Gartner Research

The Impacts of Data Ecosystems: A Cloud Architectural Perspective

Adam Ronthal, Donald Feinberg

13 April 2023

Gartner®

The Impacts of Data Ecosystems: A Cloud Architectural Perspective

Published 13 April 2023 - ID G00788724 - 15 min read

By Analyst(s): Adam Ronthal, Donald Feinberg

Initiatives: Data Management Solutions; Evolve Technology and Process Capabilities to Support D&A

Data and analytics leaders are increasingly seeking a broad and holistic approach to varied data management problems. Data ecosystems provide a preintegrated enterprise solution that allows for streamlined delivery and optimization of a broad range of data use cases.

Additional Perspectives

- Summary Translation: The Impacts of Data Ecosystems: A Cloud Architectural Perspective
  (25 May 2023)

Overview

Impacts

- Early adopters of data ecosystems may find that end-to-end preintegrated data ecosystems in the cloud don't yet suit their current urgent needs, driving them to blended approaches.
- Data gravity will increasingly drive the choice of data ecosystem delivery styles.
- Metadata and governance are the foundation of data ecosystems and drive increased focus on data fabric design.

Recommendations

Data and analytics leaders who are looking to unify and simplify their data management solutions landscape should:

- Define and document what they expect CSPs to deliver as part of their data management landscape, as well as what capabilities they expect to obtain from third-party independent software vendors (ISVs) — especially with respect to foundational and enabling components like metadata management and data integration.
- Evaluate data ecosystems based on their ability to resolve distributed data challenges and to access and integrate with data sources outside of their immediate environment.
- Give precedence to those cloud data ecosystems that support the governance, access, sharing and analysis (using standards) of metadata across various tools to support a data fabric architecture.

Strategic Planning Assumptions

- By 2025, 55% of IT will adopt data ecosystems, consolidating the vendor landscape by 40%, thereby reducing cost while reducing choice.
- By 2026, organizations adopting active metadata practices will increase to 30% across data and analytics to accelerate automation, insight discovery and recommendations.

