

---

**POWER YOUR BUSINESS  
TRANSFORMATION WITH AN  
ENTERPRISE DATA CLOUD**

**INSIGHTS**

**Insights-driven businesses grow at an average of more than 30% annually**

- Insights-Driven Businesses Set The Pace For Global Growth, Forrester Research, Inc., [10/19/18](#)

**By 2021, over 75% of midsize and large organizations will have adopted a multi-cloud and/or hybrid IT strategy**

- Predicts 2019: Increasing Reliance on Cloud Computing Transforms IT and Business Practices, Gartner, [12/13/18](#)

**By 2022, public cloud services will be essential for 90% of business innovation**

- Predicts 2019: Increasing Reliance on Cloud Computing Transforms IT and Business Practices, Gartner, [12/13/18](#)

**What does it mean to be insights-driven?**

Is data the new bacon? The world’s most valuable resource? The fuel that powers the digital enterprise? Whatever metaphor you choose, the point holds that the most successful companies in the world build their business around compelling insights derived from data. The five technology giants, known to investors as FAANG (Facebook, Amazon, Apple, Netflix, Google) use data to drive business strategies that surprise and delight their customers every day, and the market has rewarded them handsomely. As of March 2019, the market capitalization of these companies was [equal to \\$3.1 trillion](#).

Being insight-driven isn’t just in the domain of consumer tech. Banks rely on data to prevent financial crimes and reduce customer churn. The automotive industry is investing heavily in data and artificial intelligence to power autonomous vehicles. Manufacturing companies leverage data and analytics to predict and prevent machine downtime, speeding the route of new products to market.

While putting data and analytics at the center of your business strategy may sound obvious, it’s easier said than done. Digital transformation often requires a change in company culture that includes rethinking how people work with data, where it’s stored, and the tools used to access it. In fact, it can be such a struggle that many enterprises wind up stalling or abandoning their transformation strategies altogether.

This whitepaper will explore some of the challenges associated with being data-driven and how a new type of data platform - an enterprise data cloud - can address those issues.

**Why companies struggle with data**

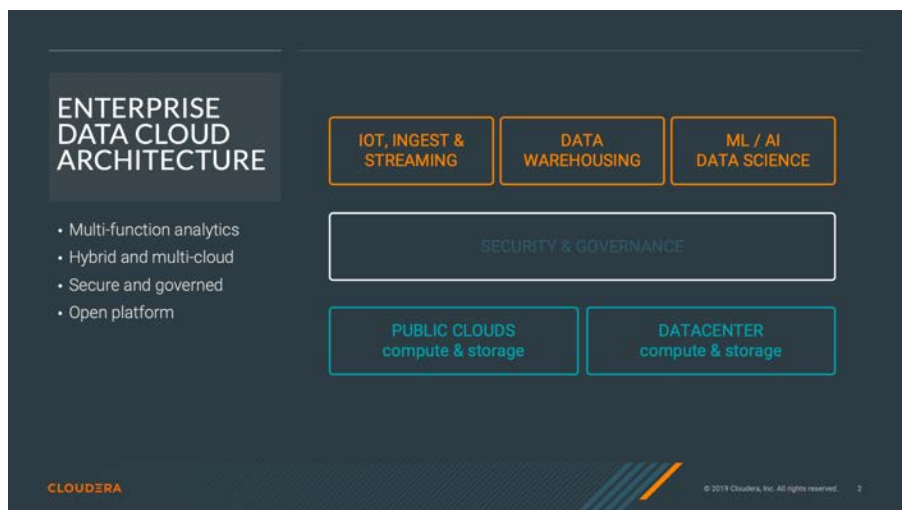
The most valuable and transformative business use cases - the ones powered by unique insights that can only be derived from a good data strategy - require multiple analytics frameworks, data science tools, and machine learning algorithms leveraging a diverse set of common data. It’s in these workloads that innovative enterprises are able to unlock immense value from their data and dominate their industries. It’s easier said than done, however, because the reality is traditional enterprises don’t operate this way. Instead:

- \_ Analytic workloads run independently in silos. Even newer cloud data warehouses and data science tools aren't designed to work together.
- \_ Data is everywhere—in data centers, public clouds, and at the edge—and enterprises have no practical way to run analytics or apply machine learning algorithms across their data.
- \_ Siloed analytics and data everywhere make a coherent approach to data privacy or IP protection nearly impossible. At best it forces onerous controls that limit business productivity and increase costs.
- \_ Traditional and even modern, cloud-based machine learning and analytics tools require the use of closed-source and proprietary storage and algorithms, meaning enterprises have to essentially hand over their most valuable asset -

their data - in order to extract meaningful insights

**Introduction to an Enterprise Data Cloud**

Being data-driven requires the ability to apply multiple analytics disciplines against data anywhere. To be able to process and stream real-time data from multiple endpoints at the edge while predicting key outcomes and applying machine learning on that same data set. To be able to take advantage of public cloud infrastructure for its agility, elasticity, and, increasingly, data gravity. And to be able to do all of this on an open platform where data security and governance are applied everywhere the data lives and analytics run. This is what the industry calls an enterprise data cloud.



An enterprise data cloud represents a new data management architecture but has commonalities with modern data platforms like Cloudera Enterprise and Hortonworks Data Platform (Cloudera and Hortonworks merged in January 2019). Key characteristics of an enterprise data cloud include support for:

**Multi-function analytics:** Lines of business must be able to rapidly identify interesting data, ingest and transform it, and deploy the analytic frameworks that make sense for the business. Addressing real-world business problems generally requires the application of multiple analytic functions working together on the same data. For example, developing autonomous vehicles requires the application of real-time data streaming from vehicle sensors (the edge), data warehousing to store, analyze and research potential automation, followed by the development and deployment of machine learning algorithms back to the vehicle to deliver the autonomous driving functionality and make intelligent decisions (AI).

**Hybrid and multi-cloud:** To provide the flexibility to deploy use cases in the environment that best suits an organization’s needs, an enterprise data cloud must operate with equivalent functionality on and off premises, supporting major public clouds and private clouds. Data has mass. It’s expensive to collect and

difficult to move, so it makes sense to analyze data where it naturally lives. That requires bringing the right compute resources to the data, whether that's in a public cloud object store or in an organization's data center.

In most enterprises, business-critical data can be found in multiple locations. For example, a company may choose to retain sensitive records that fall under compliance scope on premises, while at the same time running proof-of-concept applications using anonymized datasets in the public cloud. Hybrid and multi-cloud architectures provide greater control, putting the business in charge of where, when, and what type of analytic workloads to run.

Another important aspect of separating compute and storage is that it now provides lines of business the ability to choose between cost and performance. For example, a critical analysis may require additional compute resources from the public cloud, but for only a brief period of time. In this case, an emphasis is placed on speed rather than cost. Other analytic workloads - routine dashboards for example - can be cost-optimized by running on-premises or in a private cloud.

**Unified security, governance, and metadata:** In many enterprise environments, a data warehouse or machine learning tool will often have its own proprietary data catalog and policies that secure data and control access. In order to connect the two or more systems in service of a critical use case, the data team must extract the output from one system and reset the schema, security, and access policies, so that it's in compliance with all other systems. This takes time, introduces risk, and consumes resources that would better be applied elsewhere. The challenge is further amplified when organizations adopt hybrid infrastructures.

An enterprise data cloud delivers a common fabric for data and metadata that enables multiple analytic functions to work in concert and leverage data from across public clouds and on-premises environments. This includes unified security and governance applied to data wherever it lives to ensure the organization is meeting regulatory constraints. Companies need to be able to secure data and enforce the policies for who can access it and under what circumstances. They need to be able to track that access over time to be sure that what people do is permitted and answer questions about how and when data was used for auditing purposes.

Having a variety of analytic frameworks working together across different infrastructure and storage systems requires a consistent shared governance and security model; a single way to express a policy - applied across infrastructure and without regard to the analytic system - on who can touch the data and what they can do with it. Policies applied to data; security and governance that remains with the data as it moves between infrastructures. That's the only way to be sure that those policies are being uniformly applied.

**Open platform:** Enterprises want to avoid the mistakes of the past and not succumb to vendor lock-in. They want to own their data and the insights they unlock. They don't want their analytic decisions to be limited based on their

operating environment. Open platform means open source software, open compute architectures, and open data stores like Amazon S3 and Azure Data Lake Storage. It also means open integrations and open partner ecosystems. By avoiding lock-in, enterprises can control their data and their future.

### Use cases

Below are examples of companies that have successfully applied enterprise data cloud principles to their organization to unlock valuable insights and power their business transformation:

#### IoT and connected data

[Octo Telematics](#) provides telematics and data analytics solutions for the auto insurance industry. The company relies on data - 11 billion data points generated by five million connected cars driving more than 170 billion combined miles - to change how insurers assess risk, deliver crash and claim services, and manage customer relationships.

