



## Contents

<b>Executive Summary</b>	3
<b>Introduction</b>	7
The prize: The promise of cloud and the hyperscalers for data and analytics	7
Capturing the prize: Scope and purpose of this paper	7
<b>The fundamental mismatch: Traditional enterprises were not designed for a data-driven world</b>	8
<b>Sizing the prize: The benefits of cloud for data and analytics</b>	11
Innovation	11
Agility	13
Cost	15
Data management and control	16
<b>Big questions</b>	17
Strategy: What business and IT outcomes do we need to deliver through analytics and data?	17
Compliance: What rules must data and analytics meet and how do we ensure compliance?	19
Security: How do we keep data and analytics applications secure in the cloud?	19
Architecture: What is the overarching architecture for data and analytics in the cloud?	20
Partners: Who will be our strategic partners on our cloud journey?	21
<b>Moving from strategy to action</b>	22
Data	23
Data engineering and information management	24
A rich platform and data and analytics landing zone	25
Migration and transformation	27
Making the matrix work	28
<b>Conclusion</b>	29

# Executive summary

## The prize: Cloud and the hyperscalers promise so much for data and analytics, yet cloud is not right for everything

**The fundamental mismatch.** The world's global IT landscape is fast becoming data centric. In contrast, traditional enterprises were designed around business units and applications, with data a byproduct as opposed to a highly valued resource. The result is that traditional enterprises are out of step with the dynamic data-centric environment in which they operate, unable to “metabolize” and act on data at the required speed.<sup>1</sup>

**You will always be behind in some areas if you are not in the cloud.** The hyperscalers (Amazon, Google and Microsoft) and the venture capital industry are channeling huge investment into an ecosystem of data and analytics solutions, with cloud as the go-to deployment option. In augmented data and decision-making, for example, you cannot keep up if you are not in the cloud.

**But cloud is not right for everything.** Nevertheless, we should not view everything through the lens of cloud. Considerations of security, performance, cost and compliance can all make cloud the wrong choice for certain systems and data. Many organizations are taking a Cloud Right™ approach that utilizes both cloud and on-premises infrastructure. After the cost of transformation has been taken into account, the business case for a cloud-only approach may no longer stand; and the capacity of an enterprise to move to cloud will be limited by organizational inertia, skills availability and change bandwidth. So, what is the point of a strategy that cannot be implemented?

## Sizing the prize: The starting point for a cloud journey has to be an assessment of the business value that cloud-based solutions will bring to your particular enterprise.

**Innovation: putting data at the heart of the enterprise.** Becoming data driven of course includes making better decisions faster, based on more and better-quality data; but being a truly data-driven enterprise also entails a transformation that places data at the heart of the enterprise to drive revenue and other business outcomes. Cloud can play a key role in this transformation, by bringing innovation at multiple levels. Cloud enables new business models, for example constructing data ecosystems. Cloud also empowers machine learning (ML) and artificial intelligence (AI), through out-of-the-box models and services that support the entire ML/AI lifecycle. Cloud can open up the inner sanctum of data science, formerly held close by its high priests, professional data scientists, and put data and tools into the hands of citizen data scientists. Finally, cloud can help unlock the value in data, for example via open data sets.

**Agility: where cloud comes into its own.** The COVID-19 crisis has demonstrated the concrete value of agility: Whereas some enterprises were able to pivot seamlessly in reconfiguring operations, others struggled. Similarly, the faster pace of change across the entire economy places a premium on agility: variable costs

<sup>1</sup> See Supercharging your data metabolism, DXC Research, September 2021

vs. fixed costs; on demand vs. on order; fluid vs. static; modular vs. monolithic; test and learn vs. plan and specify. Agility is where cloud comes into its own, bringing scalability and enabling rapid experimentation. Cloud also increases data flow and fluidity, making data composable as building blocks for rapid assembly of new analytical and business capabilities. This can allow managers to ask critical questions that had not previously been considered.

**Data management: integration, automation and control.** Cloud brings many opportunities to transform and optimize data management, without which data centrality remains a dream. Cloud data management greatly facilitates integration of data from across the enterprise and beyond through data hubs and data meshes. Similarly, cloud brings the opportunity to reduce data proliferation arising from the “spaghetti and meatballs”<sup>2</sup> effect. Furthermore, through automation, cloud can impose extra discipline in data management. Finally, cloud increases the range of options for storing data, so that it can be optimized for distinct types of data and processing.

**Cost: Run the numbers early.** Cloud will reduce CAPEX and, if deployed correctly, bring greater control over costs. Yet, as highlighted in a recent paper by Andreesen Horowitz, it is by no means certain that cloud will reduce costs overall,<sup>3</sup> especially once an enterprise matures. Whether compute and storage are cheaper on premises or in the cloud will depend on your cost alternatives and on how workloads and data are used. In fact, larger savings are typically to be had from rationalizing applications and data prior to moving to cloud. Run the numbers early, since there are major implications for architecture — and make sure to account fully for networking, data transfer and transformation costs that are typically underestimated.

**Big questions: Before an enterprise can decide how and where to adopt cloud for data and analytics, answers to some big questions must be answered to act as guardrails for your journey.**

**Strategy: What business and IT outcomes do we need to deliver through analytics and data?** The most fundamental question is how you intend to compete and create value through data and analytics. For example, will your priority be analysis and exploration, or operational analytics? What are the most valuable data types? Since few enterprises have the luxury of focusing solely on the long term, your strategy should also resolve problems in the here and now — such as forthcoming regulation or performance problems that stakeholders want fixed.

**Compliance: What rules must data and analytics meet and how do we ensure compliance?** Too often cloud programs are expensively stalled by the need to retrofit regulatory rules. An essential early step is the identification of all regulations. Many have specific regulations governing cloud, while data governance rules such as the General Data Protection Regulation (GDPR) apply across all sectors. Years of gradual change may have led to some aspects of compliance being fudged, but cloud’s software-defined approach will demand the removal of all ambiguity. Since regulation now changes frequently, compliance should be future-proofed by employing a policy-driven software approach instead of hard-wired controls.

<sup>2</sup> Chris Swan, [Spaghetti and meatballs](#), Chris Swan’s Weblog, July 7, 2019.

<sup>3</sup> Sarah Wang and Martin Cosado, [The cost of cloud, a trillion dollar paradox](#), Future from a16z, May 27, 2021.

**Security: How do we keep data and analytics applications secure in the cloud?**

Although the hyperscalers deliver exceptional security for their cloud infrastructure, cloud inevitably increases the attack surface and you, not the hyperscaler, are responsible for protection of data in the cloud. You are likely to want to explore new security models to strengthen security and mitigate these risks — in particular zero trust, data-centricity, and modern identity and access management.

**Architecture: What is the overarching architecture for data and analytics in the cloud?** Choices are required around architecture. First, different primary business use cases will need different patterns for ingesting, storing and managing data. Second, there is a variety of architectural models to choose from. The lakehouse, the data integration hub-data mesh and the cloud data warehouse all have their advocates. Although new ventures will almost certainly be born in the cloud, most legacy firms will have some analytical applications and data stores that are not suited to cloud, making a hybrid of solutions inevitable.

**Partners: Who will be our strategic partners on our cloud journey?** The hyperscalers are not all the same when it comes to data and analytics. For example, they bring different capabilities in compute speed and scale, open data sets, artificial intelligence models (say, for speech and text) and services for the model development life cycle. Moreover, specialist vendors fill important niches in the landscape. Furthermore, the journey involves many novel technologies, and skills are scarce. As a result, making the right choice of IT service providers is a critical success factor in reducing risk, cost and timeframes.

**Strategy to action: Once you have addressed the big questions, you can move from strategy to action.**

**Data.** In resolving how and where data should be stored and processed, a whole range of factors has to be weighed: business needs, security, compliance, cost and technical performance. Moreover, these considerations are frequently opposed, making tradeoffs essential. Significantly, egress fees and data gravity (the pull that data exerts to attract other data) mean that your decisions will have long-term implications. Because data will most likely end up being stored in more than one location, a common metadata layer is vital to prevent the cementation of new silos.

**Data management.** Cloud presents an opportunity to enhance data and information management, but you have a narrow window to get things right. You will need to map out the principal data flows, with separate patterns defined for each — for example, operational decision support, self-service analytics and data science. Notably, these data flows may operate completely differently in cloud vs. on premises, for instance with extract, load and transform (ELT) potentially replacing extract, transform and load (ETL). Likewise, you need to define the tools, processes and governance that you will use to manage information across the end-to-end lifecycle. Key areas include data augmentation, metadata, data lineage, data catalogues and archiving.

**Rich data and analytics landing zone.** A vital prerequisite for a cloud data and analytics program is a richly functional enterprise cloud platform that includes

common components, for example single sign-on, networking, zero-trust security, monitoring and DevOps. Even though such a platform will typically take 3 to 6 months to build, this is a matter of “going slower to go faster,” ensuring each later project does not duplicate effort and delivers a more standardized solution. Similarly, a rich data and analytics landing zone is required that should contain information management and governance tools and processes such as data cataloging, data protection and data lineage. Without this, you will miss the opportunity to standardize and to bake in compliance by design and security by design. You will spend more, as successive projects reinvent the wheel, each with their own vision of roundness.