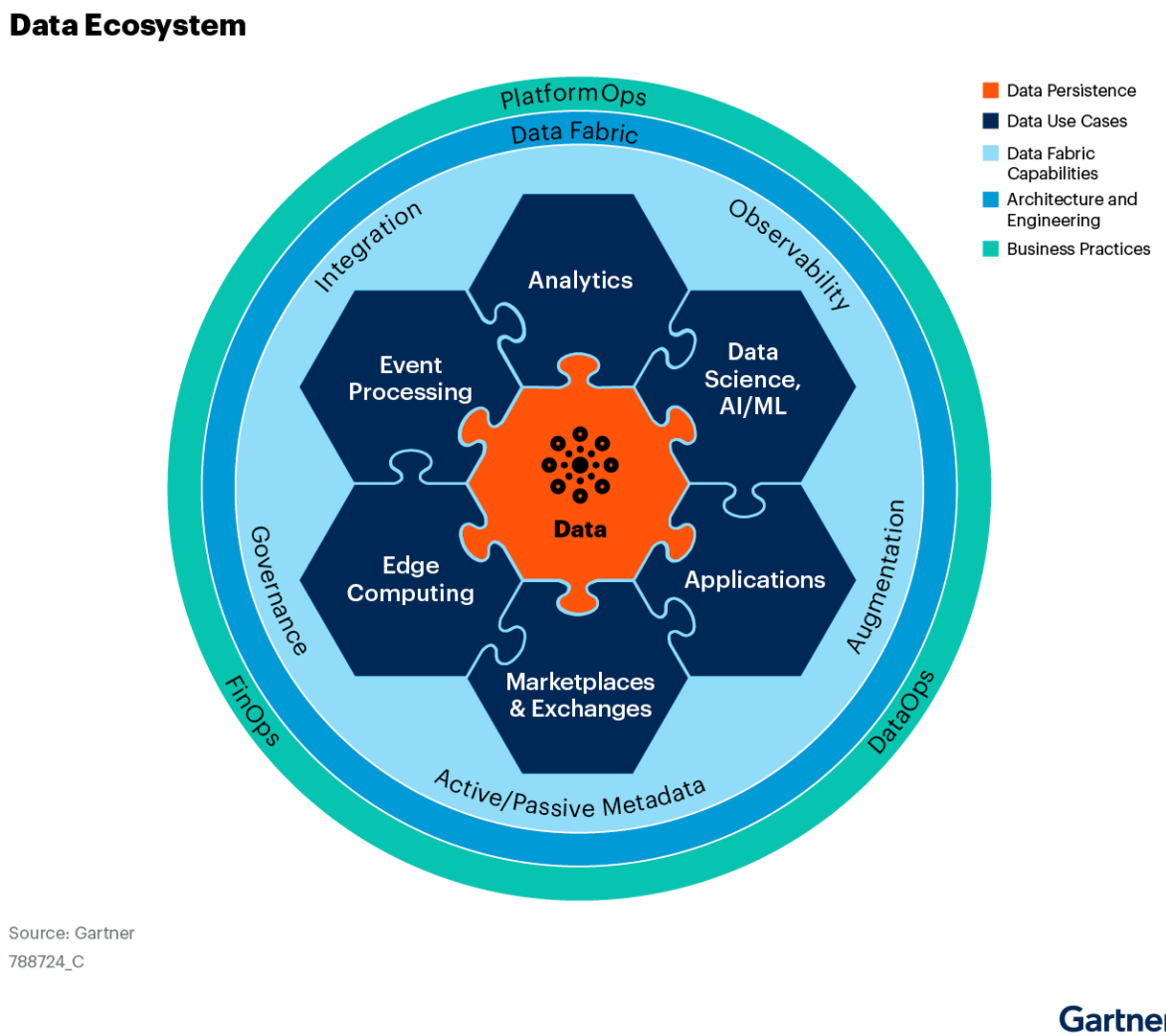
Introduction

The center of data gravity has pivoted to the cloud. In 2021, nearly 50% of the revenue in the nearly $80 billion database management systems (DBMS) market was attributed to database platform as a service (dbPaaS) cloud consumption. Additionally, nearly 90% of the 2020 to 2021 revenue growth of $14.5 billion was cloud DBMS revenue. Cloud DBMS revenue is forecast to be more than 70% of the overall DBMS market by 2026. [1] When viewed as a whole, the overall DBMS market grew at a healthy 22% in 2021. When we remove the cloud component of that growth, however, it drops to a relatively anemic 5%. As long as cloud is considered a submarket of DBMS, the DBMS market shows healthy growth. This hides the true dynamics of market growth; cloud is growing massively while the traditional DBMS market is essentially static.

More data, data use cases (see Figure 1) and supporting infrastructure are moving to the cloud and, in many cases, multicloud (see How to Plan for Optimal Multicloud and Intercloud Data Management). This has led to increased complexity surrounding data management practices, including financial governance, data and application integration, governance, and metadata management.

Along with the cloud transition, the second vector of change impacting the data management market is the diversity of data structures and the underlying database management technology used to store data in an efficient manner (see Figure 1). Data management is no longer wholly focused on relational data. Document, graph, time-series, wide-column, key-value, ledger and other targeted data stores all provide specific optimizations for different types of data and different use cases. Sometimes, these are combined in a single data management platform — a multimodel database. Other times, they remain as best-fit, targeted point solutions (see Choosing Between Multimodel DBMS and Multiple Specialized Engines).

Figure 1: Data Ecosystem



Source: Gartner
788724_C

Gartner

The space between the puzzle pieces in Figure 1 represents the connective tissue between the use cases that weave data together. Each component of the data ecosystem relates to the other components. The architecture and engineering layer (data fabric design) brings the components together into a coherent, manageable whole (see Data and Analytics Essentials: Data Fabric). By looking at sets of components together, you can describe use cases, opportunities for optimization, and scenarios for data reuse. It is from these connections that the emerging data fabric design arises.

The data fabric enables the data ecosystem through data integration, metadata management, observability, augmentation and governance capabilities. It also continuously observes and evaluates metadata to deliver recommendations that allow the data ecosystem to optimize integration workloads, workload management, metadata sharing and understanding, policy enforcement augmentation, financial governance, and other components that are essential to success.

In short, data today is distributed, disparate and diverse. Many data and analytics leaders report that the cloud experience today requires a significant integration effort to ensure that components all work well together. CSPs and ISVs are starting to respond with more refined and mature data ecosystems as the market moves from “some assembly required” to a more integrated platform experience.

Data ecosystems aim to resolve the inherent stress and complexity placed on the overall data and analytics landscape. They require new practices like DataOps, FinOps and PlatformOps. They rely on emerging, enabling architectures like the data fabric. They must leverage machine learning to self-optimize with augmented data management. And they will ultimately provide comprehensive financial governance capabilities, streamlining spend and budgeting across components.

This research will give data and analytics leaders the necessary understanding of the architectural patterns and foundations of the emerging data ecosystems and the implications for data and analytics deployments (see Table 1). This will enable them to make strategic investment choices as they plan for a future data and analytics landscape that is very different from today’s.

