Reference Architecture for Federated Analytics

Published 24 July 2023 - ID G00786991 - 28 min read

By Analyst(s): Christopher Long

Initiatives: Analytics and Artificial Intelligence for Technical Professionals; Evolve Technology and Process Capabilities to Support D&A

The growing adoption of multicloud and data mesh makes analytics architectures more likely to be distributed. This diversity is challenging for data and analytics technical professionals to govern and manage. Federated, analytics-product-focused architectures can enable agility and control.

Overview

Key Findings

- The diversity of analytics use cases has resulted in complex consumption patterns across the organization. Increasingly distributed analytics architecture — whether by design, acquisition or business unit autonomy — challenges businesses to deliver consistent, reusable metrics.
- Organizations increasingly need to employ multiple A&BI tools to satisfy the needs of diverse user groups. Therefore, a robust D&A architecture is needed to support multiple A&BI tools.
- Distributed analytics programs require a balance of control and agility. As use cases grow, the need for governance and guidance on capability execution intensifies to mitigate the adverse effects of duplicated effort, increased costs and development of analytical silos.

Recommendations

Data and analytics professionals, particularly architects, responsible for implementing and managing analytics architectures to support distributed analytical development and consumption should:

- Use the federated architecture model to enable analytics capabilities and create an analytics-product focus on the operationalization of artifacts shared across the enterprise. Develop an architectural blueprint to encapsulate baseline capabilities, policies and security for systematic implementation.

- Evaluate analytics needs of business units and developer groups to ensure appropriate technical capabilities and governance processes are provided in a distributed environment.

- Institutionalize analytics capabilities and governance within business units to promote agility while mitigating organizational risk. Leverage D&A governance as an organizational capability to enable developers, analysts and data scientists.

Problem Statement

Many organizations prefer to allocate data and analytics responsibilities among widely distributed business units or other teams. One way that technologists support this business preference is with data mesh, which requires federation at the data level.

Gartner defines data mesh as a cultural and organizational shift for data management focusing on federation technology that emphasizes the authority of localized data management.

As part of this domain-centric approach to data management, data mesh is intended to enable access to and reusability of data throughout the enterprise. Although this management technique is helping organizations focus on managing data across the enterprise, the same management techniques are not always applied to analytical assets. Analytics assets are often locally owned and managed without the same level of federated governance found in data mesh principles.

In many organizations, semantic layers have been seen as a solution to deliver consistent analytics across disparate data. However, attempts at creating and maintaining a central canonical semantic layer for analytics have often failed in the past, and have known failings, including:

- They are data engineering heavy, often requiring significant IT resources to set up and maintain

- They are often not business-user-friendly and, therefore, do not evolve at the pace demanded by the enterprise

In response to the challenges with centrally administered semantic layers, the rise of self-service analytics tools enables technical and nontechnical users alike to connect with data, model, visualize, and share analytical data across the organization. Largely enabled through the ease of implementation by the SaaS offerings of modern A&BI tools, organizations find themselves instituting decentralized analytics with haphazard plans and minimal governance. In some cases, this may even be accomplished without support from IT, which can lead to:

- Uncontrolled system implementations
- Fragmented analytics governance
- Analytical silos
- Inconsistency across reports and calculations
- New (unexpected) maintenance requirements for IT

In some organizations, the unfettered growth of self-service has produced a minefield of analytical artifacts, including ABI tool-specific semantic models that constitute significant risk and technical debt. For further details on semantic layers, their common architectures, and challenges, see Demystifying Semantic Layers for Self-Service Analytics.

Increased demands for access to data and to deliver insights to the business are principal drivers in the growth of analytical artifacts developed throughout the organization. The range of user personas requiring access to and development of analytics falls into five broad categories:

- **Analytics Consumers:** Generally passive users that access existing content in the form of reports and dashboards. Sometimes, these users may use existing data to develop new visualizations and share the resulting reports. These users do not create new analytics and generally interpret data as part of decision making or performance monitoring.
- **Business Intelligence Analysts:** Active users of analytic datasets. These users will often develop and publish new content in the form of analytical models, reports and dashboards based on existing, known data.

- **Citizen Data Scientists:** Often a business analyst with additional ad hoc query capabilities within toolsets to build and deploy analytical models. Additionally, these users may incorporate augmented analytics, often through the use of AutoML capabilities.
- **Data Scientists:** Most technical data analysts, often with direct connectivity to raw source data with the purpose of finding new insights using new/unknown data.
- **Business Application Developers:** These users are focused on integrating data and analytics directly into application workflows. The results of these users place analytics closer to the end user making decisions during operational activities.

For more details on these analytics user groups, see Note 2.

A sustainable solution to these problems is an architecture that combines enterprise data and analytics governance with some autonomy in business units to own and govern their domain of analytics. The primary goal of this document is to provide readers common architectural patterns for deploying analytics capabilities while addressing the above-mentioned concerns.

The target audience for this document includes heads of data management, heads of analytics and BI, enterprise architects, and data and A&BI solution architects. This research note may also be leveraged by chief data officers, chief analytics officers and project managers to better understand architectural patterns for implementing analytics in distributed environments.

This research focuses on three core topics when building an architecture for federated analytics:

- Identifying sample architectures for federated analytics
- Identifying guiding principles for the architecture and mapping analytics capabilities
- Shifting to an analytics product focus for deployment of analytics

The Gartner Approach

As outlined above, many organizations today implement decentralized analytics by way of deploying self-service capabilities. However, this decentralization often breeds analytical silos, redundancy and increased operational costs. By contrast, federation works to provide oversight and guidance to decentralized groups to deliver expected levels of consistency, interoperability and cost reduction and to reduce risk in the organization's analytical assets.

To provide global organizational governance regarding analytics in a distributed environment, Gartner recommends a federated architecture to support analytics and BI. This architectural pattern blends both top-down and bottom-up approaches where global guidelines are developed and enforced at a global level and business units have some autonomy over their local analytical assets. In short, the federated approach provides:

- A standardized scalable architecture
- Consistency in analytical data definitions
- Agility to deliver multiple end-user analytics products and services
- A strong partnership between IT and business
- Supported and governed self-service analytics platforms

By implementing a federated framework, D&A professionals can balance organizational control and business agility. The reference architecture then guides technical teams about where new capabilities may be implemented to support the evolution of analytics in the organization.

Along with the federated analytics architecture, Gartner recommends a product approach to delivery of analytics artifacts. As analytics artifacts are delivered to consuming users, more must be taken into consideration than just the object itself. Organizations taking a product approach to analytics delivery will find increased trust across domains and reduced redundancies in development.

The Guidance Framework

Prework

Why Implement a Reference Architecture for Federated Analytics?