**Migration and transformation.** In order to build out from this first landing zone, you will need a roadmap for scaling, as well as a robust and repeatable approach to migration, transformation and archiving, which is often best achieved by a factory approach that reduces costs and enforces standardization. As when moving operational systems, for each application or data store, you must decide if business goals are best met by simpler rehosting and replatforming, or more complex refactoring and rearchitecting. Even when the goal is transformation, you will have to choose whether to transform and migrate or to migrate and refactor/reengineer, since the cloud offers many tools to facilitate transformation and data cleansing. Finally, a robust approach is needed to archiving systems and data so that the full benefits of moving to the cloud are realized.

**Making the matrix work.** Since a successful cloud journey necessitates balancing numerous factors — cost, risk, compliance, security, technical performance and business outcomes — you need organizational structures and governance to make the matrix work. At the working level, compliance and security SMEs should be embedded within teams to ensure compliance by design and security by design, with architects likewise dispersed among project teams to ensure adherence with architecture and design principles. At the next level up, a cloud business office (CBO) applies integrated decision making where individual scrums cannot resolve issues or where an issue spans many teams. Above the CBO, an executive forum can act as a point of escalation but the buck needs to stop with a single accountable executive.

**Conclusion: Cloud will play a vital role in reorienting the enterprise around data — a vital feature of the 21st century enterprise.**

For most enterprises, cloud will certainly form a key strategy in the transformation to data centrality. Yet, we must avoid starry-eyed thinking that fails to accept this fact: Migrating to cloud is difficult, such that for large, enterprises with complex legacy systems, a hybrid solution is inevitable. Success will depend on clear top-down thinking to answer the big questions and to develop a strategy to scale adoption. The top-down approach, however, has to be tempered with bottom-up planning and action, since complex issues can only be addressed by working through the detail and learning by doing.

Competitive advantage increasingly stems from ability to exploit data to automate processes, model scenarios and make decisions at speed.

## Introduction

### The prize: The promise of cloud and the hyperscalers for data and analytics

The metaphor “data is the new oil” is mistaken at many levels. Data is not a commodity: You cannot ask for a liter of data. Equally, unlike oil, data typically decreases in value with age, does not get used up and often becomes more valuable when augmented with other data. Even so, the phrase captures the essence of the transformation that is occurring in the economy — or has occurred — where competitive advantage increasingly stems from ability to exploit data to automate processes, model scenarios and make decisions at speed.

In this new data-driven world, the companies that stand out are the hyperscalers, Amazon, Google and Microsoft. In the first place, they are the archetypes of business success through exploiting data, via APIs, machine learning and data streaming. Second, they have made many of these very same technologies available as reusable cloud services and are continuing to invest hugely in extending these services, spurred by revenue of \$25 billion and growth rates in excess of 30 percent in Q1 2021.

Inevitably, enterprises are looking at the hyperscalers and asking, “How can we emulate them by adopting the same technologies?” “How can we capitalize on the unprecedented investment that the hyperscalers are making in cloud services?” “Where will cloud add most value to our business?” Many firms already use cloud for data and analytics: A survey by MIT found that 63 percent of enterprises employ cloud services widely in their data architecture.

Nevertheless, just because cloud is the right answer in many areas does not mean it is right everywhere. Moreover, moving to cloud is not straightforward. Numerous complex decisions are required — for example, whether to move processing to the data or the data to processing. Further, it is not as if most enterprises have already cracked the wicked issues of data management. As a result, many firms fail to achieve cloud transformation that results in a data-driven enterprise.

### Capturing the prize: Scope and purpose of this paper

This paper, the third in the Leveraging the Hyperscalers series, provides a framework to harness the value of cloud-based data and analytics services. Both technologists and business people keen to understand the what, where and why of cloud for data and analytics will benefit from applying the paper’s framework.

<sup>4</sup> See David Rimmer et al, Constructing Cloud-Native Business Capabilities, LEF, June 2020 and David Rimmer, A Cloud Journey to Deliver Business Outcomes, LEF, November 2020.

**Figure 1** gives a frame of reference that defines the aspects of data and analytics that are addressed.



**Figure 1** – Aspects of data and analytics addressed in this paper

The paper is structured in four sections:

- **The fundamental mismatch** explains how a mismatch has developed between an increasingly data-centric world and traditional enterprises with systems and data centered on business units and applications; the section describes how cloud can help overcome this mismatch.
- **Sizing the prize** provides a framework to decide where cloud-based data and analytics can add most value to your business through agility, innovation, cost and data management.
- **Big questions** identifies the significant questions that must be addressed in adopting cloud for data and analytics, for example about architecture, selecting the right partners and designing in security and compliance.
- **Moving from strategy to action** explains how to implement key aspects of cloud-based data and analytics, such as data engineering and information; a richer, more functional platform and landing zone than typically built; migration and transformation approaches that will scale; and governance that makes the organizational matrix work.

### The fundamental mismatch: Traditional enterprises were not designed for a data-driven world

**The world's global IT architecture is becoming data centric.** The last 20 years have seen a transformation of the global IT architecture. In place of large-scale systems designed to process and record transactions, there are “massive, complex systems built around data.”<sup>5</sup> A defining feature of this new landscape is that it is data centric as opposed to application centric. Vast dataflows are exchanged through APIs (data services), processed in real time and used to feed ML/AI. Much of the action happens outside the enterprise, initiated by a machine or a customer holding

<sup>5</sup> Matt Bornstein, Martin Casado & Jennifer Li, *Emerging architectures for modern data infrastructure*, Future from a16z, October 15, 2020.

Traditional enterprises are completely out of kilter with the rapidly changing data-centric world in which they and their customers operate.

one of 4.88 billion mobile-phones:<sup>6</sup> at the edge and in SaaS applications, social media, the cloud and platforms. John Gage's meme "The network is the computer"<sup>7</sup> is becoming reality. Furthermore, data volumes are growing at a tremendous rate. IDC's Dave Reinsel predicts: "The amount of digital data created over the next 5 years will be greater than twice the amount of data created since the advent of digital storage"<sup>8</sup>.

It should be no surprise, therefore, that the firms that are succeeding in this data-centric world are themselves data centric: the hyperscalers (Amazon, Google and Microsoft), media streamers (Netflix and Spotify), the Chinese super-apps (Alipay and WeChat) and the social media platforms (Facebook, Twitter and TikTok).

**Traditional enterprises are out of step with the data-centric world.** The issue for enterprises that predate this transformation is that their systems are just not built this way. Typically, each business unit has its own applications, and each application produces its own data, with its own data schema. Data silos and a data sprawl are inevitable, making master data and data quality a never-ending challenge. Even now, after years of effort, how many enterprises have a true single customer view?

In addition, data is typically processed on a nightly cycle and requires substantial transformation, modeling and integration to be of value. This delays the time-to-value of data, and the often-complex data transformations can be brittle to change as they lack adaptive learning rules. Furthermore, their focus has been on numerical and structured data; the value of video, images, documents and speech has barely been touched.

The upshot is that traditional enterprises are completely out of kilter with the rapidly changing data-centric world in which they and their customers operate, and are unable to integrate, process and act on data at the speed and effectiveness required. While theory might say that data is one of a firm's most valuable assets, reality dictates that most firms face "a huge 'data value gap:' the difference between what their data is worth at present and what it would be worth if it were classified, associated with other data and made accessible to those who need it in a timely manner."<sup>9</sup>

**The hyperscalers are prime exemplars of data-driven enterprises and have profoundly shaped today's data-centric landscape.** In contrast, the hyperscalers were built from the ground up for big data, APIs, microservices, event streaming, rapid release cycles, machine learning and unstructured data — indeed, their services have fashioned, almost defined, this world. The typical 21st century scenario of applying a model to a stream of data to tailor a digital journey in real time, is highly problematic in a traditional data architecture — just applying an accurate timestamp to integrate data represents a challenge. Yet this scenario has been meat and drink for Google (ad-serving) and Amazon (recommendations) since Day 1. It is only natural, therefore, that enterprises should seek to emulate the hyperscalers and capitalize on the investment that they are making in productizing their own technology in the form of reusable cloud services.

