

Building a data-driven enterprise:

can your analytics span the hybrid cloud?

Contents

Driven by data, working in the hybrid cloud	4
Analytics in the hybrid cloud: three fundamental requirements	7
Analytics in the hybrid cloud: five considerations for successful implementation ...	9
What does the data-driven enterprise look like?	12
Cloudera and Microsoft: next-generation analytics in the hybrid cloud	14
Conclusion	16



“

Customer experience is on the verge of a new wave of transformation driven by robotic process automation, real-time interaction management, streaming analytics and AI. From fraud detection and cybersecurity to hedge funds identifying new investment opportunities, the ability to analyse data in near-real-time is transforming a wide spectrum of risk management use cases...

Driven by data, working in the hybrid cloud

For many organisations, the effort to become a data-driven enterprise is proceeding on three fronts: customer experience, risk management and operational excellence.

Customer experience is on the verge of a new wave of transformation driven by robotic process automation, real-time interaction management, streaming analytics and AI. From fraud detection and cybersecurity to hedge funds identifying new investment opportunities, the ability to analyse data in near-real-time is transforming a wide spectrum of risk management use cases. In operations, transformation is being driven by data from billions of sensors connected by the Internet of Things. Together, automation and vastly expanded data sets are revolutionising a wide range of processes, from clinical diagnosis and network management to logistics and predictive maintenance.

Two major trends – each reinforcing the other – underpin these efforts at data-driven transformation: the rise of public cloud and the exploitation of very large data sets.

Where data goes, analytics wants to follow

We live and work in a world where more and more data is being generated at the edge of networks. Increasing consumer digital activity and devices connected to the Internet of Things generate huge data sets. Much of this data finds its natural home in the cloud.

“Increasing consumer digital activity and devices connected to the Internet of Things generate huge data sets”

There are a number of reasons why this happens. For example, the low latency required for data generated by consumer IoT devices means that it's much more likely to be processed in the public cloud, rather than an on-premises data centre. In industrial IoT, the logic is different, but the outcome is the same: these data sets gravitate to the public cloud because factory-floor networks are typically air-gapped for security reasons.

Once the cloud becomes home to these data sets, data gravity starts to exert its influence. This is the process whereby an existing stock of data attracts more data and applications. Consider, for example, a large, rich data set of purchasing data stored in the cloud. If the data is sufficiently significant, it will attract two things into the cloud alongside it: applications to exploit the data, and other data sets that can usefully be analysed side by side.

Processing data close to where it lives also makes sense economically. In the cloud, you can do so without having to pre-purchase production level capacity upfront. Instead, you simply spin up the processing capacity you need, run the analytics workload and finish the job.

Running analytics in the public cloud can trigger significant benefits. Costs decline, insights flow more rapidly and a roadmap for future innovation – built on data-driven insights – falls into place. Indeed, Forrester has argued that not moving appropriate data sets to the cloud could come to be seen as an “extinction-level event for digital dinosaurs”.¹

But data gravity has its limits. For compliance and/or security reasons, some data needs to stay on-premise. At this point, the question for enterprises is no longer about choosing between public cloud and on-premises analytics. Instead, the challenge becomes how to deploy analytics in a way that spans the gap between on-premises and public cloud data sets.

The need for a hybrid data strategy

The transition to public cloud is proceeding rapidly in Europe. Some 56% of infrastructure decision-makers on the continent say their organisations have adopted public cloud. That's up from 14% in 2014.²

¹Forrester, Move your Big Data into The Public Cloud (June 2017)

²Forrester, Adoption Profile: Public Cloud in Europe Q2 2019 (April 2019)

“Eight in 10 of the enterprises using public cloud in Europe are also using an on-premises or hosted private cloud”

Today, nearly 30% of Big Data workloads are running in the public cloud, and their number is growing at pace: by around 20% annually. Ovum predicts that half of all new Big Data workloads will be cloud-based by 2019. ³

However, as we’ve already mentioned, the ties between certain data sets and the on-premises world remain strong. Eight in 10 of the enterprises using public cloud in Europe are also using an on-premises or hosted private cloud. Seven out of 10 enterprises in Europe describe their cloud strategy as hybrid. ⁴

There are good reasons for this. Not every data set is destined for the public cloud. Sensitive data connected with mission-critical processes, or subject to high levels of compliance and governance, still very often resides on-premises or in a hosted private cloud. For the foreseeable future, this will remain the case, and the vast majority of enterprises will persist with hybrid architectures.