Data ecosystems provide a cohesive data management environment that supports the whole range of data workloads, from operational and transactional data to exploratory data science and production data warehousing. Built on the foundation of the data fabric, data ecosystems have a common governance and metadata management framework, provide unified access management, and integrate augmented data management capabilities with a set of services accessible by the business user. Data ecosystems leverage distributed components that may run on multiple clouds and/or on-premises but are treated as a logical whole. This makes it easier to meet new requirements, increasing productivity and thus return on investment.

Table 1: Top Impacts and Recommendations for Data and Analytics Leaders

Impacts ↓	*Recommendations* ↓
■ Early adopters of data ecosystems may find that end-to-end preintegrated data ecosystems in the cloud don't yet suit their current urgent needs, driving them to blended approaches.	■ Define and document what you expect CSPs to deliver as part of your data management landscape, and what capabilities you expect to obtain from third-party ISVs — especially with respect to foundational, enabling components like metadata management and data integration.
■ Data gravity is driving the choice of data ecosystem delivery styles.	■ Evaluate data ecosystems based on their ability to resolve distributed data challenges and to access and integrate with data sources outside of their immediate domain through orchestration and optimization capabilities.
■ Active metadata and governance are the foundation of data ecosystems and drive increased focus on data fabric design.	■ Plan your data ecosystem environment by focusing on the metadata and governance capabilities provided by the underlying data fabric components on which it rests.

Source: Gartner (April 2023)

Impacts and Recommendations

Data Ecosystems Address Multicloud, Intercloud and Hybrid Challenges

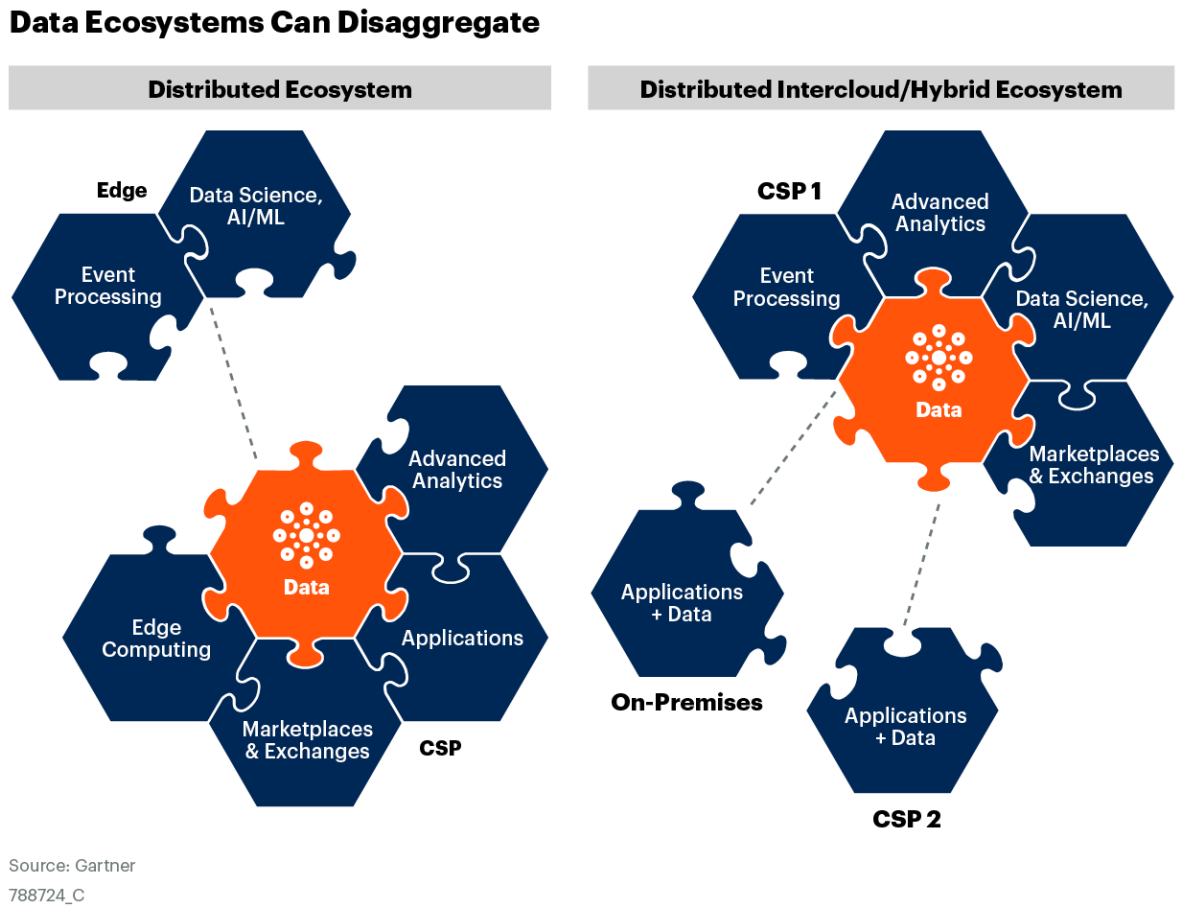
Data ecosystems leverage distributed components that run on multiple clouds and/or on-premises but are treated as a cohesive whole. The 2022 Gartner State of Data and Analytics Cloud Adoption Survey found that 85% of those using the public cloud were using more than one. [2] Further, cloud revenue as a percentage of overall revenue in the DBMS market was nearly 50% of a nearly $80 billion market in 2021. [1] However, that still leaves a sizable on-premises footprint, and the transition to cloud will have a long, drawn-out tail.