Cloud can assist in a reorientation to become data centric. As we will explore later, the cloud directly addresses many of the challenges that traditional enterprises face

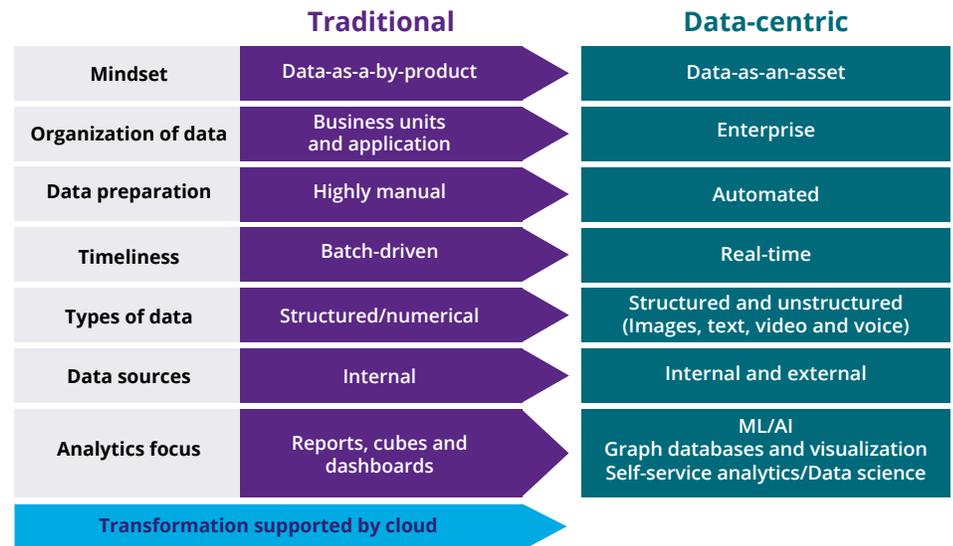
<sup>6</sup> [How many smartphones are in the world?](#) BankMyCell.

<sup>7</sup> John Graham-Cumming, [The network is the computer: a conversation with John Gage](#), Cloudflare Blog, 11 July 2018.

<sup>8</sup> [IDC DataSphere and StorageSphere forecasts](#), March 24, 2021.

<sup>9</sup> The power of digital assets and intangibles, LEF, July 9, 2019.

in their reorientation to become data centric in a data-centric world. For example, cloud naturally supports a whole range of foundational capabilities, including data streaming, event-based processing, integration of data from disparate sources, real-time processing of large volumes of data, and new analytical techniques such as ML/AI and graph databases. This transformation is illustrated in **Figure 2**.



**Figure 2** – Cloud can support the transformation from a traditional enterprise to a data-centric enterprise.

**If you stay on premises you will always be behind.** Another reason for adopting cloud is the size of investment that is being made in the ecosystem of cloud solutions for data and analytics. Competitive pressures, coupled with tremendous profit margins and revenue (more than \$25 billion in Q1 2021 alone), are leading the hyperscalers to make unprecedented investment in data and analytics services. In addition, specialist data and analytics firms are attracting huge investment: “The top 30 data infrastructure startups have raised over \$8 billion of venture capital in the last 5 years at an aggregate value of \$35 billion, per Pitchbook.” These firms typically develop solutions for the cloud first, often leveraging hyperscaler services and only later releasing on-premises versions (if at all). Much of the ecosystem of leading-edge data and analytics lives in the cloud.

A particular domain where the hyperscalers are carving out a lead is augmented analytics which Gartner defines as “the use of enabling technologies such as machine learning and AI to assist with data preparation, insight generation and insight explanation to augment how people explore and analyse data in analytics and BI platforms.”<sup>11</sup> The hyperscalers have so many customers on their platforms that their ML and AI tools are trained by huge volumes of data, resulting in a depth of automation that others cannot attain. Similar dynamics are driving the automation of many data management tasks as well. Consequently, in many areas if you stay on premises, you will always be behind the state of the art.

<sup>10</sup> Matt Bornstein, Martin Casado & Jennifer Li, *Emerging architectures for modern data infrastructure*, Future from a16z, October 15, 2020.

<sup>11</sup> [Gartner glossary](#).

The point of departure for a cloud journey has to be an assessment of the business value that cloud-based solutions will bring.

## Sizing the prize: The benefits of cloud for data and analytics

The point of departure for a cloud journey has to be an assessment of the business value that cloud-based solutions will bring in terms of innovation, agility, cost and data management and control (Figure 3).

**Business benefits of cloud for data and analytics**

Innovation	Ecosystem collaboration	Monetization	Open data sets/ external data	Data origination in the cloud
	Real-time event-driven analytics	Unstructured/ unmodeled data	Machine learning and artificial intelligence	
Agility	Scalability	Experimentation	Data flow and fluidity	
	Composable data	Data democratization	Organizational flexibility	
Cost	Compute	Storage	Automation of data management	
	Rationalization of tools	Data transfer	Early stage vs. late stage	
Data management and control	Data discipline	Data management tools	Circumventing Conway's Law	
	Less data proliferation	Choice of databases and storage options		

Figure 3 – Potential business impact of cloud on data and analytics

### Innovation

#### Business models

**“The value of data increases in ecosystems.”<sup>12</sup>** In many industries, it is common for organizations to come together to collaborate and share data. For instance, in aerospace and defence, projects typically bring together several contractors, each with their own subcontractors; while in healthcare many parties may require access to patient data and large data sets, such as a bank of genomic sequences. Putting data in the cloud simplifies data sharing from a technical perspective and creates a neutral zone for collaboration with tight control over access to data, even down to the field level. As the concept of data ecosystems grows in importance, more and more industries will seek to collaborate with partners and customers through cloud-based data sharing, often facilitated by open data sets.

#### Cloud facilitates monetization through pay-per-use APIs and analytics services.

Cloud can be used to make data and algorithms accessible for customers as a service. For instance, data services can be exposed as pay-per-use APIs that are run using on-demand cloud services. Alternatively, firms can use cloud to provide analytics services. For example, an energy exploration company might offer oil and gas firms a bring your own data (BYOData) service, along the lines of bring your own bottle (BYOB): the exploration firm provides algorithms and datasets (the food) and data processing platform (restaurant), while the customer brings their own proprietary data (bottle of wine).

<sup>12</sup> [Five technology trends in 2020 poised to transform the future of work, according to DXC Technology](#), DXC, November 19, 2019.

## ML and AI are just easier in the cloud.

### *Modern analytics*

**“No cloud means no AI.”**<sup>13</sup> Liam Maxwell, the former UK Government Chief Digital Officer, coined this phrase in his reflections on how to use technology as a catalyst to beat the “Department of No.” While somewhat of an overstatement, this captures the spirit of the point that ML and AI are just easier in the cloud. For example, ML typically requires extra-large processing capacity to train models with extensive data sets. Where traditional CPU-based processors could take days for this task, the hyperscalers provide specialized compute resources that are tailor-made for ML. Except for those running really heavy workloads (in which case cloud may be too expensive) the added flexibility of cloud will make this the natural choice for most. In addition, the hyperscalers offer model life cycle tools to automate MLOps — the end-to-end process for building, training, testing and deploying ML models. In addition, the hyperscalers offer software development kits (SDKs) and APIs so that you can embed ML/AI in operational processes.

The hyperscalers offer many services for developers of ML/AI. The hyperscalers began with the philosophy of democratizing ML/AI so that business experts could develop their own models without the need for a “translator.” A low-code/no-code approach enables citizen data scientists to access capabilities directly through a wide range of prebuilt models for contact center analytics, common statistical analyses, language translation and text extraction. More recently, the hyperscalers have extended their scope to serve specialist data scientists with tools to develop bespoke models and, therefore, prise systems away from under analysts’ desks and into the cloud.

### **For many enterprises, moving to a real-time event-driven architecture is key.**

Traditional data architectures were designed for batch processing of static data (data at rest). Conversely, event stream processing acts on a continuous flow of data (data in motion). Data streaming is valuable where immediate action is required — for example, payment fraud checks, notifications from IoT sensors and real-time customer interactions. Without cloud-based, event-based tools, it is hard to join data from several sources with accurate time stamps and correlations. The cloud can speed adoption of event streaming, since the hyperscalers offer cloud-native data streaming services and data streaming specialists offer their own cloud-based services.

### *More data, more data value*

#### **Cloud storage and processing capabilities help unlock the value of unstructured/unmodeled data.**

A figure that is commonly cited is that 80 percent of all business-relevant information originates in unstructured data. The hyperscalers all offer solutions to unlock the value of unstructured data — or unmodeled data, which is perhaps a better way to think about it. Examples include cheap object storage, natural language processing and visual analytics.

**What happens in the cloud, stays in the cloud.** Many types of data now originate in the cloud — for example, in IoT, social media and IT operations, audit trails and logs. There is a tendency for data that starts in the cloud to stay in the cloud, owing to the unstructured and real-time nature of this data together with the phenomenon

<sup>13</sup> Liam Maxwell, No cloud means no AI, Twitter feed.

In addition to technical agility, cloud can contribute to business agility.

of data gravity, whereby data exerts a gravitational pull that attracts other data. In addition, the hyperscalers facilitate use of data brokers to integrate directly with third-party cloud services, creating a single platform to exploit data derived from multiple cloud sources.