The reference architecture is a framework to design, build and deploy systems that are scalable, interoperable and supportable. As a federated reference architecture, it enables organizations to:

- Leverage the collective potential of their distributed resources
- Reduce the pressures on central teams for development
- Promote collaboration across domains
- Empower domain users to act at the pace the business demands

> **A federated model is a pattern within the enterprise architecture that allows interoperability and information sharing between semiautonomous, decentrally organized LOBs, information technology systems and applications.**
>
> — *Source: Gartner*

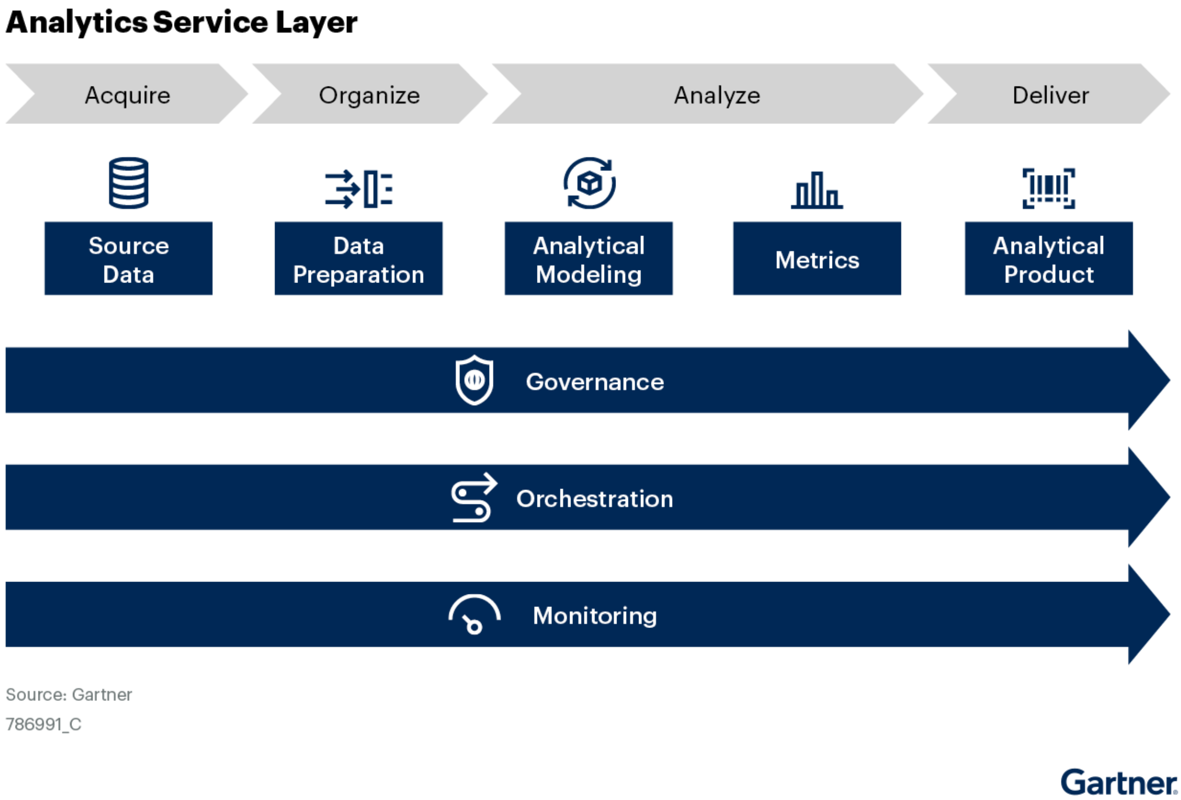
The core of this research focuses on identifying the components of and designing the reference architecture for federated analytics. After designing a reference architecture for the organization, data and analytics professionals will need to implement the architecture. See Note 1 for implementation information.

A Reference Architecture for Federated Analytics

Analytics Service Layer

The central element of the federated analytics architecture is the analytics service layer. This layer represents the combined capabilities for analytics groups to connect with data and to model, govern and deliver analytics products to the organization. Figure 1 demonstrates the pathway for data through this layer as analytics are delivered to the organization.

Figure 1: Analytics Service Layer



The analytics pipeline combines the building blocks of the analytical services layer including data preparation, analytical modeling and metrics development. This pipeline is underpinned throughout the data's journey by governance and security, orchestration, and monitoring.

Acquire and Organize

The acquisition part of an analytics architecture involves identifying data sources, types of data contained, and determining the quality of data.

Business analysts and data scientists spend a majority of their time preparing data. This is commonly known as "data wrangling." Data preparation is the process of connecting, combining, structuring and transforming data for analysis. These processes are aimed at data staging, transforming and addressing data quality issues, including:

- Identifying/removing null values
- Converting data types
- Identifying/organizing duplicate records

- Finding outliers

Data preparation capabilities are generally implemented as part of data integration tools, BI tools, data science/machine learning tools or stand-alone tool implementations.

Analyze and Deliver

The analyze stage of the analytics service layer includes the following capabilities:

- **Analytical data modeling:** Tools, features and functions are used to connect to and combine data from multiple sources, creating dimensional/tabular models to be analyzed or consumed visually.
- **Metrics stores:** A metrics store allows users to create and define business metrics as code; govern those metrics from data warehouses; and serve downstream analytics, data science and business applications.
- **Machine learning and AI:** Advanced analytics capabilities combine augmented data preparation, predictive modeling and prescriptive analytics.
- **Data exploration:** These capabilities are used to examine, summarize and visualize patterns within a dataset.

A common goal of the analyze and deliver stages is to provide analytical data models for consumption across the organization. The purpose of analytical data models (semantic layer) is to represent corporate data using simple business terms like “customer” or “account” in place of cryptic column names.

As detailed in Demystifying Semantic Layers for Self-Service Analytics, semantic layers perform the following core functions:

- Translate the underlying database structures into business-user-oriented terms and constructs, which are then intuitive to business users.
- Provide a mechanism to define and store calculations and business rules.

These semantic layer artifacts may be one part of analytics products delivered within the organization. Additional resulting analytics products shared with the organization may include:

- Metric definitions

- Reports
- Dashboards

The resulting analytics products may be deployed for use within a particular business unit or enterprisewide. They may even be monetized through delivery to external consumers. Further details on analytics products are discussed in the Related Guidance section below.

Federated Architecture Examples

As outlined, the analytics service layer represents a combined set of capabilities to connect with, prepare, model and deliver analytics to the organization. The importance of developing the analytics service layer blueprint is so that these capabilities can be implemented systematically and scaled to fit the growing needs of the organization. Implementing this layer across business domains results in a mesh-like architectural pattern similar to those seen in data mesh architectures.