Hybrid use and multicloud will be a factor in data and analytics landscapes for the foreseeable future.

As data becomes increasingly distributed, data and analytics leaders will seek a means to unify their data and analytics landscape. The data ecosystem already does this in a single environment, and it will extend its reach to secondary environments as well. Components of the data ecosystem may disaggregate and be deployed outside of the primary center of data gravity, while still participating in the cohesive whole (see Figure 2).

An Internet of Things (IoT) environment may deploy a centrally trained machine learning model at the edge, while still feeding data into a centralized data store residing in the primary environment. A data ecosystem may ingest data from multiple sources in secondary clouds or on-premises environments, or both, and can manage them as a cohesive whole.

Figure 2: Data Ecosystems Will Be Implemented in Multiple Environments



Just as metadata and governance serve as the foundation of a single-environment data ecosystem, they will be instrumental in unifying the distributed data ecosystem as well.

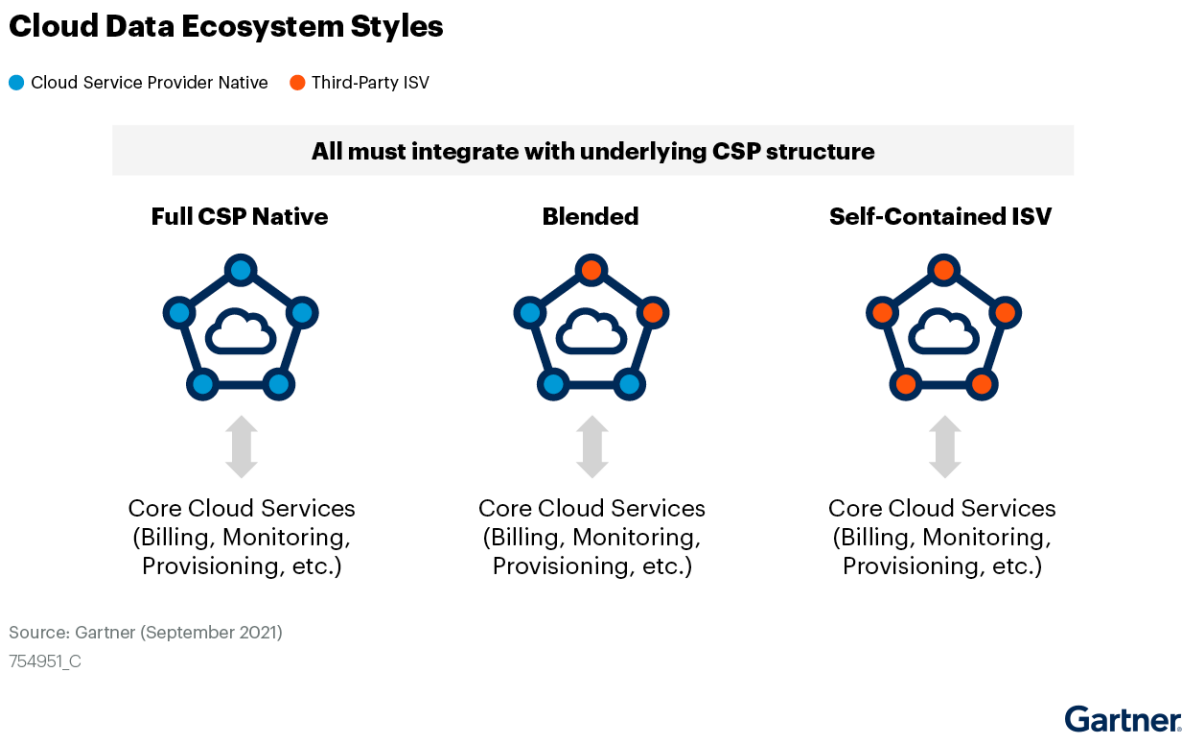
Recommendation for data and analytics leaders:

- Use data ecosystems to resolve distributed data challenges by evaluating their ability to reach beyond their primary deployment environment and to integrate with data sources outside of their immediate domain.