As a result, the single most important challenge confronting most enterprises involves making the hybrid approach work, generating insights that span data sets resident on-premises and in the public cloud. To achieve this in a way that aligns with the analytics workloads of today and tomorrow, enterprises need to address three challenges:

- How do you establish consistent enforcement of security and governance policies across all infrastructures so that self-service access to data and analytics becomes possible?
- How can you exploit the full potential of data-driven transformation by driving machine learning and AI to production, regardless of infrastructure?
- How can you configure and process complex analytics queries with the kind of speed and flexibility that anticipates the requirements of real-time and near-real time analytics? ■

³Ovum, Harnessing the Cloud to Deliver Big Data Insights (June 2018)

Next:
Analytics in the hybrid cloud: three fundamental requirements

Analytics in the hybrid cloud: three fundamental requirements

Any solution that works across on-premises and public cloud also needs to take account of how analytics itself is evolving as existing trends accelerate and new possibilities emerge.

Self-service access

The democratisation of functionality in business intelligence has a long history. Freeing data from silos so that users can find it is the first step. This significantly improves productivity because users no longer have to wait for IT to deliver data sets – a process that can take months. Quicker access and self-service enable better, faster understanding of markets, business performance, competitors and opportunities: it has become a necessity for business analysts, superusers and line-of-business managers. In the public cloud or on-premises, your strategy needs to take account of this profound change so that the business can build upon existing progress.

The emergence of predictive and prescriptive analytics

Any strategy for hybrid cloud analytics also needs to take account of how rapidly analytics itself is evolving. For years, the focus of traditional analytics has been descriptive and diagnostic. This kind of analytics answers two questions: “what has just happened?” and “why did it happen?”

“In the public cloud or on-premises, your strategy needs to take account of this profound change so that the business can build upon existing progress”

Today, thanks to AI and machine learning, analytics is increasingly able to turn its attention to the future. As it does so, it is acquiring the potential to steer the business in the optimal direction. The first stage of this evolution involves predictive analytics. This enables a new question: “what is likely to happen next?” Predictive analytics answer this question to allow enterprises to anticipate routine events, whether the context is network traffic and, with that, capacity or the likelihood that the tyres on a truck in a delivery fleet will need replacement at a specific point in time.

“Today, thanks to AI and machine learning, analytics is increasingly able to turn its attention to the future.”

Next, there’s the potential for prescriptive analytics. Here, analytics asks questions like “What needs to be done?” and “What should we do to make x happen?” When paired with appropriate forms of automation, prescriptive analytics can allow technology – rather than humans – to optimize for improved outcomes on the fly. Examples include the ability to offer reluctant e-commerce customers a next best offer, or improved pricing. For trucks in a delivery fleet, predictive analytics predicts the required timing of a tyre change. Prescriptive analytics then go one step further, automatically scheduling a maintenance stop for the truck at a nearby depot as part of the regular delivery schedule.

Real-time on the horizon

According to IDC, more than a quarter of the data created by 2025 will be real-time in nature. A series of use cases are rapidly maturing, including logistics, supply chain management, precision medicine, fraud prevention and network optimisation. The ultimate goal is self-management, courtesy of decision-making algorithms that work upon a constant stream of data. ■

Next:
Analytics in the hybrid cloud: five considerations for successful implementation

Analytics in the hybrid cloud: five considerations for successful implementation

Be-coming a data-driven organisation is not easy. The upside is the potential for lasting competitive advantage, transformational innovation and powerful business insight. However, to reach that goal, organisations first need to tackle five challenges at the heart of all next-generation analytics operations.

Taken together, these challenges suggest one solution: a unified platform for data management, analytics, security and governance that opens up controlled access to data, minimises demands on IT, automates as much as possible and gives users access to the data and tools they need in a self-service manner.

“Can IT security professionals and responsible executives define user access rights in an organised, reliable and consistent way?”

Are you master of your own data?

Public cloud offers unparalleled benefits. But it won't provide you with a data strategy by default. It won't, for example, format your data or applications so that you can avoid being locked into proprietary templates of cloud service providers. In addition: do you know where your data lives? Do you know who has access rights at any one time? Can IT security professionals and responsible executives define user access rights in an organised, reliable and consistent way? In the cloud and on-premises, you need a platform that goes beyond analytics, enabling consistent data governance, security and control.

“...the technologies within standalone systems from different vendors aren't always compatible”

The need to break down silos

The basic processes of analytics include ingesting and transforming data, working with data warehouses and operational databases, and applying queries informed by data science. However, the technologies within standalone systems from different vendors aren't always compatible. Individual systems may run on different infrastructure, and often have their own definitions for security and governance. As a result, data often has to be replicated across systems to enable analytics. In turn, this makes integration and management a nightmare, sapping manpower and driving up operational costs. The solution is a silo-free approach: a single coordinated data platform that allows analysts to move seamlessly between different types of analysis while security and governance remains consistent in the background.

