ingest, transfer, store and analyze streaming data
- network analytics and optimization
- security monitoring
- data engineering tools to track and provision workloads
- and a data science workbench for developing and deploying projects that leverage new technologies

We can provide communications service providers with the flexibility to work where you want to work — either on premise or taking advantage of any public cloud. It's a strategy for a true hybrid cloud, meaning you can take advantage of a range of deployment options, depending on your demand, security, compliance, or processing needs.

**70% –
90%**

of repetitive network maintenance work can be automated via machine learning, improving consistency and efficiency⁹

⁹ Deloitte, The age of Telecom Network Automation



The Benefits of Cloudera Agile Telco Data Cloud



The flexibility of hybrid cloud

Work where you want, on whatever infrastructure you choose, as well as centralize strategies to accelerate innovation and optimize costs.



An agile platform with the ability to ingest, store, and analyze streaming data from real-time data sources

Greater insight into operational silos and a centralized view of all data, enabling faster working and the ability to leverage data to create, deliver, and monetize data-driven business services.



An end-to-end, cloud-based platform for AI, ML, and other advanced data use cases

A platform that covers the whole data lifecycle – from analytics to storage, with the ability to add new feature sets as needed.



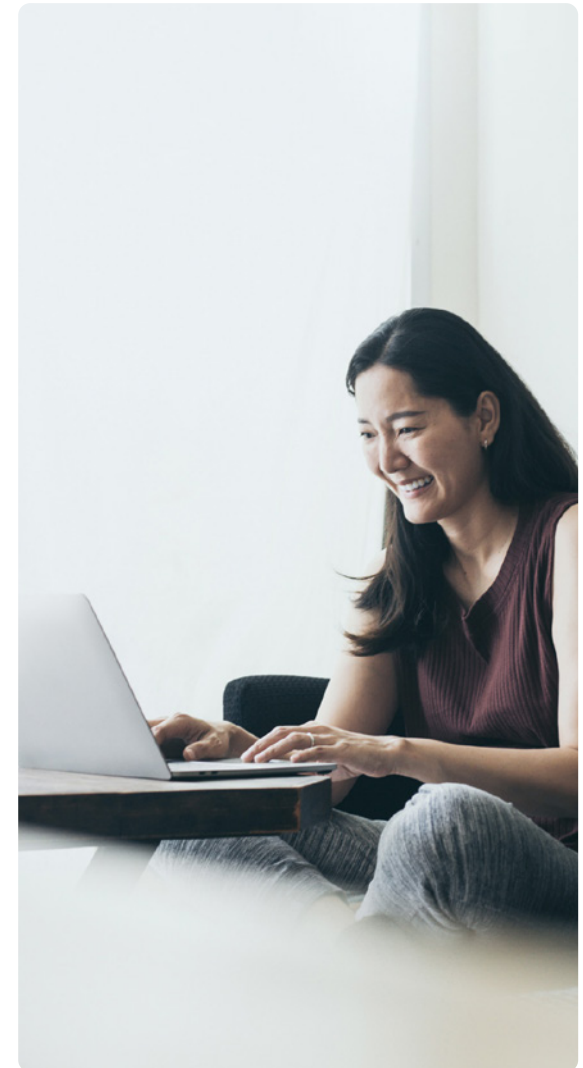
Open and interoperable to move data across clouds

Open architecture, interoperable with a wide range of ecosystems and data management vendors, along with support for all major cloud providers to avoid vendor lock-in.



Consistent security and unified governance

Our shared data experience (SDX) design philosophy provides consistent, enterprise-strength security with unified governance and compliance.



About Cloudera

Cloudera is the only true hybrid platform for data, analytics, and AI. With 100x more data under management than other cloud-only vendors, Cloudera empowers global enterprises to transform data of all types, on any public or private cloud, into valuable, trusted insights. Our open data lakehouse delivers scalable and secure data management with portable cloud-native analytics, enabling customers to bring GenAI models to their data while maintaining privacy and ensuring responsible, reliable AI deployments. The world's largest brands in financial services, insurance, media, manufacturing, and government rely on Cloudera to be able to use their data to solve the impossible—today and in the future.

To learn more, visit [Cloudera.com](https://cloudera.com) and follow us on [LinkedIn](#) and [X](#). Cloudera and associated marks are trademarks or registered trademarks of Cloudera, Inc. All other company and product names may be trademarks of their respective owners.

CLOUDERA

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

© 2025 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. EB_021 January 10, 2025