**An increasing range of open data sets is available in the cloud.** In the struggle against COVID-19, a key weapon has been the publication of open data sets for open collaboration on COVID-19.<sup>14</sup> As the open data movement grows, through initiatives such as the Open Data Institute,<sup>15</sup> the ability to tap into these open data sets will be of increasing importance, especially as the hyperscalers (Google has gone furthest here) are placing open data sets in their clouds. Moreover, the hyperscalers are offering more industry-specific data sets and building out industry solutions that reduce the time and cost of migration.

### Agility

**Cloud enables scalability.** The nature of many analytics jobs (where capacity is often required over short periods of time, whether to run models, produce reports, train ML/AI models or perform calculations) inherently makes a good fit for cloud. Moreover, the ability of cloud services to autoscale seamlessly aligns well with the less-predictable nature of many analytics workloads. As an example, analytics jobs may be initiated by corporate customers of a bank in a self-service model, with real-time calculation of risk, in order to price products. Finally, being on-demand makes it possible to scale across the entire stack as the needs of the enterprise grow.

**Cloud is made for the experimentation that drives growth.** Amid the uncertainties of COVID and an unpredictable geopolitical environment world, the ability to experiment with new products and services becomes a vital survival skill. The faster development, training, testing and deployment cycles of cloud enable you to experiment, test, learn and tune in much shorter timeframes, essentially collapsing the evolutionary timescale. In addition, cloud makes it easy to scale in line with business growth and to set up data and analytics sandboxes to experiment with data sets and new tools and methods, many of which now come with prebuilt integration provided by the hyperscalers. Especially when enterprises or new capabilities are in their growth phase, cloud provides a logical option to reduce capital outlay and experiment with requirements. A later move on premises may make sense to reduce cost.<sup>16</sup>

### Business agility

In addition to technical agility, cloud can contribute to business agility by making data and analytics more widely available and reducing the time for decision makers to take action.

**Data flow and data fluidity are critical to reducing time to action.** In the same way that enterprises are attempting to reduce time-to-market for new products and services, there is an equal desire to collapse cycle times in data and analytics. But this need goes a step further: It is also about increasing the flow of data, work and decisions across an enterprise and its ecosystem. This is achieved by data that is

<sup>14</sup> Martin Woodward, [Open collaboration on COVID-19](#), The GitHub Blog, March 23, 2020.

<sup>15</sup> [Open Data Institute](#), 2021.

<sup>16</sup> Sarah Wang and Martin Cosado, [The cost of cloud, a trillion dollar paradox](#), Future from a16z, May 27, 2021

The ability to ingest new data, produce innovative models and provide new operational information is a prerequisite for developing resilience and antifragility.

closer to real time, by a faster flow of data through data pipelines, more automated enrichment of data to create information, and enabling analytics customers to self serve. Ultimately this amounts to compressing the cycle time between events, data, information, knowledge and action. Moreover, the timeframe of analytics is shifting from history (reporting), to the present (real-time monitoring) and increasingly the future (modeling).

The effect of this increased data flow and fluidity is felt across an enterprise and its ecosystem, in faster customer interactions, deeper automation and speeding up the evolution of analytics. Cloud solutions can increase data flow and fluidity, by virtue of being API-driven, on-demand and inherently designed for DataOps – the adoption of DevOps practices for data.

**Composable data makes it easier to construct new capabilities quickly.** A key to increasing data flow is composable data, i.e. building blocks of data and analytics that are openly available for self service, allowing capabilities to be constructed quickly and easily.<sup>17</sup> Developers will want composable data to incorporate disparate data sources; data scientists to access ready-to-use data so that they can spend their time on analytics as opposed to manipulating data; decision-makers to self-serve and create their own analyses; and teams from across the enterprise to collaborate by using common data sources. Cloud supports composable data by allowing data to be accessible via APIs underpinned by microservices, containerization and cloud-based data management approaches such as a data fabric. This componentization of data and analytics is a manifestation of the trend towards constructing cloud-native business capabilities,<sup>18</sup> which is the essence of cloud.

**Cloud supports data-democratization, but you need to guard against loss of control.** Cloud has the potential to democratize data and analytics, putting the power to make decisions into the hands of customers and business users. For customers, mobile access to data and point-and-click tools can offer a different level of engagement and empowerment. Meanwhile, business users will be able to self-serve, through internal data marketplaces supported by data catalogs and simple-to-use tools, such as drag-and-drop, no-code AI, visualization and graph tools that represent a more intuitive way to understand relationships. Democratization can turn into anarchy, however, so guardrails are essential, in the shape of governance and controls on data access.

**Cloud provides flexibility for resilience and antifragility.** We have seen how the fight against COVID-19 has depended on the urgent acquisition of new data and modes of analysis. Modeling has ceased to be a task for back-room boffins, and instead has been at the forefront of guiding policy. Indeed, whenever the environment changes, data and analytics requirements are bound to change too. As a result, the ability to ingest new data, produce innovative models and provide new operational information is not a 'nice-to-have'; it is a prerequisite for developing resilience and antifragility.<sup>19</sup> In comparison to on-premises analytics with its dependence on physical hardware and manual processes, cloud makes it that much faster to build new models and reports, add new compute capacity and incorporate new data items.

<sup>17</sup> [Gartner top 10 data and analytics trends for 2021, 2021.](#)

<sup>18</sup> Constructing cloud-native business capabilities, LEF, June 25, 2020.

<sup>19</sup> Krzysytof, Daniel et al, Shock treatment: developing resilience and antifragility, LEF, October 2, 2020.

## Cost

**Cloud can save compute costs for analytics applications that are used periodically.** If ever there was an instance where either-or thinking is deeply unhelpful, it is in the cost comparison of running workloads in the cloud vs. on premises. For example, even within a single type of computing such as high-performance compute, the equation will differ widely. Reports, calculations, models and occasional jobs can often be run more cost-effectively in the cloud, whereas jobs that run almost continuously (for example, long-running calculations and trading) could be very much more expensive in the cloud, hence the need for a Cloud Right approach. In addition, the cost of running workloads in the cloud will vary substantially according to how they are engineered — for example, as IaaS Infrastructure as a Service (IaaS), containers or serverless. On the other hand, solutions that process huge amounts of data 24x7x365 will likely be a very poor fit. Examples include autonomous driving and some pharmaceutical jobs.<sup>20</sup> The key to deciding whether cloud will save money or not is to have an accurate assessment of demand and a bill of materials — a list of all the cloud components for each specific workload.

**Saving on storage will depend on your cost alternative.** The always-on nature of storage does not at first sight make this a rich hunting ground for cost savings. In addition, moving to the cloud often results in more copies of data (for example, where sensitive customer data is retained on premises and anonymized data is stored in the cloud for analytics). The financial equation will depend very much on the on-premises alternative, with large enterprises often able to achieve substantial scale on premises and therefore match or better the price point of the hyperscalers. Furthermore, cloud storage costs typically vary according to speed of access, so (for example) the fast response times offered by some cloud storage services are an unnecessary luxury for a requirement such as archiving.

**Automation will reduce people costs.** Cloud offers the potential to save money through automation of data management and integration tasks. In practice, however, enterprises end up doing more in the cloud, so automation rarely translates to the bottom line. Nevertheless, given that data scientists are likely to be one of your scarcest resources, freeing their time to do analytics instead of data manipulation is invaluable.

**Often the largest area for savings is in rationalization of tools, data and applications.** As a result of mergers and acquisitions and the decentralization of analytics, most enterprises have a proliferation of systems that perform a similar purpose. The move to cloud can act as a trigger to rationalize the application landscape. There will be comparable duplication in data repositories. Whereas a lift-and-shift approach will replicate cost and complexity, rearchitecting will uncover numerous opportunities to rationalize data stores and analytics applications. Moreover, it will be an opportunity to standardize data management and integration tools, in particular by using the tools that are native to the hyperscalers.

**Additional costs in data transfer, networking and transformation should not be underestimated.** There will most likely be increased data transfer and network costs offsetting any potential savings in compute and storage. Often, client-side network

<sup>20</sup> Sooraj Shah, [How long until cloud becomes the preferred environment to run HPC workloads?](#) Computer Weekly.com, January 27, 2021.

## Cloud can be used to enforce data discipline.

costs of data loading and movement are underestimated. In addition, there will be fees for transferring data out of the cloud and so-called shadow costs for interacting within the cloud. The cost of migration and transformation is also frequently underestimated owing to longer-than-expected timescales and greater complexity.

**Cloud is the natural option for capabilities in their early stages, but it may be less appropriate as capabilities mature.** A recent report from a16z argues, “It’s becoming evident that while cloud clearly delivers on its promise early on in a company’s journey, the pressure it puts on margins can start to outweigh the benefits, as a company scales and growth slows.”<sup>21</sup> The issue here is that once a capability matures and usage can be projected accurately, the value of on-demand, pay-per-use services decreases and may no longer offset the cost premium that can come with on-demand services.