Automated and repeatable execution

Many organisations are increasing investment in AI projects and attempting to move them into production, realizing that operationalizing ML and AI are the key to achieving their data-driven goals and objectives. Across the enterprise, this typically starts to occur in a fragmented way. Knowledge is not always shared. Process isn't always formalised. Security, governance and compliance are gradually undermined. As a result, it becomes increasingly challenging to get data science projects into production. The solution required here is a repeatable, predictable way of building, deploying and scaling machine learning and AI solutions.

Functionality to the power of three

Today, analytics involves three end-user groups, each with their own goals, requirements and processes: data engineers, data scientists and business analysts. Each of these groups needs common basic functionality. Beyond this, however, needs diverge. Your solution for hybrid cloud analytics needs to give each of these three user communities what they need.

Next:
What does
the data-driven
enterprise
look like?

Help IT to be useful

Traditional IT organisations often spend 70%-80% of their cash keeping the lights on. So resourcing is always tight. Provisioning new on-premises capacity for analytics always takes time: but it can take longer if additional funding has to be secured. In these circumstances, IT can easily become the Department of No. One classic response involves shadow IT: a data scientist running a modest cluster under her desk or an unofficial public cloud instance – with all of the resulting security risks. Even when analytics are going to run in the cloud, it can take time for IT to provision compute power and data for analytics. The best way of avoiding these risks and time sinks – both on-premises and in the cloud – is to give IT the tools they need to provide enterprise wide data and analytics access, with consistent and appropriate security, governance and control of data. ■

Case Studies:

What does the data-driven enterprise look like?

Transforming business intelligence with the approaches discussed in this white paper allows enterprises to grow revenues, protect data and to connect products and services in new ways.

A successful platform for hybrid cloud analytics will accommodate the rapid evolution of analytics as a discipline and contribute to attaining all three business objectives. The following case studies illustrate how data-driven organisations are implementing these new strategies to achieve real competitive advantage:



Grow



Mission: uncover new possibilities, accelerate time to market, improve customer engagement, reduce churn

Bank Danamon, Indonesia's seventh-largest retail bank, generates revenues of \$1 billion annually. A challenger in a market dominated by larger institutions, Danamon has embarked on radical digital transformation. Results include a 30% reduction in fraud incidents and a 300% increase in the conversion rate for marketing campaigns, enabled by Cloudera's platform. On a daily basis, Danamon's data platform now ingests over 1TB of structured and unstructured data from more than 50 previously siloed systems and third-party external data sources.

For more on Bank Danamon, go [here](#)



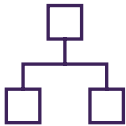
Protect



Mission: detect fraud, manage risk, mitigate security vulnerabilities and underwrite compliance with advanced analytics

Although it's not often discussed, telecom carriers suffer from significant levels of fraud: around \$30 billion-worth annually worldwide. ⁵Fraudsters typically benefit by setting up illegitimate premium rate numbers, and then encouraging telecoms customers to call or text. By using machine learning and AI to analyse network data, Deutsche Telekom has reduced fraud-related revenue losses by up to 20%. By applying the same technologies to the customer experience, Deutsche Telekom also reduced customer churn by between 5% and 10%.

For more on Deutsche Telekom, go [here](#)



Connect



Mission: understand, leverage and react to the vast quantities of data generated by the Internet of Things

“Data empowers us to understand the specific gene make-up of a given patient and what care path is best,” says Jawad Khan, director of knowledge management and data sciences at Rush University Medical Center (RUMC) in Chicago. Using machine learning, RUMC seeks recommendations for personalized treatments among years-worth of clinical notes, medical device data and genomic data. It no longer takes 48 hours to analyse this data. By using Cloudera's platform, RUMC now delivers near-real time analytics-as-a-service to hospital administrators, clinicians and researchers.

For more on RUMC, go [here](#)

Next:
Cloudera and
Microsoft:
next-generation
analytics in the
hybrid cloud

⁵ Communications Fraud Control Association, 2017 Global Fraud Loss Survey (April 2018)

Cloudera and Microsoft: next-generation analytics in the hybrid cloud

Only Cloudera delivers an enterprise data cloud for hybrid deployment that addresses the challenges of next-generation analytics. Ready for one-click deployment on Microsoft Azure or in a private cloud, Cloudera's enterprise data cloud allows you to manage, analyse, secure and govern all of your data, wherever it resides – on-premises, hybrid cloud or multi-cloud.