This research identifies two common patterns organizations may use to create a federated architecture for implementing distributed analytics. These reference patterns include:

- Federated analytics based on decentralized data
- Federated analytics based on centrally managed data

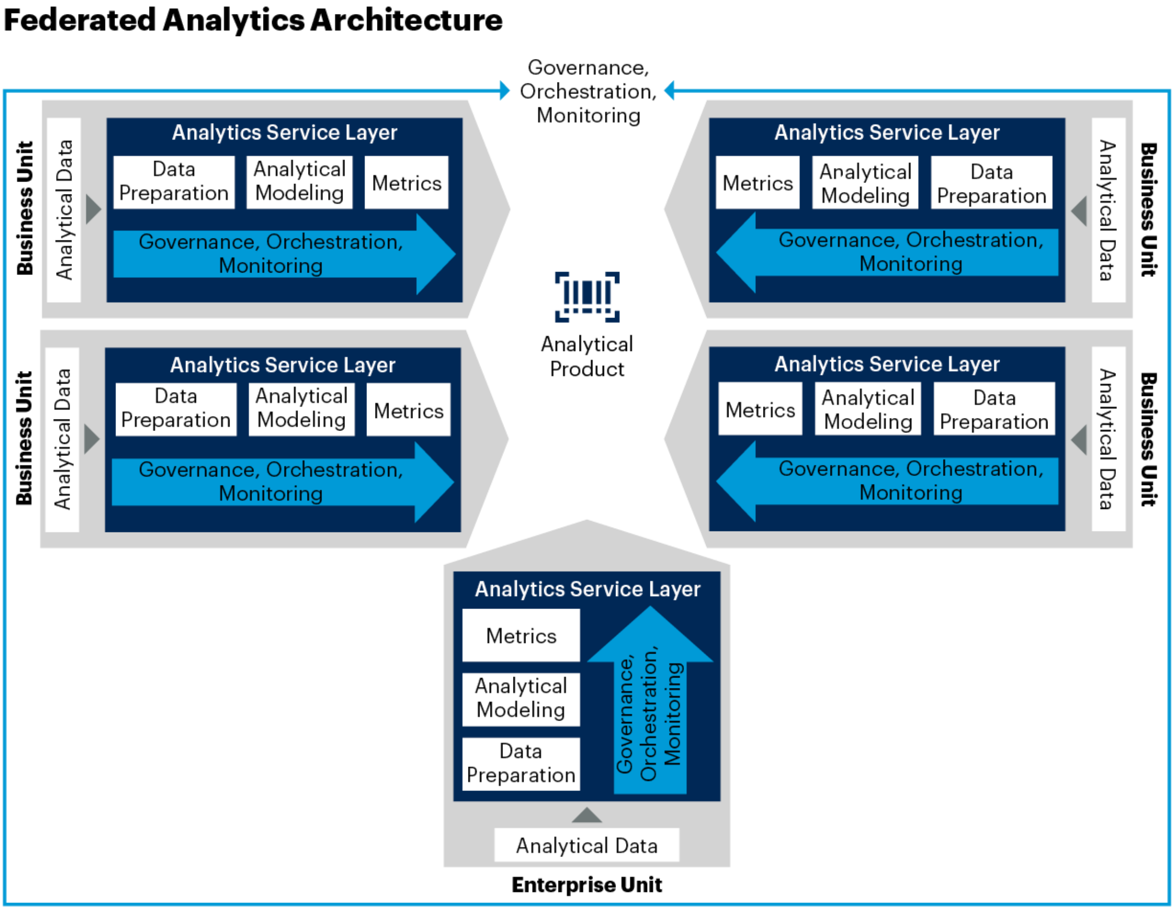
These base patterns may be adapted within your organization and are intended to provide a common framework to establish systematic, scalable implementations.

Federated Analytics Based on Decentralized Data

This federated analytics architecture builds on the increasingly distributed development, management and deployment of analytics assets across the organization. The result is a mesh-like architecture pattern that closely resembles the data mesh paradigm driven by federating data management responsibilities and through distributed governance (see Data and Analytics Essentials: Data Fabric and Data Mesh). The federated analytics architectural pattern empowers business domains to manage analytics artifacts and products with guidance from a centralized IT group (without the associated bottlenecks).

Figure 2 depicts a federated analytics architecture based on a distributed data management architecture. In this pattern, business domains share analytics as products to be consumed and reused across the enterprise. In this mesh-like architecture, analytics and data management capabilities are deployed both centrally and in business units across the organization through the analytics service layer.

Figure 2: Federated Analytics Architecture



Source: Gartner
786991_C

Gartner

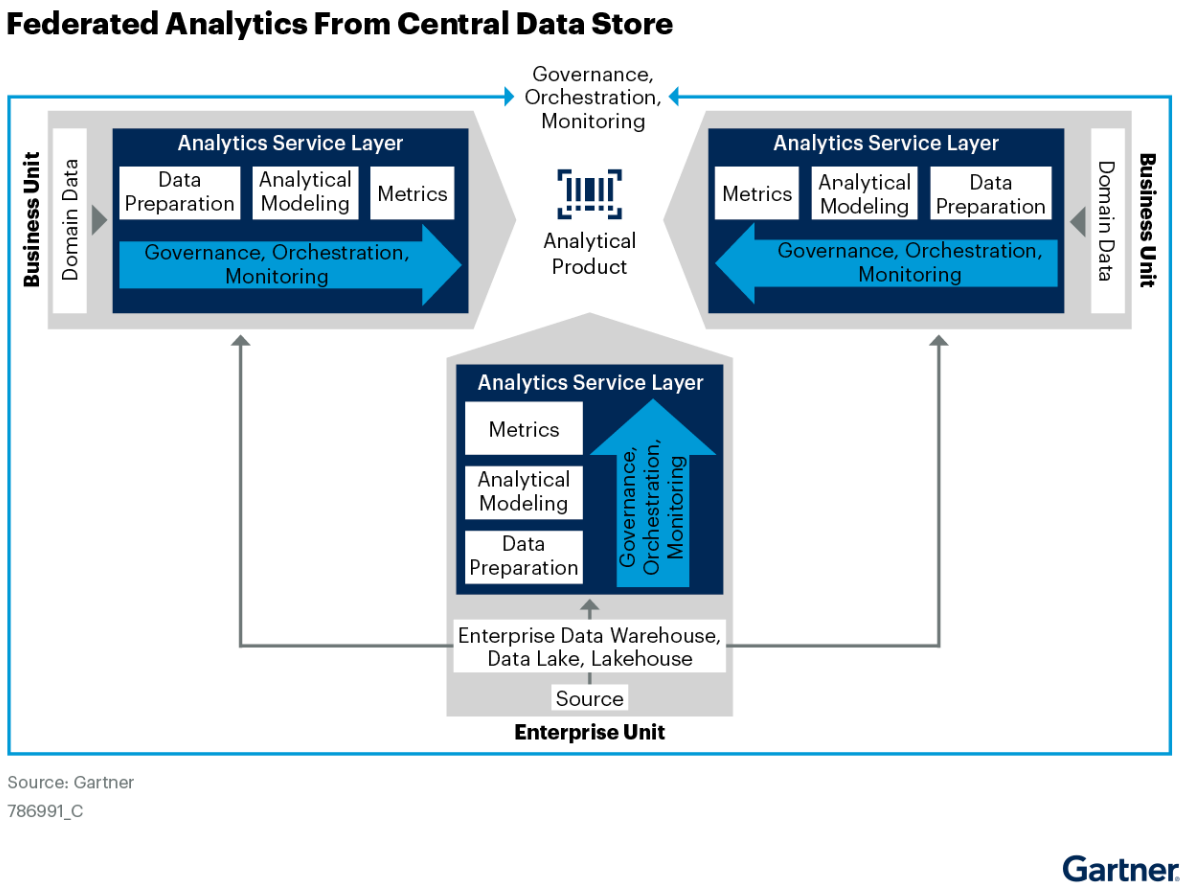
The wide-spread adoption of self-service analytics, coupled with the growing implementation of data mesh architectures, empowers business units to control and cultivate their domains of data and analytics and share resulting content as products to the wider organization.