Three Styles of Data Ecosystem Have Emerged

Three styles of data ecosystems have emerged (see Figure 3).

Figure 3: Cloud Data Ecosystem Styles



Source: Gartner (September 2021)
754951_C

Gartner

In a *full CSP native stack*, all native CSP services are used with no third-party ISV components. The expectation is that these components will all work well together, and that minimal integration effort is required — something that has not always held true in the cloud to date.

A *self-contained ISV ecosystem approach* will be followed by vendors with broad and deep portfolios that are already well-integrated. These portable ecosystems will run on the end user's cloud of choice. Examples include Cloudera Data Platform, IBM Cloud Pak for Data and the SAP Business Technology Platform. Even in this scenario, the ISV must be a good citizen of the larger cloud ecosystem on which they operate. Ideally, the ISV would build connections to underlying management infrastructure and services and leverage core offerings like the cloud object store and infrastructure as a service (IaaS) effectively.

A *blended approach* leverages both native CSP service offerings and ISV offerings. It takes a CSP-centric or ISV-centric approach, depending on where the center of data gravity lies between these components. Point solutions may be replaced with CSP or third-party ISV alternatives. For example, the data warehouse component on Amazon Web Services (AWS), Google Cloud Platform (GCP) or Microsoft Azure could be replaced with a third-party offering like Databricks, Snowflake or Teradata. Similarly, a blended ISV-centric ecosystem approach like Cloudera Data Platform could be augmented by native CSP offerings.

When a blended approach is followed, there are two primary implications:

- The expectation of integration ease with the surrounding data ecosystem is not a given. CSPs will often prioritize connectivity and seamless integration within their own service product portfolio first, and third-party ISV offerings may lag behind. Similarly, in a blended ISV-centric approach, priority will be given to managing data within the ISV offering.
- Third-party ISV offerings have historically taken a broader view of the cloud. They often run on more than one cloud and include multicloud or intercloud capabilities (see Understanding Cloud Data Management Architectures: Hybrid Cloud, Multicloud and Intercloud). An ISV data integration tool, for example, will see various clouds and the offerings that run on them as additional sources or targets to support. An ISV metadata management tool views each offering in a variety of clouds as simply another source of data. Native CSP offerings have historically had a more myopic view of the world because they focused on getting data into and managing data in their own cloud.

This focus has started to shift though. CSPs are increasingly taking a broader view and introducing offerings that can access data in other clouds and on-premises environments for metadata purposes, as well as for querying and analytics purposes. Examples include Google BigQuery Omni and Microsoft Azure Purview.

Most notable, however, is that adoption of full CSP native deployments is expected to increase by 11% over the next three years (see Figure 4). This indicates a trend toward CSP-led approaches with a corresponding decline in blended approaches, and it reflects the growing maturity of native CSP offerings. In short, most components of these offerings are now “good enough” to meet end-user needs and SLAs, thus reducing the requirement to look elsewhere for core functionality. Interestingly, self-contained ISV deployments remain relatively static, reflecting the challenge of replacing a preintegrated all-in-one ISV platform or migrating to a CSP or blended approach.