### Data management and control

**It shouldn’t have to be this way, but cloud enforces data discipline.** So long as data is kept on premises, there will be a temptation for database administrators to “pop downstairs” and make a few tweaks. Likewise, data scientists will keep private systems under their desks to maintain bespoke data sets and models, and business units will insist on their own personal plate of “meatballs.” Conversely, cloud can be used to enforce data discipline. By bringing everything out into the open, an enterprise can automate, standardize and bring transparency to data management.

**Best-in-class data management tools are now built for cloud first.** A commonly used benchmark is that 80 percent of analysts’ time is spent discovering and preparing data. This lost time can be reduced through cloud-based data management capabilities that typically make it easier to manipulate and manage data in the cloud than on premises.

**You can use the cloud to circumvent Conway’s Law.** Conway’s Law states that any organization that designs a system ... will produce a design whose structure is a copy of the organization’s communication structure, which is often simplified to say that the structure of IT systems reflects organizational structure. So, if you have four products, each with their own development team, you will end up with four customer records and four databases. Moving data to the cloud can circumvent Conway’s Law, by providing effective ways to aggregate and integrate data from across the enterprise — for example, data integration hubs or data meshes that aggregate disparate data from on-premises and cloud sources. Nevertheless, although cloud makes it simpler to integrate and consolidate data, this does not happen automatically by dint of moving to the cloud. Sharing and governance mechanisms must be put in place and new behaviors adopted whereby project teams look beyond their own introspective requirements to the enterprise as a whole.

**Spaghetti and meatballs: more integrated data, less data proliferation.** In his blog Spaghetti and meatballs,<sup>22</sup> Chris Swan explains how limits around the size of database machines and the desire of business units to control “their” data have resulted in enterprises having multiple copies of data. This creates a complex web of links to extract and transform data. Having “lots of meatballs tied together with

<sup>21</sup> Sarah Wang and Martin Cosado, [The cost of cloud, a trillion dollar paradox](#), Future from a16z, May 27, 2021

<sup>22</sup> Chris Swan, [Spaghetti and meatballs](#), Chris Swan’s Weblog, July 7, 2019

The most fundamental question is how you intend to compete and create value through data and analytics.

spaghetti” makes master data management a challenging — and unending — task. As Chris Swan puts it: “The cloud has infinitely large meatballs, so no need for spaghetti ... and master data management boils down to the good management of one giant database.” Here cloud is not a silver bullet: without data management discipline, integration issues and silos will persist in the cloud, but the potential for data consolidation is real.

**One size does not fit all: Multiple-choice data storage.** Until recently, enterprises have relied on very few tools to manage data — typically relational database management systems (RDBMS) that store data in tables. Nevertheless, a one-size-fits-all approach, constrains how an organization stores, retrieves and analyzes data. In contrast, the hyperscalers, as well as a host of specialist vendors, offer a very wide range of data storage options in the cloud. To quote Chris Swan again, “The RDBMS has its place, and that place isn’t everywhere for everything. In fact, I’d go so far as to say that the rational architect wouldn’t choose RDBMS from the plethora of available choices for all but a few projects starting from scratch today.”<sup>23</sup> As a result, many enterprises will adopt a multiplatform storage portfolio across the cloud and on premises.

### Big questions

Before you work through the details of how and where you will adopt cloud for data and analytics, you need to answer some big questions that will shape and guide your strategy, plans and projects teams (**Figure 4**).

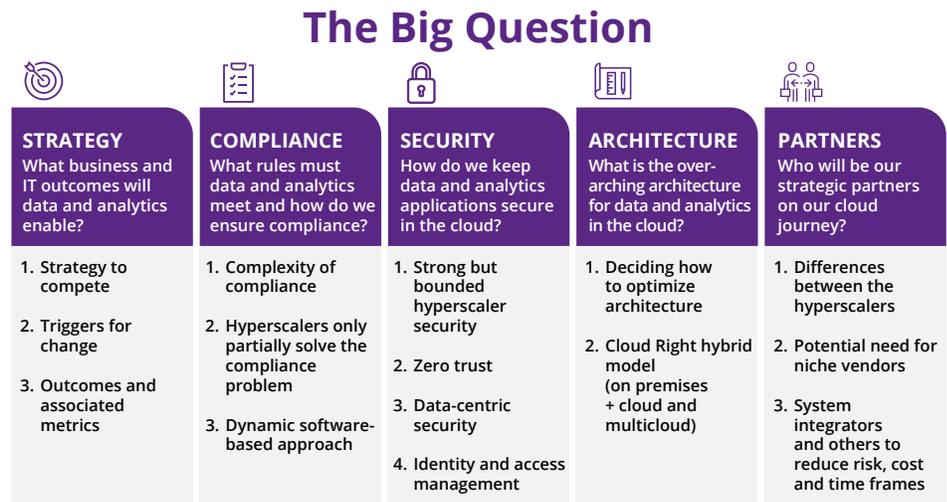


Figure 4 – The big questions

### Strategy: What business and IT outcomes do we need to deliver through analytics and data?

**Determine your strategy to compete and create value through data and analytics.** The most fundamental question is how you intend to compete and create value through data and analytics. Will your priority be analysis, to identify trends in customer preferences or to optimize operations? Or will the focus be

<sup>23</sup> Chris Swan, *Not only SQL*, Chris Swan’s Weblog, May 7, 2010.

It is essential for cloud journeys to have clearly defined outcomes to guide and measure progress.

on operational analytics, say, for autonomous driving, credit risk or preventative maintenance? Equally, you should ask which data sets will be most valuable and whom the principal customers will be for data and analytics: managers, specialist data scientists, customers or AI/ML-based systems. A boil-the-ocean strategy is not required, but there needs to be clarity around the foundational use cases — a high-level set of requirements. Without this business context, it is impossible to make decisions about architecture, hyperscalers and other key partners.

**Identify triggers for change that will drive immediate action.** While your strategy will frame the journey to cloud, any migration will need to demonstrate value by addressing problems in the here and now. This is why it is important to identify triggers for change, say, forthcoming regulation, or IT performance problems, such as scaling and reliability, that must be fixed. You can build out the new architecture by piggybacking on projects to address these triggers.

**Define and agree clear business and IT outcomes upfront.** In *A Cloud Journey to Deliver Business Outcomes*,<sup>24</sup>, we explained why it is essential for cloud journeys to have clearly defined outcomes to guide and measure progress. When it comes to data and analytics, you should define outcomes on several dimensions to ensure that you have a balanced program. Examples are given in **Figure 5**.

Dimension	Outcomes and measures
Innovation	Acquisition of new business capabilities, e.g., autonomous driving for an automotive company or real-time tailored risk pricing in a bank; monetization of APIs and data services; outcome metrics for important operational analytics, e.g., cost reduction and revenue growth
Agility	Time to build new reports; time to repair problems; speed of data flow; access to data on a self-service basis; time to make key decisions (e.g., to gain market feedback to make product changes); feedback from users on data accessibility and ease of use
Cost	All-in run costs for the new service: people, software, data, cloud hosting and storage services, egress and networking costs
Data and information management	Standard metrics for data quality encompassing accuracy, consistency, completeness, integrity, timeliness (ACCIT); measures of time from acquisition of data to action
Technology and Architecture	Addressing triggers for change, e.g., scaling; establishing key features of your new architecture, e.g., a cloud data warehouse or an integration hub
Data	Measurement of actual use of particular data and data sets as a proxy of business value

<sup>24</sup> David Rimmer, [A cloud journey to deliver business outcomes](#), LEF, November 2020.

**Figure 5** – Example of business outcomes

Just because the hyperscalers provide strong security, this does not mean security is solved.

### Compliance: What rules must data and analytics meet and how do we ensure compliance?

#### **Compliance will be the most complex issue to resolve for many enterprises.**

Cloud introduces new paradigms for data storage and recovery that will not map easily onto existing on-premises approaches to compliance. Moreover, many industries (for example banking, healthcare and defense) have specific regulations governing cloud. An extra layer of complexity comes from the fact that compliance is often not black and white: regulations are often risk-based and there are various ways in which data can be de-identified and access to data controlled. In addition, there may be conflicts between the regulations of different jurisdictions or differing requirements that necessitate one solution for one country and a different one for another. Finally, whereas the status quo around data storage and processing will have evolved over many years, the move to cloud will shine a spotlight on compliance. Here, the software-defined approach of cloud will act as forcing function that drives out ambiguity.

**The regulatory-compliant solutions of the hyperscalers only partially solve the problem.** Compliance is easier than it was. The hyperscalers are adding data residency functionality into storage solutions, expanding their data center network and have invested in solutions that comply with privacy and industry regulations, such as PCI and HIPAA. All the same, whenever an enterprise builds its own services on top of a hyperscaler's certified solution, there is a need to revalidate regulatory compliance. Moreover, regulators will want to know exactly where in the cloud data is processed, resides and is backed up.

