Expert Insights

—

A blueprint for data in a multicloud world

IBM **Institute for Business Value**

IBM

Experts on this topic

Tony Giordano

Senior Partner and VP,
CBDS Data Platform Services
Global Leader, IBM Services
linkedin.com/in/
tony-anthony-giordano-3111091
tonygior@us.ibm.com

Tony has more than 25 years of global professional services experience in performance marketing, big data and analytics, information governance, customer relationship management, and program management, with industry experience in financial services, life sciences, retail, and automotive sectors. He has authored two books on Information Management.

Mehdi Charafeddine

Associate Partner,
Executive Solutions Architect,
IBM Services
linkedin.com/in/mehdichara
mchara@ibm.com

Mehdi is a technical leader with a track record for delivering large complex projects that require a new way of looking at problems and solving them. As a strong Open Source advocate, Mehdi experiments with new technologies with a value potential to clients. This requires a unique combination of a deep understanding of software versus the risk/return ratio for IBM clients.

Dan Sutherland

Distinguished Engineer and CTO,
Data Platforms, IBM Services
linkedin.com/in/
dan-sutherland-4661a67
dsutherl@us.ibm.com

Dan is an expert at transforming a client's strategic business vision into innovative solutions while exploiting leading edge technologies. Dan is comfortable resolving complex data architecture and design issues and regularly advises on very large, high volume data-centric projects where sound data architecture and design techniques are key components to a successful solution implementation.

Becky Carroll

Associate Partner,
Cognitive and Analytics
Cognitive Business Decision
Support, IBM Services
linkedin.com/in/beckycarroll
rlcarrol@us.ibm.com

Becky has over 25 years of experience helping business leaders in large enterprises transform traditional business processes to new business models. She has led projects covering big data and analytics, cognitive and AI, digital and social business strategy and analytics, CRM, and customer service. Becky's work spans many industries, including High Tech, Telecommunications, Entertainment & Media, Financial Services, Airline, Automotive, Industrial Products, Life Sciences, Gaming, Education. Becky holds a BS in Electrical Engineering and Computer Science and a MBA in Marketing Management.

A well-planned and executed data strategy helps avoid unwelcome surprises in a multicloud environment.

Talking points

A data strategy and architecture addressing applications and data spread over multiple clouds are essential for companies operating in today's multicloud environment.

A business-driven data strategy should be fully integrated with a company's evolving multicloud architecture to keep data from becoming siloed and, therefore, not easily accessible to applications.

A strong data strategy will focus on business transformation opportunities, avoid "lock-in" to any particular cloud, and use DataOps, a method to automate preparatory work of data scientists, to derive the most insight possible from available data.

Moving data in a multicloud world

As companies move beyond initial forays into the cloud, many are experiencing the advantages—and the challenges—of a multicloud environment. Benefits of multicloud are manifest: increased ability to innovate, enhanced products and services, and agile business processes.[1]

In fact, according to a recent Institute for Business Value (IBV) study, 85 percent of companies globally are already operating in a multicloud environment.[2] By 2021, 96 percent plan to be using multiple clouds.[3]

Companies are remaking themselves as "digital" and making artificial intelligence (AI) central to their core business processes. As digitization continues, the movement to multiple clouds will increase, because it often involves obtaining cloud-based services from different cloud vendors. But the migration is unlikely to be as simple as merely shifting a current architecture to a new location in the cloud.

Each cloud vendor has its own architectural methods. How each organizes data could be incompatible with other clouds. If addressing such incompatibility is not accounted for in a well-designed strategy, a company may experience poor performance and higher-than-expected cost.

At the root of the challenge is the nature of data and its rapid growth. When companies expand the number of services and applications they use, their data grows exponentially. But it isn't the amount of data that is the real issue—it's where that data resides. And where it resides is often based on where it is collected and created.