Figure 4: Data Ecosystems Adoption Increasing 11% Over the Next Three Years

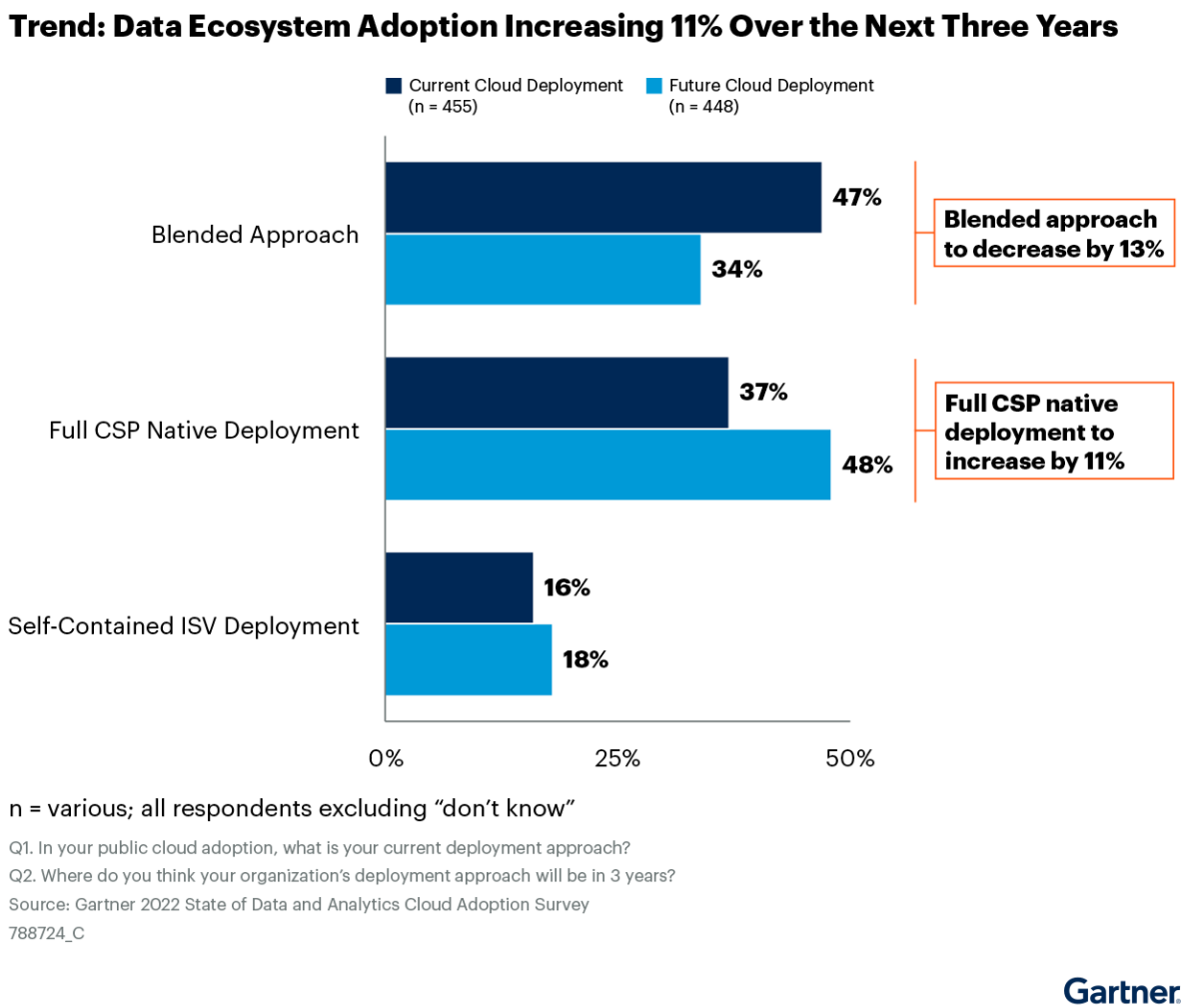


Table 2 summarizes core strengths and weaknesses of each delivery style.

Table 2: Data Ecosystem Strengths and Weaknesses

(Enlarged table in Appendix)

	Strengths	Weaknesses
Full CSP Native 	■ Deepest integration "out of the box" with a CSP environment and approach ■ Lowest integration cost ■ Best potential for rapid time to value for "greenfield" implementations	■ CSP lock-in concerns ■ More difficult to integrate with data residing in other clouds ■ May require some functionality trade-offs
Blended 	■ Best potential for broadest functionality and capabilities because less capable native offerings can be replaced with alternatives ■ Common path to distributed data ecosystem	■ Initial integration with the combined CSP/ISV offerings may require more effort and cost
Self-Contained ISV 	■ Easiest "lift and shift" from on-premises into the cloud ■ Best for end users with an existing strategic commitment to an ISV's offerings	■ More effort may be required to extend to surrounding CSP ecosystem ■ ISV lock-in concerns

Recommendations for data and analytics leaders:

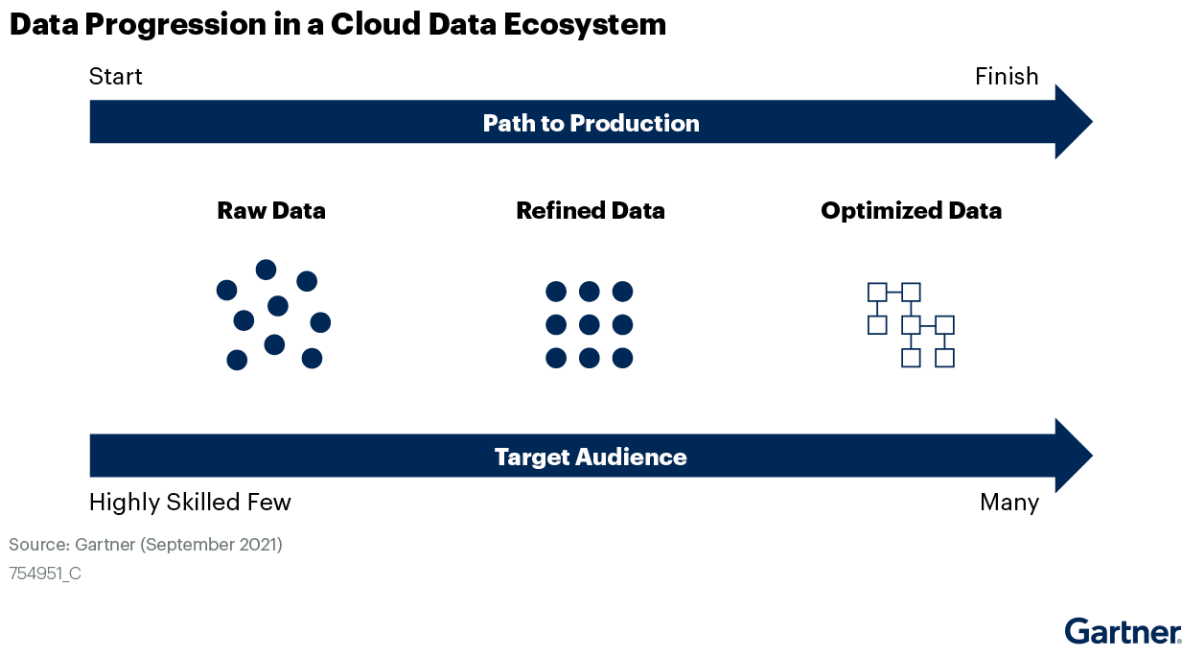
- Evaluate CSP native ecosystems on both immediate capabilities to address current needs and long-term vision. Many data ecosystem approaches are still maturing and may not provide a fully comprehensive environment yet.
- Use a self-contained ISV or blended data ecosystem when there is a strategic investment and relationship with the vendor, and when continuity with on-premises or other cloud environments is important.

- Assess ISV offerings to determine what specific value they add over and above what is available from CSP native offerings. Weigh this differentiated functionality against the increased overhead of integration, governance and financial governance when deciding whether or not to use a blended approach.
- Do not put much emphasis on vendor lock-in concerns (either CSP or ISV). In reality, you will have some degree of lock-in, regardless of which vendors you choose. Focus on capabilities and placing your point of lock-in where it will be most beneficial to you.

Active Metadata and Integration Capabilities Are the Foundation of Data Ecosystems

Data ecosystems define and strengthen the path to production for data by providing a cohesive environment that builds on common metadata and augmented data management capabilities. Data progresses from its raw state into a refining process and ultimately to production-optimized delivery (see Figure 5).

Figure 5: Data Progression in a Data Ecosystem



Data ecosystems build connections between the exploratory systems of innovation and discovery that explore unknown data and unknown questions and also between production systems that deliver on known data and known questions (see Data and Analytics Essentials: Logical Data Warehouse).

The unification of disparate data sources, types and deployment environments starts with metadata. Understanding the location, quality, frequency of use/access and security aspects of data is essential to the effective use of that data. As such, metadata and governance systems cannot afford to take a myopic view of data and analytics landscapes. They must seek out data of relevance wherever it resides (see The State of Metadata Management: Data Management Solutions Must Become Augmented Metadata Platforms). Data fabrics use active metadata analysis to deliver the most optimal cloud data ecosystem rendering for the use case demand.

For the data ecosystems to realize their full potential, data fabrics must provide certain key capabilities (see Figure 1). These include the ability for data ecosystems to utilize augmented data catalogs to create a unified inventory of metadata spread across distributed data stores. These capabilities must go beyond simply collecting technical metadata to also encompass performance, social and even business metadata.

Once the metadata has been collected, data fabrics must also provide capabilities to perform continuous analysis of this metadata to automate parts of data integration design and delivery. As data ecosystems begin to integrate (and analyze) metadata across a multicloud/hybrid ecosystem — and use this “active metadata” for decision insights and automation — they leverage the data fabric design patterns to support augmented data management and become increasingly self-optimizing.

Recommendation for data and analytics leaders:

- Plan your data ecosystem architecture by focusing on metadata and governance capabilities in distributed data ecosystem environments associated with data fabric design.
- Evaluate cloud service provider (CSP) native ecosystems on both immediate capabilities to address current needs and long-term vision. Many data ecosystem approaches are still maturing and may not provide a fully comprehensive environment yet.