**A dynamic software-based approach to compliance is required.** Whereas regulation was once stable and predictable, geopolitical pressures and mounting concerns over privacy have made regulation more fluid. Compliance should, therefore, be policy-driven through software that provides a line of sight from regulation to code and enables rapid reconfiguration in response to rule changes, as opposed to brittle, hard-wired bricks-and-mortar controls. Folk interpretations of regulation will often have grown up as received wisdom, so it is particularly important to tie back to the actual wording of the regulation.

### Security: How do we keep data and analytics applications secure in the cloud?

#### **The hyperscalers provide exceptional security, but cloud increases the attack surface.**

The investment of the hyperscalers in security dwarfs that of anyone, bar a few national defense organizations — and even they use the hyperscalers. Consequently, most IT security specialists accept that the hyperscalers provide stronger security than is possible on premises.<sup>25</sup>

Nevertheless, just because the hyperscalers provide strong security, this does not mean security is solved. The hyperscalers operate a shared responsibility model, whereby they guarantee the security of their cloud, while the customer is responsible for security of what is in the cloud. In addressing their half of cloud security, enterprises need to be aware that cloud introduces new vulnerabilities. Owing to the modular, componentized nature of cloud services, the attack surface is increased.

<sup>25</sup> David Linthicum, [IT pros agree: security is better in the cloud](#), InfoWorld, March 31, 2017, and David Mitchell Smith, [Cloud strategy leadership](#), Gartner, 2017.

Furthermore, in adopting cloud, disparate data is often brought together, raising the stakes in case of a data breach. This means that new approaches to security are required, in particular, Zero Trust security, data-centric security and cloud-based identity management.

**Zero Trust models are becoming the standard for cloud and hybrid solutions.**

The principle of Zero Trust is now recognized as fundamental: “Zero Trust assumes that everything around a network asset is hostile, including network assets from the trusted zone. All access to the network asset is, by default, not trusted.”<sup>26</sup> The premise of Zero Trust is that to be secure, enterprises must verify and authenticate access in a continuous manner based on numerous data points, including identity, location, device, service and data classification.

**Security should be data centric.** Alongside the move to Zero Trust, an equally essential transformation in security models is embracing a data-centric approach. The rationale is that once data is stored in the cloud and moved backwards and forwards between clouds and on premises, security has to be applied to the data itself. Controls will be required around data access and authorization: access controls dictate whether people or systems have any access at all to a type of data, whereas authorization controls such as attribute-based access controls (ABAC) determine which aspects of the data they can see and in what format.

**Identity and access management may well need to be overhauled.** “On-premises [identity and access management] systems are becoming less relevant with data and IT services spread across data centers, cloud infrastructures, mobile devices and the expanding internet of things (IoT).”<sup>27</sup> As a result, adopting cloud for data and analytics often requires an overhaul of identity and access management. Although the hyperscalers all provide advanced cloud-native identity management capabilities, cloud-based identity-as-a-service (IDaaS) options are also beginning to find favor, especially where enterprises operate in a hybrid model, with a range of cloud service providers and on-premises solutions.

<sup>26</sup> TM Ching, [Zero Trust for maximum security](#), DXC, May 2020.

<sup>27</sup> Martin Reilly et al, [Identity and access in the cloud: The future of the secure enterprise](#), DXC, June 2021.

<sup>28</sup> Matt Bornstein, Martin Casado & Jennifer Li, [Emerging architectures for modern data infrastructure](#), Future from a16z, October 15, 2020.

<sup>29</sup> Ted Friedman, [How data warehouses, data lakes and data hubs differ in focus and work better together](#), IProPortal, August 7, 2020.

**Architecture: What is the overarching architecture for data and analytics with a Cloud Right approach?**

**You can't optimize for everything.** Since it is not possible to optimize for every eventuality, choices are required around architecture based on the primary business use cases that you have defined. This theme is explored in detail in *Emerging Architectures for Modern Data Infrastructure*,<sup>28</sup> where three blueprint architectures are outlined to fit different primary uses cases: modern business intelligence, which focuses on cloud-native data warehouses and analytics use cases; multimodal data processing, covering both analytic and operational use cases built around the data lake; and operational systems and the emerging components of the AI and ML stack.

Within these overarching architectures, a number of concepts are emerging. A new orthodoxy has yet to emerge. Indeed, there may end up being no standard orthodoxy: A Cloud Right approach deploying “horses for courses” may be the answer. Prominent among these emerging architectural concepts are the lakehouse, the data integration hub/data mesh and the cloud data warehouse (**Figure 6**).<sup>29</sup>

Most large traditional enterprises will operate Cloud Right solutions on premises and in the cloud.

Architecture model	Definition	Find out more
The lakehouse	A cross between a data warehouse and a data lake	Welcome to the Lakehouse
Data integration hub or data mesh	Integrates and centralizes enterprise data, connecting business applications and other data sources with analytics repositories and tools	What is a Data Hub: Shucks — Piccadilly Circus!
Cloud data warehouse	Cloud data warehouse typically combines storage and processing, but allows storage and processing to be scaled independently.	What is a cloud data warehouse, and why is it important?

Figure 6 – Three emerging architectural concepts

**A hybrid model is inevitable for most legacy enterprises.** Simply because cloud is the answer in many areas, it is not necessarily the right answer everywhere. No doubt startups will look straight to the cloud. Legacy enterprises, however, will usually have some analytical applications and data stores for which there is no business case for cloud, necessitating a Cloud Right approach for the right platform at the right time. The obstacles will include technical performance, security and compliance in highly regulated industries, and costs of data storage and processing in the cloud, or the investment required for migration and transformation. Consequently, most large traditional enterprises will operate Cloud Right solutions on premises and in the cloud.

Most architectures will also be a hybrid of cloud service providers. In fact, according to a survey, 34 percent of enterprises will operate multiple clouds for data and analytics.<sup>30</sup> As a result, a hybrid control plane that provides visibility into all workloads, whether on premises or in the cloud, becomes an all-but-essential capability.

Finally, the time taken to migrate and transform analytical systems and data repositories is such that you will in practice be operating in a hybrid world for several years.

### Partners: Who will be our strategic partners on our cloud journey?

#### The hyperscalers are not all the same when it comes to data and analytics.

The services offered by the three principal hyperscalers become more differentiated as they move up the stack from IaaS to PaaS and to cloud-native capabilities.<sup>31</sup> For instance, in areas such as data integration and ML/AI, the hyperscalers bring different capabilities to the table. Consequently, we often see enterprises choosing a hyperscaler specifically for data and analytics, to go alongside the hyperscaler(s) that they have selected at the enterprise level.

<sup>30</sup> Building a high-performance data and AI organization, MIT Technology Review, April 15, 2021.

<sup>31</sup> David Rimmer et al, Constructing Cloud-Native Business Capabilities, LEF, June 25, 2020.

Even in something as basic as storage, the range and cost of the options can vary widely between the hyperscalers.

As always, it is a matter of being clear about requirements. For example, Google's Big Query is a serverless data warehouse that enables scalable analysis of petabytes of data, but do you actually require lightning-fast processing? Even in something as basic as storage, the range and cost of the options can vary widely between the hyperscalers, reflecting different response times and distinct visions as to why you are storing data in the cloud. Another example difference between the hyperscalers is that they come with a variety of open data sets and have struck distinct partnerships with data vendors.

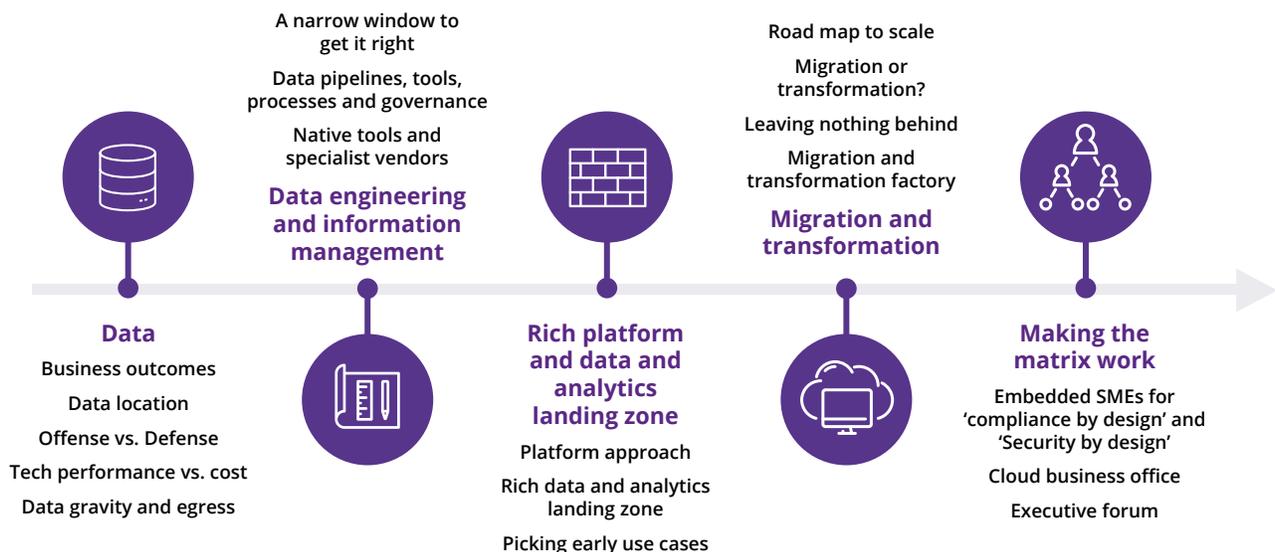
**Large and complex enterprises are likely to require specialized software partners to complement the hyperscalers.** Although the hyperscalers offer cloud-native services across the piece and are making acquisitions fast, there are likely to be aspects of your architecture that are best filled by a specialist — say for next-generation data lakes, cloud data warehouses, data catalogs, data access control and self-service reporting tools and data governance.