Federated Analytics Based on Centrally Managed Data

Although a federated analytics architecture is a natural outcropping of a data mesh architecture, this federated analytics pattern is also possible from centralized data stores such as an enterprise data warehouse, data lake or lakehouse. In this architectural pattern (see Figure 3), business domains develop and manage analytics semiautonomously with some oversight. This oversight often includes defined analytics tools from a central authority.

Note that, in this pattern, business domains principally connect with and model enterprise managed data while leaving some room to connect with domain-centric data. It is common for business units to include data specific to their domains to add context to the enterprise data used in analytics. This is a distinction that promotes the agility required by business units to develop and deliver analytics without the bottlenecks commonly found in highly centralized architectures.

Figure 3: Federated Analytics From Central Data Store



Gartner client interactions suggest this architectural pattern commonly emerges organically as organizations implement self-service analytics and multiple BI tools. However, as stated previously, this often emerges without specified guidance or governance leading to inconsistencies in implementation, duplication of efforts and a lack of interoperability in the analytics products delivered. The implementation of a reference architecture assists organizations with the planning necessary to scale the deployment of capabilities across the organization.

In both of the federated architecture examples provided, organizations decentralize domain analytics ownership, management and governance to business units for domain-centric concerns while guiding on enterprise-level standards and policy. This mesh-like architecture is less a technical construct and more a logistical (or organizational) construct requiring matured capabilities in data and analytics management and governance programs throughout the organization.

Successful implementations of this analytics mesh architectural pattern require the following:

- **Mature, federated data and analytics governance:** Governance policy, procedure and tools are implemented at both global and local levels throughout the architecture. Collaboration between business unit and enterprise governance teams ensures that analytics products delivered meet both specific business needs and adhere to the global principals to meet organizational strategic goals.
- **High data literacy:** Both developer and consuming users have a comprehensive understanding of the analytics they manage, deploy and consume. Increased data literacy directly improves the quality of analytics produced and builds trust for analytical consumers. For further details, see Tackle Data Literacy Head-On to Avoid Data and Analytics Program Failure.
- **Willingness to accept responsibility:** Because this represents a cultural shift in management, it is imperative that organizations are willing to accept and be responsible for the management and governance of analytics assets.

Following an understanding of the reference architecture for federated analytics, the next section focuses on identifying the guiding principles for the architecture.

Analytics Architecture Guiding Principles

D&A, BI and AI architects within the organization responsible for designing an analytics architecture must establish a set of guiding principles the pattern must deliver on. These principles represent the highest-level guidelines for governance of the analytics architecture. Guiding principles must be connected to business goals and provide actionable guidance.

Guiding principles should be:

- Clear and unambiguous
- Prescriptive
- Broadly applicable across the reference architecture

Additionally, the guiding principles will be used to identify the technical capabilities to be considered within the architecture. Data and analytics architects should consider the following categories as part of outlining the guiding principles as part of their architecture for federated analytics:

- Effective Data Storytelling
- Analytical Data and Metrics Modeling
- Governance Framework
- Accessible and Interoperable Analytics

Effective Data Storytelling

Data storytelling is the practice of building a narrative around a set of data and its accompanying visualizations to help convey the meaning of that data in a powerful and compelling fashion. Analytics storytelling combines characteristics as shown in Figure 4.

Figure 4: The Data Storytelling Flow

Technical professionals delivering on the principle of effective data storytelling will need to consider technical capabilities that include:

- Enterprise reporting
- Interactive dashboards
- Automated insights
- Augmented analytics

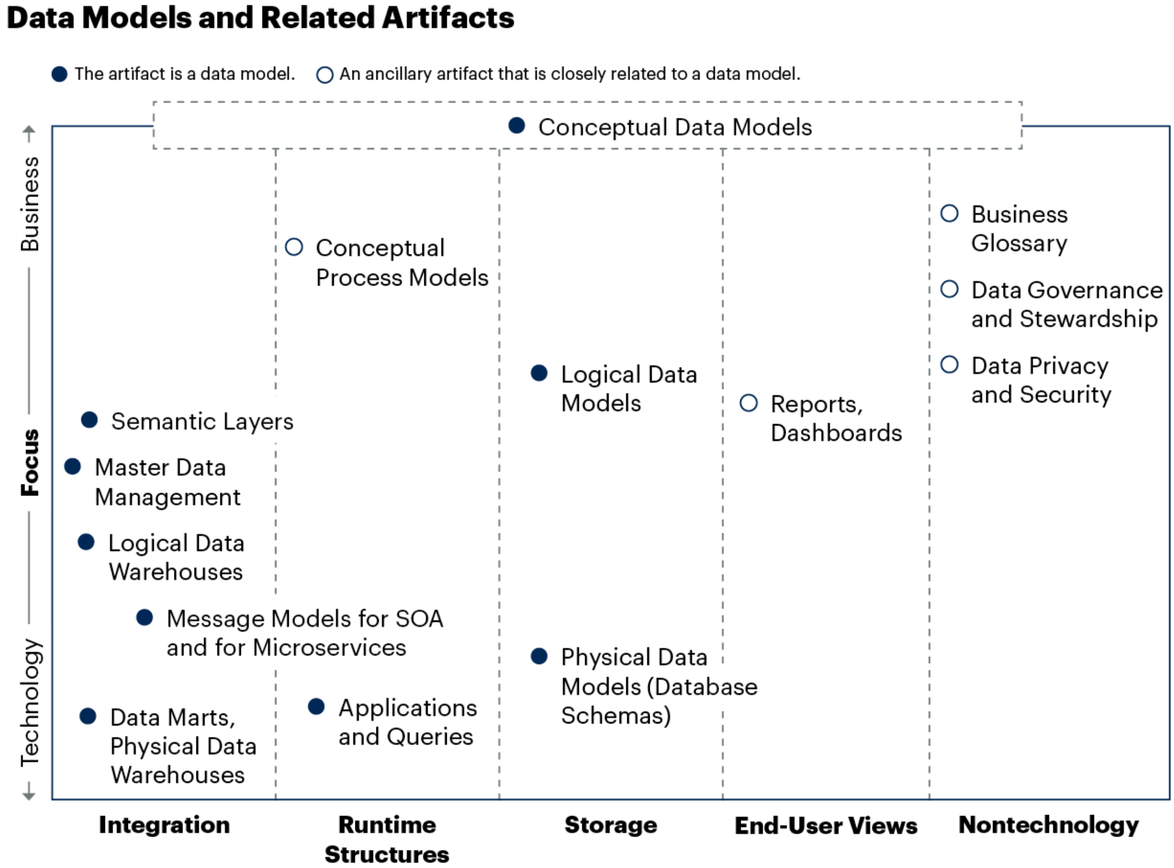
D&A technical professionals must also consider the uses of generative AI for data storytelling. Generative AI enables more self-service users not only to develop compelling interactive content, but also to enhance the end-user experience through natural language narratives and contextual "next step" prompts. As of writing, generative AI is being incorporated into many ABI vendors' tools. Review Quick Answer: What Are the Short-Term and Midterm Implications of ChatGPT for Data and Analytics? for additional insights into how this technology can impact the analytics architecture and product delivery.