The ability to flexibly use major cloud service providers such as AWS, Google Cloud Platform, and Microsoft Azure means Octo customers can support transient but compute-intensive projects, such as testing new pricing algorithms or risk model development, on a usage-based commercial basis.

Since moving to an edge to AI data platform on hybrid cloud architecture, Octo has reduced time to market for new product launches by greater than 50%. The company has also significantly enhanced the customer experience by driving greater efficiency in claims processing and faster and more accurate fraud identification.

#### Digital transformation

[Deutsche Telekom](#) is a leading European telecommunications provider, delivering services to more than 150 million customers globally. The company built its modern data platform on the cloud to accommodate massive, streaming datasets while providing an environment that would deliver machine learning and fast analytics, and enterprise-grade reliability. A key component for success is the trusted, fully governed platform and ecosystem where users are empowered to exchange and analyze data and develop multi-function, data-driven applications easier and securely.

Deutsche Telekom's modern data platform is driving tangible results across the business:

- \_ Reduced fraud: Deutsche Telekom leverages large-scale, high-speed data processing and interactive querying to improve network quality and detect fraudulent activities in real time. Deutsche Telekom estimates this visibility helps reduce revenue loss caused by fraud activities by 10 to 20 percent.
- \_ Better customer satisfaction: They have a deeper understanding of customer

needs and desires. Deutsche Telekom has built a single enterprise view of customers, which has led to more targeted campaigns, generating revenues by the tens of millions of Euros while also reducing customer churn by five to 10 percent.

\_Improved operational efficiencies: The business is moving faster with the modern data platform in place, and overall operational efficiencies have improved by 50 percent as a result.

### **Predictive and preventative care**

[Clearsense](#) helps healthcare providers realize measurable value from their data. They have developed a secure, cloud-based healthcare data ecosystem that rapidly consumes data from an array of sources to provide a real-time view of any healthcare environment. The company's flagship product, Inception™, is designed specifically for the clinical, financial and operational needs of healthcare providers. Inception efficiently detects patient health deterioration, tracks key performance indicators to drive costs, and leverages data to improve operational efficiencies.

The edge to AI capabilities inherent in Inception led to the development of a clinical surveillance model that can predict whether a patient is likely to go into cardiac arrest within 4 to 12 hours and notify caregivers accordingly. The company is also applying predictive algorithms to reduce cases of sepsis.

### **Summary**

These customers highlight the scope of what's possible for the insight-driven enterprise. Companies must think first about the needs of the business and then design the insight-driven use cases that power the business strategy. Once use cases have been designed, you can start to think through your data and analytics needs and ultimately the environment in which your data management architecture will run. An enterprise data cloud helps companies extract meaningful insights from data and fuel innovation.

Cloudera offers the breadth of data analysis disciplines needed to solve the most demanding business use cases. Data disciplines from the Edge to AI—edge analytics, streaming analytics, data engineering, data warehousing, operational analytics, data science, and machine learning—all working together, securely, and operating across your choice of environments—data centers, multiple public clouds, and hybrid cloud.

Our current offerings - Cloudera Enterprise, Hortonworks Data Platform, and Cloudera Data Flow - offer the key capabilities of an enterprise data cloud—hybrid and multi-public cloud, multi-function analytics, shared security and governance services, and open-source platforms with choice of compute and storage.

The Cloudera Data Platform (CDP) - available in the second half of 2019 - will combine the best of Hortonworks' and Cloudera's technologies to deliver the industry's first enterprise data cloud. CDP will:

- \_ Empower customers to extend current Cloudera and Hortonworks deployments with native cloud services on AWS and Azure
- \_ Offer a full complement of open-source data management and multi-function analytics, with the agility, elasticity, and ease of use of a public cloud-like experience
- \_ Provide a single control plane to manage infrastructure, data, and analytic workloads across hybrid and multi-cloud environments
- \_ Extend SDX shared services to safeguard data privacy, regulatory compliance, and cybersecurity threats across all cloud environments
- \_ Be 100 percent open source, supporting enterprise objectives to avoid vendor lock-in and accelerate enterprise innovation

For additional detail on enterprise data cloud and CDP availability, please visit [www.cloudera.com](http://www.cloudera.com).