Microsoft's deep partnership with Cloudera is based on significant investment and a shared belief in making open source technologies work together seamlessly. Collaborating closely with customers, Microsoft Azure and Cloudera provide the expertise, guidance and support required to realise sustainable competitive advantage. Both companies provide you with a single point of contact, and work together from inception to delivery of your project – and beyond.

On Azure, Cloudera's customers have access to accelerated provisioning, scalable compute and storage plus an industry-leading compliance portfolio and state-of-the-art security, all available on a global scale. In addition, this partnership allows rapid integration of new and emerging Cloudera technologies.

Cloudera Data Platform on Microsoft Azure

Cloudera Data Platform is a single platform designed for the full spectrum of analytics, including real-time data ingestion, data engineering, data warehousing and operational data management and data science tools. You can run it in the location of your choice, in the

“Collaborating closely with customers, Microsoft Azure and Cloudera provide the expertise, guidance and support required to realise sustainable competitive advantage”

public cloud or a private cloud. Either way, this unified platform allows users to work on the same data sets with different tools. This removes the need for isolated data silos, unlocking the data and insights they contain and amplifying both through combination. By enabling the re-use and sharing of data between users, Cloudera Data Platform drives higher rates of analytics productivity, accelerating time-to-insight. Combined with the extensive portfolio of Azure platform services, Cloudera Data Platform on Azure is the ideal foundation for beginning the drive toward future IoT, AI, machine learning and data science initiatives

- Deployed on premises, in a private cloud, or in the public cloud of choice, Microsoft Azure
- Available as an enterprise edition designed to support multiple workloads, with special editions for data engineering, data science, and operational database
- Supports a wide range of workloads and integrates with Microsoft data and analytics software

Cloudera SDX (Shared Data Experience)

Cloudera SDX is a fundamental part of the Cloudera Data Platform, managing consistent security and governance: it's the data fabric that allows enterprises to avoid the risks associated with transient, fragmentary workloads running on isolated infrastructure. SDX organises data, enforces access policies, enables secure self-service and underwrites data governance. With Cloudera SDX, both Chief Security Officers and Chief Data Officers can rely on IT being able to define, apply and enforce consistent data security and governance as well as the control policies that enable agile, self-service analytics.

- By applying stateful, centralised and consistent security and governance, SDX enables hundreds of workloads to run against shared or overlapping sets of data
- Shared catalog defines what data is available and how it should be used, independent of where it lives
- Access control: a unified view of who has rights to what types of data

Conclusion

No-one said that becoming a data-driven organisation is easy. Mastering data management, unifying silos and offering analytics-as-a-service to different users across the enterprise is challenging. Balancing access with control requires fine judgement. The need to build a managed, compliant and secure architecture across hybrid cloud adds a further layer of complexity.

However, early adopters have taken their risks, learned their lessons and deployed their systems. As the experience of Deutsche Telekom, Bank Danamon and Rush University Medical Center demonstrate, the rewards of achievement can be significant. All three organizations have addressed the key strategic and operational requirements outlined here. In doing so, they have unlocked the full potential of analytics today and lay the foundations for the emergence of AI tomorrow. ■

Next:
Find out more



Want to discover how Cloudera and Microsoft Azure can help you to exploit the potential of next-generation analytics?

Learn more from Cloudera & Microsoft here:

<https://www.cloudera.com/partners/solutions/microsoft>.

CLOUDERA

At Cloudera, we believe data can make what is impossible today, possible tomorrow. We deliver an enterprise data cloud for any data, anywhere, from the Edge to AI. We enable people to transform vast amounts of complex data into clear and actionable insights to enhance their businesses and exceed their expectations. Cloudera is leading hospitals to better cancer cures, securing financial institutions against fraud and cyber-crime, and helping humans arrive on Mars – and beyond. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises



Microsoft (Nasdaq "MSFT" @microsoft) enables digital transformation for the era of an intelligent cloud and an intelligent edge. Its mission is to empower every person and every organization on the planet to achieve more.



IDG Connect is the demand generation division of International Data Group (IDG), the world's largest technology media company. Established in 2006, it utilises access to 44 million business decision makers' details to unite technology marketers with relevant targets from any country in the world. Committed to engaging a disparate global IT audience with truly localised messaging, IDG Connect also publishes market specific thought leadership papers on behalf of its clients, and produces research for B2B marketers worldwide. For more information visit: www.idgconnect.com