Analytical Data and Metrics Modeling

Analytical data and metrics modeling is the practice of designing data models and developing metrics for analytics use cases. This includes the development of data models in the data layer (such as in the data warehouse) and in the analytical layer through data virtualization and BI tools. This modeling involves creating logical representations of data entities, relationships and attributes to support varying analysis activities. Analytical modeling includes capabilities that allow enterprise and self-service developers to connect with, prepare and summarize data, and build metrics to be used within domains and shared across the organization.

Figure 5 identifies multiple artifacts resulting from data modeling activities.

Figure 5: Data Models and Related Artifacts



Source: Gartner
753676_C

For additional details on data models, see Data Modeling to Support End-to-End Data Architectures.

As part of the development process, whether at an enterprise or self-service level, effective analytics architectures will include processes to operationalize analytics. These processes create delivery and change management of your analytics artifacts. See Demystifying XOps: DataOps, MLOps, ModelOps, AIOps and Platform Ops for AI for additional details on applying DevOps-type processes to data and analytics development.

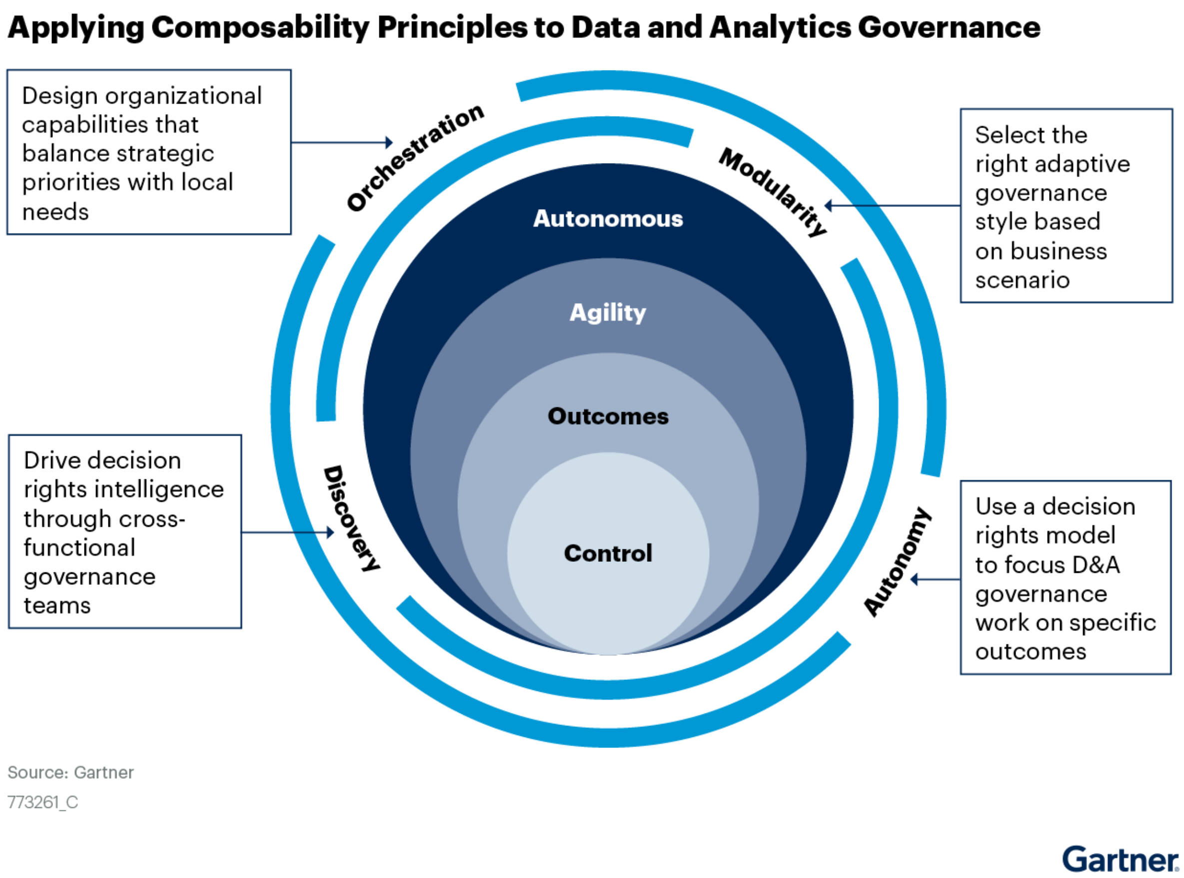
Governance Framework

> **Data and analytics governance is the specification of decision rights and an accountability framework to ensure the appropriate behavior in the valuation, creation, consumption and control of data and analytics.**

As a sociotechnical system, governance is not purely a technical platform, but a series of people, policies and procedures that is supported for operationalization and automation by technology capabilities. Federated analytics architectures require mature governance programs within the organization. The governance framework will provide a set of processes, implemented and used by stakeholders, that leverage technology to ensure that critical data is protected and well-managed. Adaptive methods of governance provide global-level guidance to the organization while providing for domain-level discretion that is foundational to successful federated architectures.

As organizations move to federated architectural patterns, governance practices must shift along with them. Figure 6 highlights how composable principles may be applied to adaptive governance frameworks to balance your data and analytics governance operating model.

Figure 6: Applying Composability Principles to Data and Analytics Governance



For more information on creating a data and analytics governance program, see:

- Building a Comprehensive Governance Framework for Data and Analytics
- Data and Analytics Governance Approaches for the Technical Professional
- Quick Answer: How Can I Apply Composable Design Principles to Data and Analytics Governance Organization Capabilities?

Accessible and Interoperable Analytics

Distributed analytics, primarily through self-service, has given rise to analytics silos throughout organizations. Analytics silos are a natural and to-be-expected consequence of self-service analytics. Common causes of these silos can include:

- The use of multiple analytics and BI tools in the organization
- The lack of sharing analytics across business units

- The inability of users to find and consume analytics assets in the organization

Successful federated analytics is predicated on the ability to share, find and consume analytics products within the organization. Users must be able to share developments with consumers within their business unit and with the wider organization. This may come in the form of:

- Analytics hubs for sharing curated data, analytical models and metrics. Hubs themselves may be a platform provided centrally by IT or may even be maintained by the business unit depending on the overall data and analytics strategy within the organization. This will often be a means of providing data in a consumable form to be utilized or integrated in other data-oriented processes.
- Analytics catalogs for providing users access to reports, dashboards and metrics that are focused on end users or decision makers.
- Uniform, organizationwide naming conventions for datasets and analytical artifacts. Without such conventions, catalogs and hubs will be of little use. See Improve Data Literacy and Governance by Using Accurate, Meaningful Names for Data Artifacts for additional details.