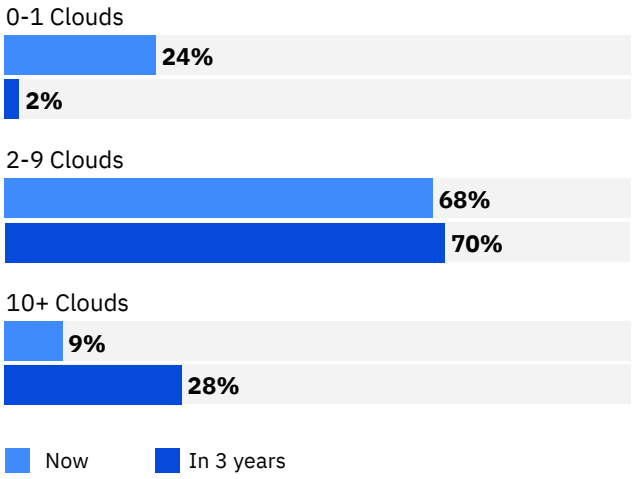
A data fabric helps make data visible—and usable—across an entire enterprise.

Such dispersed data created by distinct business units or functions is known as a "data silo." Silos drastically diminish the data's usability. The data may be redundant, but organized differently, and therefore difficult to correlate. There may be discrepancies in the data because of the way it was collected. And the data may be unknown or otherwise inaccessible to applications in other parts of the company.

Since almost all companies will soon be functioning in the multicloud reality, negative effects of data siloed within a company will only be heightened as companies spread processes, applications, and pockets of data across multiple clouds (see Figure 1).

Figure 1

Ninety-eight percent of surveyed organizations say they plan to use multiple hybrid clouds within three years



Source: IBV 2018 Multicloud management survey, Q2 & Q3.

Q2 How many cloud services and platforms are you currently using across your entire business?
Q3 How many cloud services and platforms are you planning to use across your entire business in the next three years?

These new circumstances lead to an important question. How can companies successfully navigate the complexity, cost and latency issues operating in a multicloud environment generates?

The answer? By properly defining a "data fabric" or "blueprint for data" in a multicloud world.

A data fabric to navigate multiple clouds

A data fabric is a conceptual representation or architecture of how data assets will be organized. This blueprint is a formal structure to define and view data across an enterprise, and is independent of particular infrastructure or cloud requirements.

A data fabric helps make data visible—and usable—across an entire enterprise. It guides how the data will be maintained and governed. It's founded on the concept that while business processes may change, the data underpinning them should be stable.

Data is grouped into collections called "hubs" or "lakes" for visibility and access. Data fabrics can help avoid or mitigate data silos, low reliability and scalability, reliance on legacy systems, and cost inefficiencies. But to work in a multicloud environment, a data fabric needs to accommodate different patterns of use.

Consider three examples of how a company might choose to distribute its workloads, applications, and data across different clouds.