**Implementation partners can reduce risk, cost and timescales.** The journey to cloud for data and analytics is complex and involves many new technologies and patterns. As a result, while some enterprises will decide to go it alone, most will seek assistance from systems integrators and other IT service providers. By drawing on lessons learned, skilled practitioners and proprietary tools in areas such as architecture and migration planning, cost, risk and time to market can all be lowered.

<sup>31</sup> David Rimmer et al, Constructing Cloud-Native Business Capabilities, LEF, June 25, 2020.

### Moving from strategy to action

Having answered the big questions, you can move from strategy to action. Here are five areas that are critical for execution (**Figure 7**).



**Figure 7** – Critical areas for execution

Just because analytics are performed in the cloud, it does not follow that data has to be stored in the cloud.

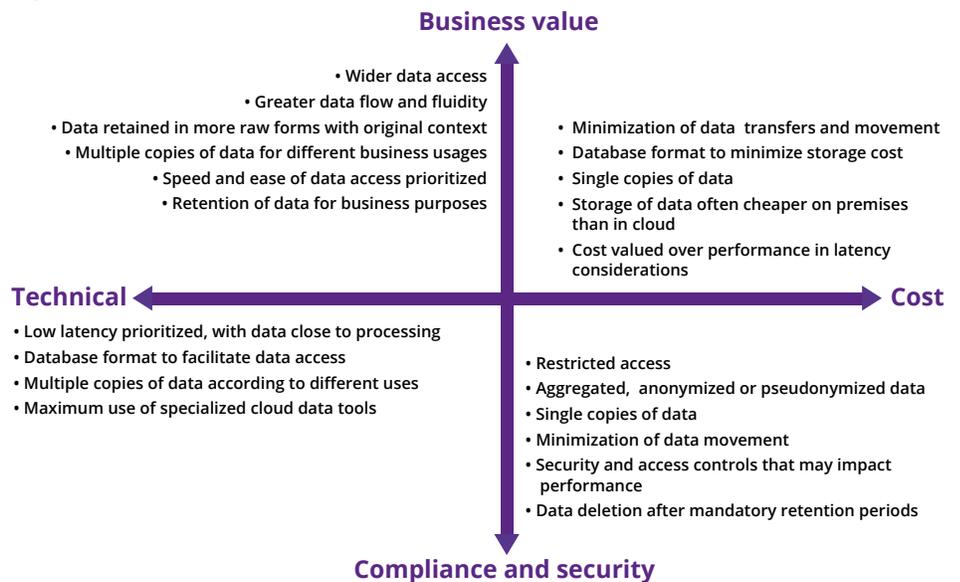
## Data

**Your data strategy will follow directly from the business outcomes that you have defined.** Becoming a data-centric enterprise rests on recognizing that data is an asset, whose primary value comes from being an input to processes, instead of an output. Moreover, while processes, applications and infrastructure will change over time, an enterprise's data, in aggregated or atomic form, will endure. So, having defined business outcomes, the next step is identifying the data required to deliver these outcomes, whether sourced internally or externally. These requirements form the basis for a data-centric architecture that defines where, when, how and by whom data is required. This in turn shapes whether data should be stored or processed in the cloud or elsewhere.

**Data location is a complex issue in which tradeoffs are often required.** Just because analytics are performed in the cloud, it does not follow that data has to be stored in the cloud. Data can be stored on premises and moved to the cloud when needed. Likewise, although enterprises will be looking to cloud to offer a more integrated view of data, it does not follow that data has to be centralized physically in the cloud. Rather, data centricity is achieved by having a common metadata layer that allows data to be utilized as a shared asset, wherever it is stored.

In resolving how and where data should be stored and processed, a whole range of factors has to be weighed: business needs, security, compliance, cost and technical performance. Furthermore, these dimensions often pull in different directions, so tradeoffs may be needed. It is important to recognize that, as much as anything, this is an organizational issue, where you need to make the matrix work to weigh apples against pears to take decisions.

The multi-dimensional, push-me-pull-you character of the problem is illustrated in **Figure 8**.



**Figure 8** – The multiple dimensions of data location

**Offense vs. defense: Business value vs. compliance and security.** In *What's your data strategy?*<sup>32</sup>, the authors argue that a data strategy needs to strike the right balance between offense and defense. Offense focuses on growth-related business objectives such as increasing revenue, profitability and customer satisfaction. Defense is about minimizing downside risk and ensuring compliance with regulations. The choice matters because offense optimizes for flexibility: more open access to data, more atomic data, data that is stored along with its context, longer retention of data, and multiple versions of the truth; while defense brings a compliance and security perspective and optimizes for control, arguing for the reverse positions, for example a single source of the truth.

While enterprises will want to define their overall posture (their balance between defense and offense) and consequent data management policies, in reality there are countless decisions that will need to be made use case by use case.

**Technical performance vs. cost.** A balance frequently has to be found between cost and technical factors. For example, performance may require data to be stored close to processing or call for storage to be optimized for accessibility. Alternatively, minimization of cost may pull in the opposite direction, dictating different location and storage options. Other technical considerations include the “chattiness” of applications, the frequency of data interchange, and integration of data from multiple locations. Many of these technical constraints can be overcome at a price, for instance by caching data or retaining multiple copies. It is important that performance issues are considered early in the design and architecture process, and if needed assessed through POCs; otherwise, there is a risk of wasting money on solutions that look good on paper but prove not to be viable in practice.

**Data gravity and egress fees dictate that your decisions will have long-term implications.** When systems and data become more distant, latency grows, IT performance decreases, and the cost of moving data increases. Consequently, a gravitational pull (data gravity) is exerted on data and other IT systems to be located near to existing data and systems. This is an observable effect that can be calculated through the formula:

$$(Data\ mass \times Data\ activity \times Bandwidth) / Latency^2 = Data\ gravity^{33}$$

Owing to data gravity, decisions about where to store and process data have long-term implications — another reason why cloud architecture needs to start with the data.

Gravitational pull is heightened by the egress fees that the cloud providers charge, making moving data to the cloud not just a one-way street, but one in which it is hard to reverse.

### **Data engineering and information management**

**You have a narrow window to get data management right.** Without an absolute focus on data engineering and information management, much of the promise of moving to the cloud will be lost. In moving to the cloud, you have a narrow window to get information management right before a new sprawl is created. It is important to realize that data engineering and information management are not just there

<sup>32</sup> Leandro DalleMule and Thomas Davenport, [What's your data strategy?](#) Harvard Business Review, May-June 2017.

<sup>33</sup> [Data gravity index DGx](#), Digital Realty, 2020.

Data flows may operate in completely new ways once you move to cloud.

for compliance: they create the trust that is essential for customers and business users to rely on data to make decisions. A group of IT leaders found: "Although risk management and compliance departments are often its first proponents, all departments benefit from a robust [data governance] framework. Whether they see it as such is where the challenge lies."<sup>34</sup>

**Map data pipelines to tools, processes and governance.** In order to design data pipelines, you will need to map out the data flows that will be affected by the move to cloud, with separate patterns defined for each. An enterprise will typically have five to ten distinctive patterns (examples include operational decision support; self-service analytics; data science; asynchronous decision support; and data streaming). Data flows may operate in completely new ways once you move to cloud, for instance, with a reversal of extract, load and transform (ELT) to extract, transform and load (ETL) in order to speed processing. Likewise, there has to be utter clarity on the tools, processes and governance that will be used to manage information across the end-to-end life cycle. Key areas will include data augmentation, data masking, addition of metadata, data lineage and data catalogs.

**Leverage native tools wherever possible, but that may not be enough.** The hyperscalers offer native services and you will obviously want to take these as your starting point and map data flows to these services. Even so, this is a highly specialized domain where you may well require additional vendor tools, especially where scale, industry regulation and multiple geographies drive extra complexity. In order to prevent data in the cloud becoming a new silo, data engineering and information management should be handled in a holistic manner across all platforms, both on premises and in the cloud.