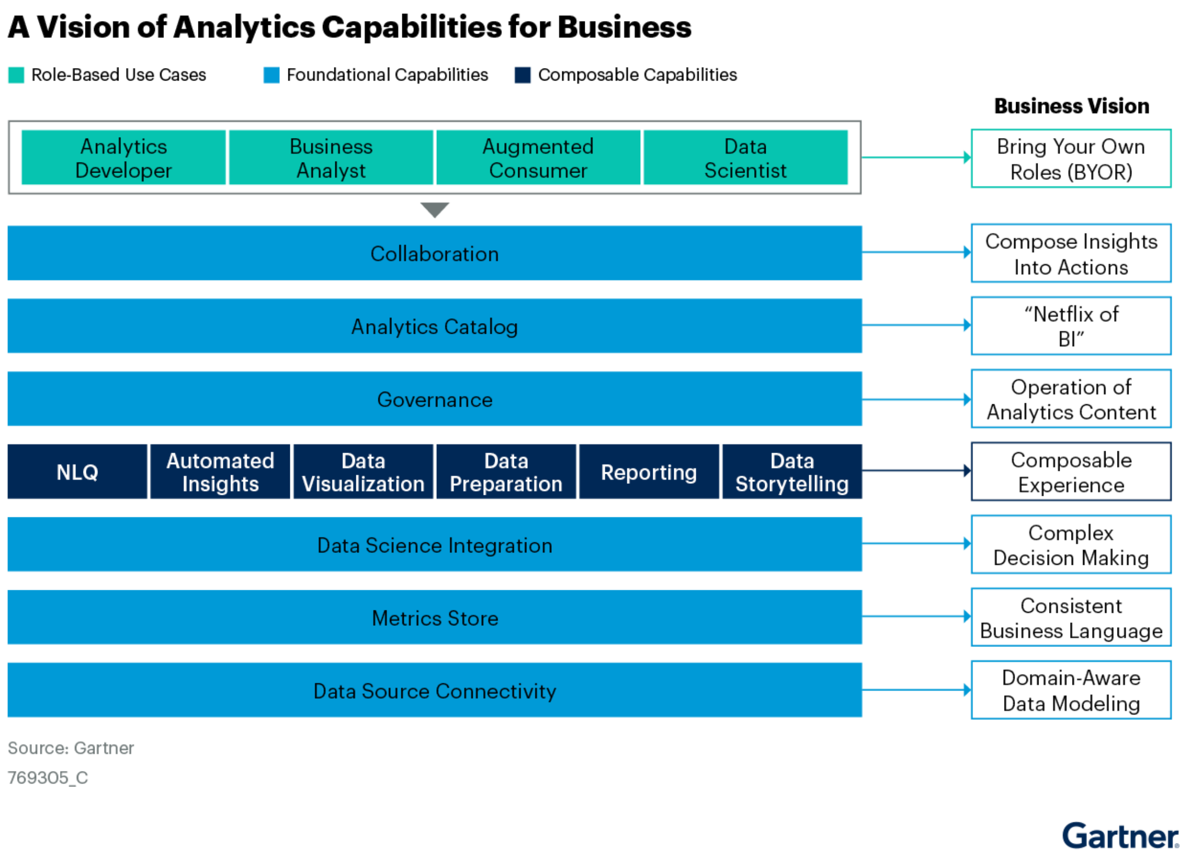
The goal of accessibility and interoperability as a guiding principle is to ensure consuming users know where to find the analytical products provided, prevent duplicated analytics development efforts and promote consistency in the use of analytics across the organization.

To operationalize these guiding principles, organizations must identify related technical capabilities. The next section identifies and maps technical analytics capabilities to the guiding principles that support the reference architecture.

Analytics Architecture Capabilities

Figure 7 details the distinct capabilities that will work to operationalize the previously identified guiding principles for your reference architecture.

Figure 7: A Vision of Analytics Capabilities for Business



Foundational Capabilities

Foundational capabilities are those solutions implemented to support analytics development and delivery across all potential use cases. These are capabilities that promote the standardization required for a federated system's success.

- **Analytics Catalog:** This refers to the product's ability to display analytic content to make it easy to find and consume. The catalog is searchable, makes recommendations and targets business users.
- **Collaboration:** Analytics collaboration is the application of collaboration capabilities to analytics workstreams for organizations that want to provide an environment where a broad spectrum of users can simultaneously co-produce an analytics project.
- **Data Science Integration:** These capabilities enable augmented development and prototyping of composable data science and machine learning (DSML) models by citizen data scientists and data scientists with integration into the broader data science and machine learning ecosystem.

- **Data Source Connectivity:** Data source connectivity capabilities enable users to connect to and ingest structured data contained in various types of storage platforms, both on-premises and in the cloud.

- **Governance:** Governance capabilities track usage and manage how information is shared and promoted.

- **Metrics Store:** The metrics store provides a virtualized layer that allows users to create and define metrics as code; govern those metrics from data warehouses; and service all downstream analytics, data science and business applications. This also includes capabilities such as goal management.

Composable Capabilities

Composable capabilities are those typically provided through the analytics and BI platforms that contribute to the analytics experiences for developers and consumers. D&A technical professionals can assess the needs of analytics personas and the composability of their ABI platforms to provide the necessary capabilities while retaining the ability to expand as maturity grows throughout the organization.

- **Automated Insights:** A core attribute of augmented analytics, this is the ability to apply machine learning (ML) techniques to automatically generate insights for end users (for example, by identifying the most important attributes in a dataset).

- **Data Preparation:** Data preparation includes support for drag-and-drop, user-driven combination of data from different sources, and the creation of analytic models (such as user-defined measures, sets, groups and hierarchies).

- **Data Storytelling:** Data storytelling is the ability to combine interactive data visualization with narrative techniques to package and deliver insights in a compelling, easily understood form for presentation to decision makers.

- **Data Visualization:** Data visualization involves support for highly interactive dashboards and exploration of data through the manipulation of chart images. Included is an array of visualization options that go beyond those of pie, bar and line charts, such as heat and tree maps, geographic maps, scatter plots and other special-purpose visuals.

- **Natural Language Query:** The natural language query (NLQ) capability enables users to ask questions of the data using terms that are either typed into a search box or spoken.

- **Reporting:** The reporting capability provides pixel-perfect, paginated reports that can be scheduled and bursted to a large user community.

The above definitions are provided by Gartner's Critical Capabilities for Analytics and Business Intelligence Platforms, which also provides insight on how modern ABI platforms implement and support these technical capabilities.

To summarize the technical aspects discussed, Table 1 maps the above capabilities to the federated architecture's associated guiding principles. Note that each capability may support more than one guiding principle.