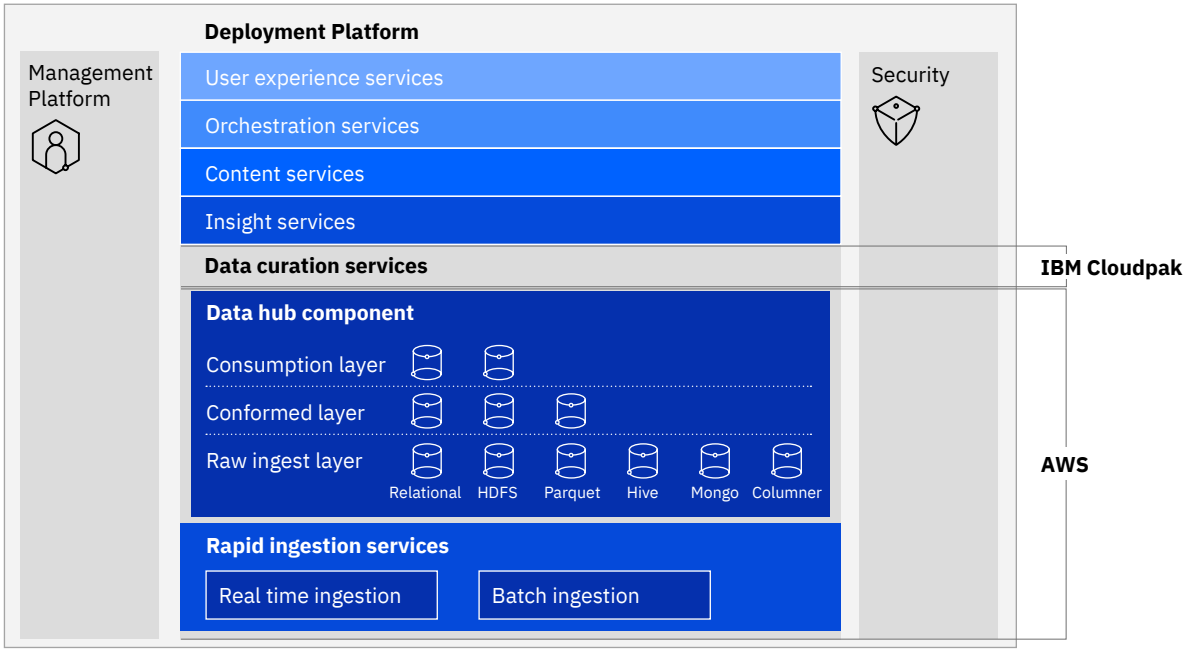
Example 1

In the first instance, a company wants to run an analytics package offered on IBM Cloudpak, using data it already has on AWS. Here, its data ingestion and curation services reside on one cloud, with only a specific analytics service calling on that data from a different cloud (see Figure 2).

—

Figure 2

Managing data across a multicloud by architectural component



To work in a multicloud environment, a data fabric must accommodate different patterns of use.