Evidence

[1] Forecast: Public Cloud Services, Worldwide, 2020-2026, 4Q22 Update; Forecast: Enterprise Infrastructure Software, Worldwide, 2020-2026, 4Q22 Update; Market Share: All Software Markets, Worldwide, 2021

[2] **2022 Gartner State of Data and Analytics Cloud Adoption Survey:** This survey was conducted to validate and understand how end-user organizations are practicing and planning their journeys to perform data and analytics in the cloud, and what the key drivers are for such a transition. The research was conducted online from October through November 2022 among 461 respondents from North America, EMEA and Asia/Pacific. The respondents were from the manufacturing, natural resources, healthcare provider, banking and finance, IT, retail and wholesale, government, education, media and communications, transportation, and utility industries, in organizations with more than 250 full-time employees. The respondents were screened for being in director roles and higher, having responsibility for adopting cloud or planning it, and having some visibility or involvement in financial decision making for cloud adoption. Disclaimer: Results of this survey do not represent global findings or the market as a whole, but reflect the sentiments of the respondents and companies surveyed. There is no need to italicize this.

Document Revision History

The Impacts of Emerging Cloud Data Ecosystems: An Architectural Perspective - 9 September 2021

Cloud Data Ecosystems Emerge as the New Data and Analytics Battleground - 29 January 2020

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Innovation Insight: Data Ecosystems Will Reshape the Data Management Market

Predicts 2023: Data Management Solutions Finally Leverage Foundational Concepts

How to Plan for Optimal Multi Cloud and Intercloud Data Management

Understanding Cloud Data Management Architectures: Hybrid Cloud, Multicloud and Intercloud

Data and Analytics Essentials: Data Fabric

Data and Analytics Essentials: Logical Data Warehouse

Data and Analytics Essentials: Cloud

Forecast: Public Cloud Services, Worldwide, 2021-2027, 1Q23 Update

Forecast: Enterprise Infrastructure Software, Worldwide, 2021-2027, 1Q23 Update

Market Share: All Software Markets, Worldwide, 2021

Table 1: Top Impacts and Recommendations for Data and Analytics Leaders

Impacts ↓	*Recommendations* ↓
■ Early adopters of data ecosystems may find that end-to-end preintegrated data ecosystems in the cloud don't yet suit their current urgent needs, driving them to blended approaches.	■ Define and document what you expect CSPs to deliver as part of your data management landscape, and what capabilities you expect to obtain from third-party ISVs — especially with respect to foundational, enabling components like metadata management and data integration.
■ Data gravity is driving the choice of data ecosystem delivery styles.	■ Evaluate data ecosystems based on their ability to resolve distributed data challenges and to access and integrate with data sources outside of their immediate domain through orchestration and optimization capabilities.
■ Active metadata and governance are the foundation of data ecosystems and drive increased focus on data fabric design.	■ Plan your data ecosystem environment by focusing on the metadata and governance capabilities provided by the underlying data fabric components on which it rests.

Source: Gartner (April 2023)

Table 2: Data Ecosystem Strengths and Weaknesses

	Strengths	*Weaknesses*
Full CSP Native 	■ Deepest integration “out of the box” with a CSP environment and approach ■ Lowest integration cost ■ Best potential for rapid time to value for “greenfield” implementations	■ CSP lock-in concerns ■ More difficult to integrate with data residing in other clouds ■ May require some functionality trade-offs

	Strengths	Weaknesses
Blended 	■ Best potential for broadest functionality and capabilities because less capable native offerings can be replaced with alternatives ■ Common path to distributed data ecosystem	■ Initial integration with the combined CSP/ISV offerings may require more effort and cost

	Strengths	Weaknesses
Self-Contained ISV 	■ Easiest “lift and shift” from on-premises into the cloud ■ Best for end users with an existing strategic commitment to an ISV’s offerings ■ Potential cloud portability ■ Less complex integration effort	■ More effort may be required to extend to surrounding CSP ecosystem ■ ISV lock-in concerns

Source: Gartner (April 2023)

Actionable, objective insight

Position your organization for success. Explore these additional complimentary resources and tools for Data & Analytics leaders:

Research

3 Essentials for Starting and Supporting Master Data Management

Adopt an architecture most suited to fulfill your current and future goals.

Download Now

Webinar

Is Your Data AI Ready? Critical Questions You Must Ask Data Engineers

Explore the interactive relationship between traditional and AI data processes.

Watch Now

eBook

What Role Data Fabric Will Play in Your Data Management

Assess what data fabric means to your business.

Download Now

eBook

Innovation Insight: How Generative AI Is Transforming Data Management Solutions

Transform the data management practice with GenAI.

Download Now

Already a client?
Get access to even more resources in your client portal. Log In

Connect With Us

Get actionable, objective insight that drives smarter decisions and stronger performance on your mission-critical priorities. Contact us to become a client:

U.S.: 1 855 811 7593

International: +44 (0) 3330 607 044

Become a Client

Learn more about Gartner for Data & Analytics Leaders
gartner.com/en/data-analytics

Stay connected to the latest insight

Attend a Gartner conference
View Conference

Gartner