Table 1: Connecting Technical Capabilities With Guiding Principles

(Enlarged table in Appendix)

	Effective Storytelling	Analytical Modeling	Governance and Security	Accessibility and Interoperability
Analytics Catalog		Y	Y	Y
Automated Insights	Y			
Collaboration		Y	Y	Y
Data Preparation		Y		
Data Science Integration		Y		
Data Source Connectivity		Y		
Data Storytelling	Y			
Data Visualization	Y			
Governance			Y	
Metrics Store		Y		
Natural Language Query	Y			Y
Reporting	Y			

Source: Gartner (July 2023)

Risks and Pitfalls

Federated Versus Decentralized Analytics

When using decentralized ownership in a distributed architecture, some organizations erroneously assume federation will arise organically. Decentralized analytics are often the outcome of distributed, autonomous implementations of analytics and BI tools across an organization. In the name of agility and reduction of bottlenecks, business units want this autonomy. However, without some common semantics or guiding principles, organizations will not fully realize the value and potential of these distributed groups.

Where decentralization promotes unfettered autonomy and agility, federation balances local autonomy with broader standardization and collaboration. Such balance allows organizations to scale the developments of individual business units across the wider enterprise.

Table 2 highlights key differences between federated and decentralized approaches to analytics.

Table 2: Federated Analytics Versus Decentralized Analytics

	Federation	Decentralization
In Practice	Systems and practices allow multiple domains to work as a whole focusing on achieving the business goals while retaining some autonomy regarding local decision making.	Systems and practices dictate that business domains work autonomously regarding local decision making.
Governance	There is a joint effort and agreement of governing principles from a central authority and business domains.	Control rests with business domains.
Decision-Making Authority	Central authority may make decisions affecting the organization as a whole. Business domains retain some autonomy for local decision making.	Decision making is distributed without a central authority.
Interoperability	Business domains agree to standards for development and sharing analytics.	Business domains develop analytics independently and define their own standards.
Coordination and Collaboration	A central governance group facilitates collaboration and coordination across business domains.	There is no central group for coordination. Collaboration is independent across domains.

Source: Gartner (July 2023)

Related Guidance

Take a Product-Centric Approach to Analytics

Organizations creating and implementing a federated analytics architecture should also combine this with a product approach to analytics asset delivery. Analytics assets delivered by business units can take many forms. The consumer of the asset may be a developer taking the analytical data for input for an additional use or may be an end-user consuming a dashboard for decision support. The key element to note is that an analytic product is not a specific type of delivered artifact. Nor is the product the artifact alone.

In addition to its inherent consumer focus, multiple characteristics will coalesce to create a complete product, including:

- **Meets business needs:** Analytics products should be designed to support business requirements. By building artifacts to support business needs, analytics developers can quantify artifact value and justify continued investment.
- **Defined ownership:** Analytics delivered to the organization, whether for internal business domain use or shared with the broader organization, must have clear ownership and responsibility. This ownership includes succession planning as individuals move around or out of the organization.
- **Supported:** As a product, there must be a support mechanism to ensure it is operating as expected without errors. As errors are discovered, clearly defined SLAs should be in place to ensure continued service to the consuming users.
- **Discoverable and shareable:** Developers of analytics artifacts must be able to share, within guidelines, with the organization. Examples include through analytics hubs providing subscription-based access or links to analytics workspaces. Users in the organization must also then be able to search for and identify the analytics assets delivered. When users are unable to locate analytics, there is an increased risk of redundant development efforts.
- **Governed:** As part of a mature data and analytics governance program, analytics products must adhere to the global standards and guidelines expected as well as any business-domain-specific requirements. This includes providing consuming users with expected and acceptable use cases for the analytics.

Analytics products are generally conceived and delivered for internal customers. Often, these characteristics are not fully considered or matured for these internal consumers due to resource constraints or shifting organizational focus. Analytics products may also be externally focused. Such products have several distinct concerns.

External-facing analytics:

- Are revenue generating and should follow a typical product-development life cycle
- May be subject to additional regulations
- Will require increased attention from D&A governance programs
- Are subject to lower tolerance (or complete intolerance) for downtime and require in-place disaster-recovery procedures and CI/CD processes that allow the analytical ecosystem to be improved without any interruption of service

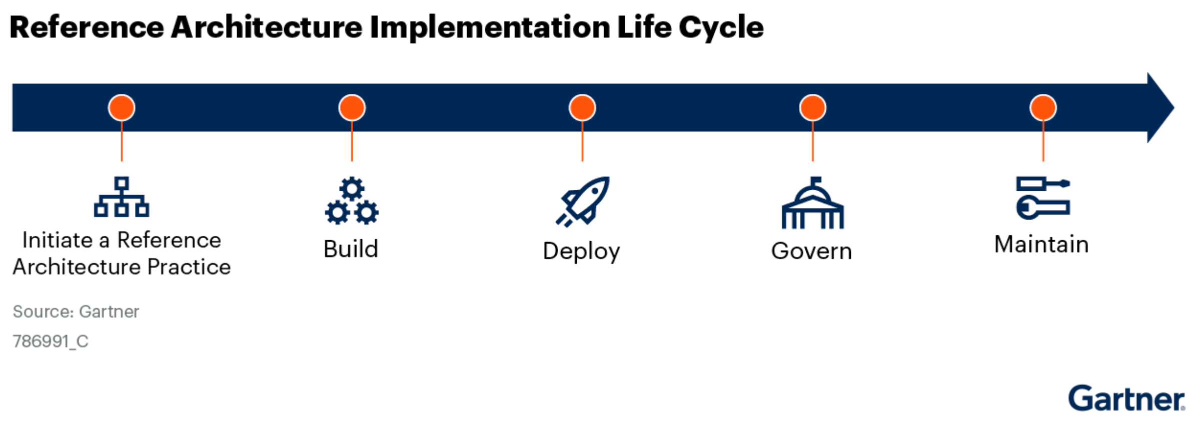
Gartner research shows organizations that need pipelines to support external analytics products often decide to maintain entirely separate technical stacks for those pipelines.

A product focus on delivery of analytics ensures that the artifact delivered meets the needs of its consumers and can be trusted to provide value to the organizations. Because in a federated architecture, ownership and management of analytics is distributed, governance and trust are essential to the success and acceptance of analytics delivered. Taking a comprehensive approach to delivering analytics not as artifacts, but as products, provides the assurance to users that these assets can be used reliably and built on throughout the organization.

Note 1: Implementing a Reference Architecture

This research focuses on identifying a reference architecture for federated analytics, its related guiding principles and capabilities. However, data and analytics leaders, architects, and technical professionals must also implement plans to build, deploy, govern and maintain these architectures in order to reach their full potential. Figure 8 identifies a typical reference architecture life cycle.