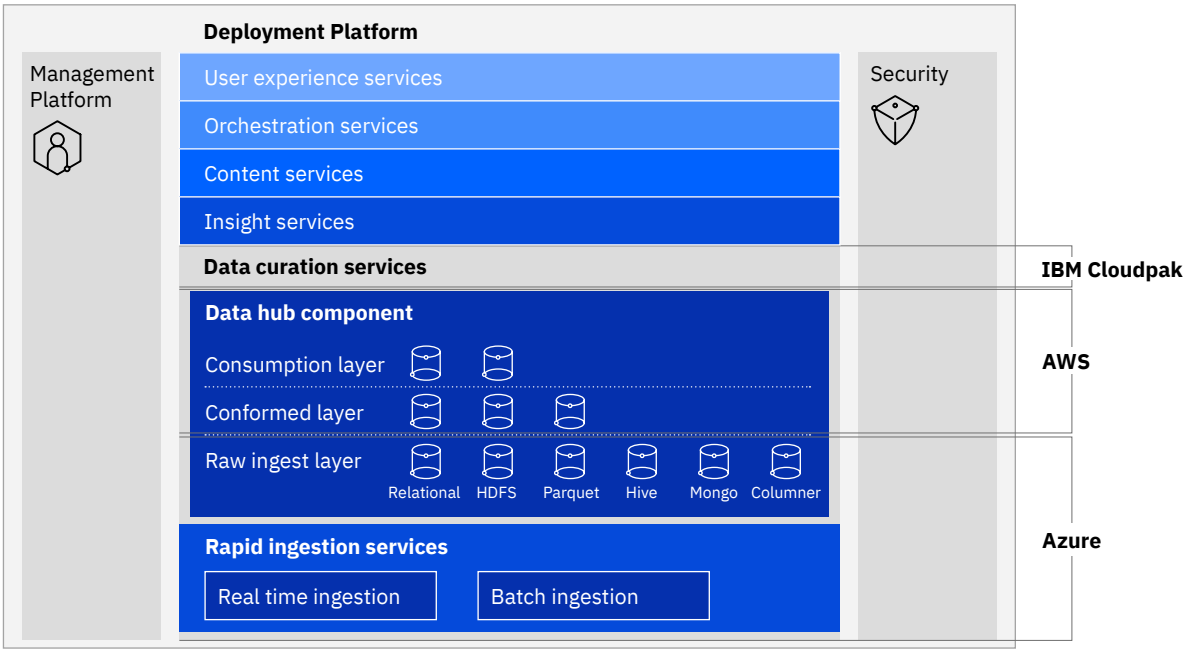
Example 2

In a second example, the company wants to take advantage of attractive pricing on applications offered on Azure, but keep other existing processes on AWS. Its ingestion of raw data and its ability to conform and consume data, then, are distributed over different clouds. It also subscribes to an analytics service on IBM Cloudpak (see Figure 3).

—

Figure 3

Managing data across a multicloud by data hub layer



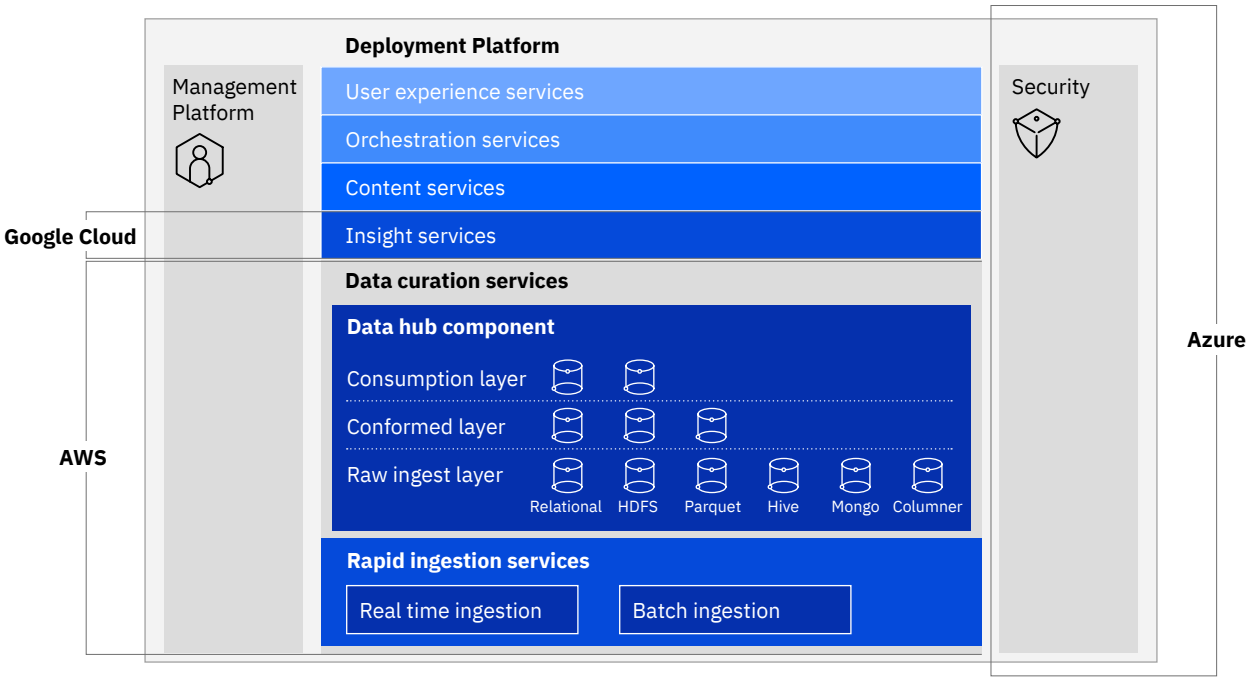
Example 3

In a third example, the company needs to have its various business divisions functioning on different clouds because of their unique needs, but chooses to keep a critical cross-business function such as security on Azure operating across all its other clouds (see Figure 4).