### A rich platform and data and analytics landing zone

**A platform approach: Going slower to go faster.** A vital prerequisite for a cloud data and analytics program is a richly functional enterprise cloud platform. By this we mean the set of common components that will be required to manage applications and workloads in a particular hyperscaler's cloud. For example, the platform should include single sign-on, networking, security models such as Zero Trust, monitoring, DevOps and regulatory compliance. These are all functions that a data and analytics landing zone will require, but since they will be needed across the enterprise, they should be defined in an enterprise cloud platform.

For a large enterprise, such a platform will typically take 3 to 6 months to build and goes further than what many would consider to be an initial landing zone. If these capabilities are included in the platform then each project is spared the effort of making its own decisions, a more standardized solution is delivered, and above all, the task of evolving compliance and security response is simplified and more robust because changes are made in the common platform. Ultimately, this is about going slower to go faster.

<sup>34</sup> Winston Thomas, [Are we doing data governance backwards?](#), CDO Trends, February 9, 2021.

**A rich data and analytics landing zone is required to ensure compliance by design and security by design.** In the haste to make rapid progress there is a temptation to build a basic data and analytics landing zone that addresses security, networking, dropping a data file, and maybe a sandbox, but which leaves everything else to later projects to resolve. This approach is an expensive mistake. You will miss the opportunity to standardize data engineering and information management and to bake in compliance by design and security by design. You will spend more, as projects will reinvent the wheel.

Instead, just as with the cloud platform, a data and analytics landing zone with a richer set of functionality is required, in particular to address data engineering and information management. This data and analytics landing zone is an adjunct to the cloud platform, not an independent entity. The data and analytics landing zone should cover the end-to-end data-engineering pipeline, from the sourcing of data to data ingestion and data exchange.

Moreover, the landing zone should include all the tools and processes necessary for information management and governance, such as data cataloging, data protection and data lineage. Downstream processes, such as archiving, deletion and e-discovery, should not be forgotten. Since many enterprises choose a hybrid approach, it is essential to consider how data flows across the cloud and on premises.

**Choose the use case for the first landing zone with care.** The landing zone should address a high-value use case with a set of specific requirements, rather than an abstract generic requirement, otherwise there is no guarantee that the landing zone is truly functional, nor that there will be a positive business case — not a good way to start! A use case that is overly complex will put back the overall timelines, yet one that is too simple may not actually prove very much — and the next step will be another POC or landing zone.

## Migration and transformation

**You need a roadmap to achieve migration and transformation at scale.** While most enterprises are able to build POCs and MVPs and move limited numbers of applications and data stores, many struggle to move and transform enough of their estate — and some end up in a worse position than before, having introduced added complexity without commensurate benefit. One of the issues here is the lack of a clear strategy to scale beyond the initial implementations.

Key considerations in developing a roadmap are shown in **Figure 9**.

Planning consideration	Description
Key milestones	Delivering the major outcomes defined for the journey, especially tangible milestones such as fixing problems in the current environment and establishing transformational capabilities or features of the target architecture
Industrialization	Identifying blocks of repeatable applications and reports where scale can be achieved through a factory approach
Avoiding migration greed	Migration greed is where unrealistic aspirations for migration end up disappointing stakeholders and wasting time in constant replanning cycles. A Cloud Right approach recognizes early on that some workloads will not be a good fit for cloud, instead of wasting effort in attempts to squeeze square pegs into round holes.
Business systems	Focusing on business systems (e.g., repositories of inventory data) to transform an area of the business
Gravitational pull	Targeting significant data sets and applications for migration and transformation in order to exert a gravitational pull towards the target architecture
Sequencing	Deciding the sequence for building out the patterns of data pipelines and the applications & data stores that will subsequently follow these patterns
Iterative process	No plan survives contact with the enemy, so development of the roadmap should be seen as an iterative process, with hands not thrown up in the air when the plan needs to change

**Figure 9** – Key planning considerations

**Migration or transformation?** As with moving operational systems, decisions need to be made around what moving entails. Is it a migration, involving rehosting or replatforming? Or a transformation, entailing refactoring and rearchitecting so that applications and infrastructure are modernized and re-engineered to operate on a fully cloud-native basis? The key here is to be clear about what business outcomes you are trying to achieve and what degree of change will unlock those outcomes.

Even when the goal is transformation, you will have to choose whether to transform and then migrate or to migrate and refactor/re-engineer later. The cloud offers

many tools to facilitate transformation, so frequently either strategy both delivers benefits in the short term and reduces transformation costs in the long term, in particular for complex analytics applications. Sometimes, an approach where you migrate and then see what happens can work, especially for small files and reports. On the other hand, for many reports and databases, the lack of an easy read-across leads to the path of transformation and then migration. Moreover, a degree of data cleansing, if not restructuring, pays dividends prior to a large-scale move, such as to a cloud data warehouse.

**Don't leave any bits and pieces behind.** A common mistake in migration and transformation is failing to pick up the whole of a process or application. In any migration or transformation, it is essential to identify and migrate all the elements that comprise an application or process, including all upstream and downstream data sources and dependent applications. If this is not done the move will prove not to be minimally viable, resulting in both a poor customer experience and the ongoing cost of operating multiple systems.

**A migration and transformation factory can reduce costs and bring standardization.** There will be parts of your estate that have a substantial degree of commonality (similar applications, migration patterns, databases and reports). In these instances, the most cost-effective approach is often to put them through a data and analytics migration and transformation factory. A further benefit of this approach is the standardization that comes from having a common team that can work with business and application teams. The increasing range of tools to automate migration reinforces this industrialized approach.

### Making the matrix work

**Embed SMEs to ensure compliance by design and security by design.** A challenge in solving the multi-dimensional data problem is that each dimension is represented by different stakeholders. As a result, you need to find a way to make the matrix work at various levels of the organization. At the project level, compliance and security experts should be involved, not at the end of a build as the person who likes to say, "No," but throughout as members of the team and as the person who says, "Yes, if." The terms DevSecOps and DevSecRegOps are often used to convey this approach to compliance by design and security by design. Similarly, data architects should be dispersed among project teams to ensure adherence with data architecture and design principles, instead of acting solely as a centralized control gate.

**A cloud business office brings integrated decision-making.** In *A Cloud Journey to Deliver Business Outcomes*<sup>34</sup>, we explained how a cloud business office (CBO) can act as an integrated operating model that brings together the business, IT, finance, procurement, security, compliance, legal and architecture to make rounded decisions. In data and analytics, the CBO can play an important role in weighing perspectives, for example, ensuring that run costs are fully factored in to design decisions.

**An executive forum can act as a point of escalation but the buck needs to stop somewhere.** At the executive level, there needs to be a forum to agree how the big questions are going to be addressed and to provide subsequent oversight. This forum will also serve to resolve issues that cannot be addressed lower down the

<sup>34</sup> David Rimmer, *A Cloud Journey to Deliver Business Outcomes*, LEF, November 2020.

Success will depend on clear top-down thinking to address the big questions and to develop a strategy to scale adoption.

organization. Within this executive team it is wise to have one person with overall accountability for the data and analytics aspect of a journey to cloud. The buck has to stop somewhere.

## Conclusion

Reorienting the enterprise so that it is centered on data is a colossal undertaking — but one that is essential in order to thrive in a data-centric world. It is hard to see how cloud would not form a key strategy within this transformation, given the range of capability that is already available and the depth of investment. All but the most risk-averse enterprises will surely try to avoid being one step behind along a crucial dimension of competition. But equally, the response must not be starry-eyed thinking that fails to recognize that migrating to cloud is difficult and that for most mature enterprises it is unlikely to be the right solution everywhere.

Instead, success will depend on clear top-down thinking to address the big questions and to develop a strategy to scale adoption. At the same time, the top-down approach has to be tempered with bottom-up planning and action, since complex issues (such as compliance, data management and transformation) can only be addressed by working through the detail and learning by doing.

---

### About the authors

**David Rimmer** is an industry advisor for Banking and Capital Markets at DXC Technology. His role is to understand customers' business problems and develop digital solutions that can be delivered rapidly to achieve their goals, often through rapid POCs and pilots. He previously worked at Hewlett Packard Enterprise Services, responsible for the Mid-Tier UK Financial Services sector.

**Dave Wilson** is an account chief technologist and researcher at DXC Technology, focusing on innovation and new business. He previously worked as an enterprise architect at Hewlett Packard Enterprise Services, responsible for architecting digital solution for customers.

**Contributors:** James Coleman, Michael Conlin, Mamoun Hirzalla, Sebastian Kloeser, Andriy Sas, Chris Swan, David Whitehead

Learn more at  
[dxc.com/cloudright](https://dxc.com/cloudright)

Get the insights that matter.

[dxc.com/optin](https://dxc.com/optin)



### About DXC Technology

DXC Technology (NYSE: DXC) helps global companies run their mission critical systems and operations while modernizing IT, optimizing data architectures, and ensuring security and scalability across public, private and hybrid clouds. The world's largest companies and public sector organizations trust DXC to deploy services across the Enterprise Technology Stack to drive new levels of performance, competitiveness, and customer experience. Learn more about how we deliver excellence for our customers and colleagues at **DXC.com**.