—

Figure 4

Managing data across multiple clouds



The right data strategy will anticipate change to help a company maintain its flexibility.

As the three examples illustrate, this new multicloud reality creates new options and opportunities. But it also creates new challenges for IT leaders. Indeed, recent market intelligence suggests that 82 percent of IT leaders are concerned about how they will connect all of these clouds with their traditional IT.[4] Seventy-three percent say they need better ways to move apps, workloads, and data more effectively across these clouds. And 67 percent worry about how they will manage this new mix of cloud and environments in a consistent way, across vendors, without introducing new security and compliance risks.[5]

To face these challenges, we have developed three guiding principles to help companies create and execute a data strategy for moving to, and operating in, a multicloud environment.

Principle one: Moving to the cloud is a time for business-driven transformation

Our data shows that cost reduction continues to be of strategic importance in investing in multicloud environments. However, some organizations experience difficulties and delays in capturing cost savings from their multicloud investments. Based on our experience with clients, many face cost increases—in some cases, by as much as 300 percent. The cause?

Moving to the cloud magnifies existing faults in a legacy IT landscape. Cloud vendors charge for data transfers and network usage at a much higher rate than most internal IT organizations. So siloed data—an inevitable reality for most legacy systems—generates enormous cost and performance issues as it moves from cloud to cloud to be used by various applications.

Therefore, a move to the cloud is the time to examine current business processes and data management, evaluate how changes to both can help keep costs in check, and best exploit the potential of the new multicloud environment. This is not simply an IT discussion or a restructuring of the current data. Rather, it should be a close collaboration between the line of business and IT, with business needs, value, and a line-of-business executive leading the way.

A related challenge is the rapidly changing technology landscape: today's tools may not work tomorrow. The transformation opportunity, then, also means that moving the legacy implementation will likely require new technologies, instead of continuing to reuse current ones, such as Hadoop.

Principle two: Anticipate and plan for future movement, changes and innovation

When migrating to a multicloud environment, companies should avoid "vendor lock-in." Cloud providers and offerings are changing rapidly, as are the technologies they support and offer. Which offering or service is best for a company could likewise change quickly.

The right data strategy will anticipate change to help a company maintain its flexibility. At some point, it's likely a company will want to move its data and apps from one cloud to another, scale its storage and processing without encumbrance, and store data in the optimal place for data science and workloads.

To anticipate change, a company's data strategy needs three key components: containerization, serverless capabilities, and pragmatic data design.

"Containerization" refers to a company packaging its data applications into "containers" that are not dependent on a single cloud implementation. Containerized applications can run across clouds and operating environments. This enables the organization to move their applications and data as their ecosystem evolves, especially as processes increasingly interact across clouds. The company will also not be entirely reliant on a specific cloud to run certain parts of its business.

Another aspect of change is the large volumes of data generated by new applications and services. It's practically impossible to scale the teams responsible for managing that data and the costs associated with its storage and processing. A serverless strategy can help.

Instead of reserving a set resource allocation—for example, 1,000 servers—a company's cloud vendor assumes responsibility for scaling resources required as its data and usage needs rise gradually, spike, or ease. The company pays for the resources it uses, and focuses on the business case and the code necessary to support it, rather than the IT resources required to run it. Together with containerization, this reduces the risk of new deployments in terms of complexity, skills, and costs. This approach helps a company become not just cloud vendor-agnostic, but also technology-agnostic.

A third component in anticipating change is proper data structure design. When designing data structures such as for traditional reporting, data science, digital, and operational use cases, it's important to keep the workload in the right place relative to the data to reduce network traffic and cost. Data design also should include data policies for security, compliance, and data lifecycle management across multiple cloud providers. These data design factors can be integrated into a set of managed and orchestrated data workflows that can span multiple cloud providers.

Yara: Data and a multicloud digital farming platform

Norway-based Yara, one of the world's largest fertilizer producers, wants to help fulfill the vision of a sustainable world without hunger. To lead the innovation of its core business models through digitalization, Yara is building the world's leading digital farming platform.

In building the platform, Yara focused on creating and realizing a cloud-agnostic strategy that enabled consistent data governance and data security. It also focused on DataOps—automating data functions that allow its data scientists to focus on data models and innovation.

The platform provides holistic digital services and instant agronomic advice to farmers across the globe, ultimately avoiding deforestation by increasing food production on existing farmland. Together, the Yara digital platform aims to cover 7 percent of all arable land worldwide.

The cloud-agnostic digital farming data platform follows a pay-as-you-go commercial model and provides Yara with two Data Services: Weather Data and Crop Yield as a Service. These accelerators are the first of many; an Open Innovation layer will enable Yara to create new ground-breaking algorithms providing farmers knowledge and decision-making insights.

Principle three: Add DataOps to DevOps

Companies should take a lesson from DevOps and how it has revolutionized application development over the past decade or so. Automating aspects of deployment has accelerated how developers test and prove their work. Data scientists now need a similar revolution, and they are finding it in DataOps: the automation of the kinds of work that slow them down.

Today, data scientists spend an inordinate amount of time preparing, validating, and cleansing data sources, then training their data models on them. They spend a surprisingly small part of their time designing the data models themselves—the highest value work of a data scientist. When you automate the data prep and training, you free the data scientist to bring to the business more insight and, ultimately, new value.

Data strategy, culture, and people

Just as book-filled libraries would be useless without a culture of reading, troves of data and the most innovative data tools require people with the right skills to put them to work, and the culture has to support them. People need data and the insights gleaned from it, and they need it in the specific context of their work. The same data can mean different things to different users. What insights a scientist finds useful will likely differ from those needed by a developer, a product manager, a marketer or a process guru.

Especially when AI is applied to huge data sets, meaning and context become critical. What are the right questions to ask, and is this the right data to provide the answers? If this is the right type of data, can we trust it? Can we trust both the AI algorithms operating on the data, as well as the training the AI is receiving—including the actual training data it is being fed?

A successful data strategy will include an inventory of needed skills and training, in addition to a long-term plan for cultivating a vibrant data culture: one where people are energized to get the most from data, and have the ability and organizational support to do so.

How to get started

Creating a data strategy integrated with your multicloud plans and applying the principles outlined above might seem challenging, but there are immediate steps you can take to prepare.

1. Build a business transformation strategy that explicitly factors in and takes advantage of multicloud capabilities, and manages data to exploit those capabilities.
2. Investigate technologies that keep data closer to where it will be used, and make sure it is properly governed across multicloud environments.
3. Identify any lower-value tasks consuming the majority of your data scientists' time and automate them. This will free data scientists to focus on more valuable and strategically important work.

Are you ready?

— Which core business processes have you prioritized and explored for transformation?

— What steps has your organization taken to build a data culture adapted to a multicloud environment?

— How will you navigate the world of regulations associated with data resident in different clouds?

About Expert Insights

Expert Insights represent the opinions of thought leaders on newsworthy business and related technology topics. They are based upon conversations with leading subject matter experts from around the globe. For more information, contact the IBM Institute for Business Value at iibv@us.ibm.com.

Notes and sources

1 IBV survey Assembling your cloud orchestra: A field guide to multicloud management (2018). Steve Cowley, Lynn Kesterson-Townes, Arvind Krishna, Sangita Singh. https://www.ibm.com/downloads/cas/EXLAL23W

2 Ibid.

3 Ibid.

4 IDC Cloud Forecast 2018-2020 – Market sizing + CAGR; BCG & McKinsey Study conducted for IBM – Multicloud + Priority concerns.

5 Ibid.

6 Source: IBM Institute for Business Value hybrid cloud survey (2016).

7 Internal IBM measurement, IBM Global Business Services.

8 IBM client experience

IBM Corporation
New Orchard Road
Armonk, NY 10504
Produced in the United States of America
October 2019

39028739USEN-00