Figure 8: Reference Architecture Implementation Life Cycle



For additional resources on implementing reference architectures, see the following Gartner research:

- Reference Architecture Implementation Guide
- Reference Architecture Components Cheat Sheet

Note 2: Understand Your Analytics Users

A typical organization consists of the following four basic business user groups, which are grouped into two categories depending on the functions they perform:

Innovators and experts:

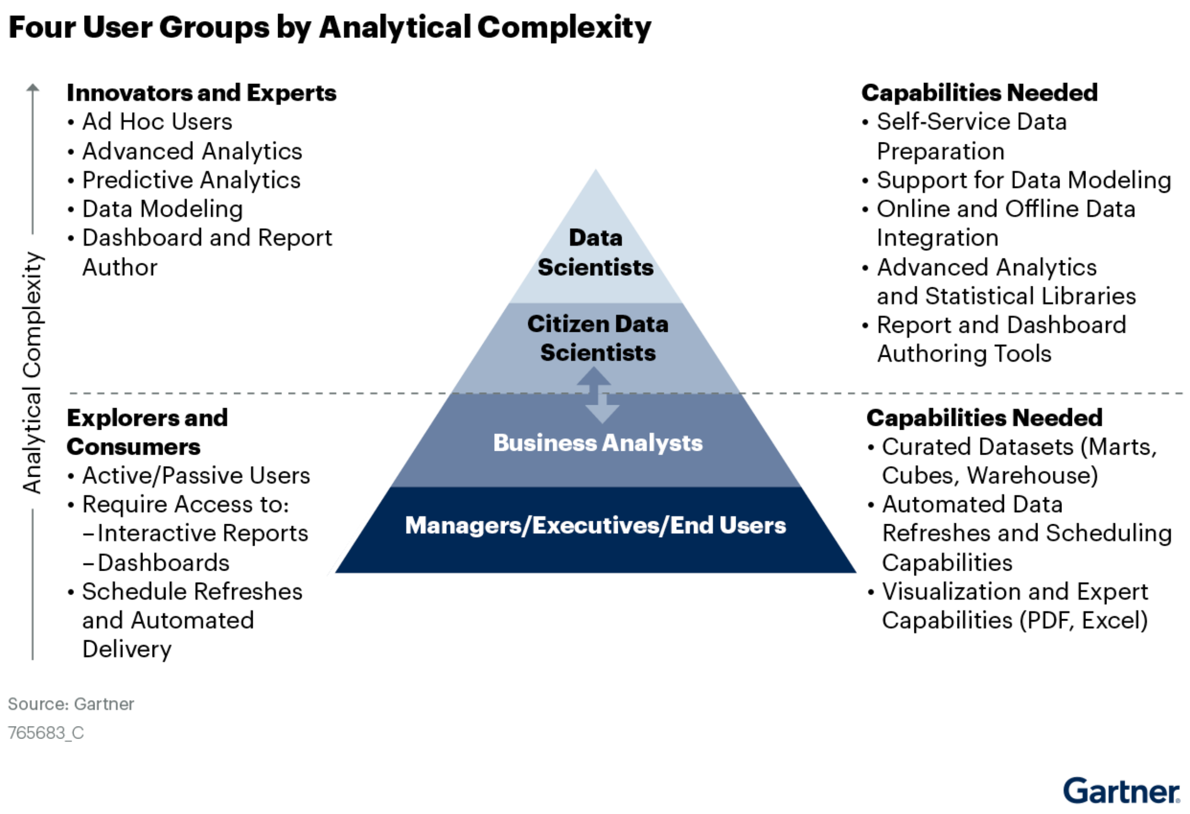
- Data scientists
- Citizen data scientists

Explorers and consumers:

- Business analysts
- Managers, executives and end users

Figure 9 describes the various business user groups identified by Gartner, organized by analytical complexity and skills.

Figure 9: User Group by Analytical Complexity



Note the bidirectional arrows in the above figure. These represent the blurred boundaries and range of skill sets between citizen data science and business analyst personas based on the vast spectrum of skills and analytical complexity between advanced citizen data scientists and novice business analysts. Business analysts are increasingly empowered to deliver augmented analytics due to the advancements within modern A&BI tools, thus blurring these analytical personas.

Data and analytics technical professionals must design an architecture for a diverse audience with different needs rather than relying on a notion of users with fixed boundaries of engagement. See Table 3 for a summary of these four main user personas and their capabilities in a self-service data and analytics environment. Also refer to Roles and Skills to Support Advanced Analytics and AI Initiatives for an examination of the skills scope for data scientists, citizen data scientists and related ML personas.

Table 3: Analytic Persona Model

(Enlarged table in Appendix)

Analytic Personas	Description	Capabilities Within the Self-Service Environment
Consumer	Consumers may be frontline workers, executives or external customers who view analytic content periodically. They are "conversational" from a data literacy perspective.	Consume and interact with production content. No formal content creation rights, and typically do not request them. Unable to validate or nominate any new content to be promoted to pilots.
Explorer	Explorers may be business users who are looking to do more diagnostic-type analytics, or junior analysts who are looking to expand their understanding of specific domains. They are "literate" from a data literacy perspective.	Able to duplicate and then modify production content. Able to nominate their content to be promoted to pilots, but unable to self-validate their work or the work of others.
Innovator	Innovators may be more analytic-savvy business users, business analysts or citizen data scientists. They are "competent" from a data literacy perspective.	Can leverage any curated, approved or "sandbox" data source to build prototypes. May also bring in domain-specific data, flat files or third-party data to enhance analyses. Able to nominate their content to be promoted to pilots, but unable to validate the work of others.
Expert	Experts could be power users, experienced analysts, data engineers or data scientists. They are "fluent" or "multilingual" from a data literacy perspective.	Similar to the innovator, with the added ability to bring new data into the data sandbox. Able to promote their content to pilots, self-validate and validate the work of others based on agreed-upon life cycle practices.

Recommended by the Author

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

Reference Architecture to Enable Self-Service Analytics

Reference Architecture Implementation Guide

Solution Path for Building Modern Analytics and BI Architectures

Building an Analytics and AI Architecture Using Amazon Web Services

Building an Analytics and AI Architecture Using Microsoft Azure

Table 1: Connecting Technical Capabilities With Guiding Principles

	Effective Storytelling	Analytical Modeling	Governance and Security	Accessibility and Interoperability
Analytics Catalog		Y	Y	Y
Automated Insights	Y			
Collaboration		Y	Y	Y
Data Preparation		Y		
Data Science Integration		Y		
Data Source Connectivity		Y		
Data Storytelling	Y			
Data Visualization	Y			
Governance			Y	
Metrics Store		Y		
Natural Language Query	Y			Y
Reporting	Y			

Source: Gartner (July 2023)

Table 2: Federated Analytics Versus Decentralized Analytics

	Federation	Decentralization
In Practice	Systems and practices allow multiple domains to work as a whole focusing on achieving the business goals while retaining some autonomy regarding local decision making.	Systems and practices dictate that business domains work autonomously regarding local decision making.
Governance	There is a joint effort and agreement of governing principles from a central authority and business domains.	Control rests with business domains.
Decision-Making Authority	Central authority may make decisions affecting the organization as a whole. Business domains retain some autonomy for local decision making.	Decision making is distributed without a central authority.
Interoperability	Business domains agree to standards for development and sharing analytics.	Business domains develop analytics independently and define their own standards.
Coordination and Collaboration	A central governance group facilitates collaboration and coordination across business domains.	There is no central group for coordination. Collaboration is independent across domains.

Source: Gartner (July 2023)

Table 3: Analytic Persona Model

Analytic Personas	Description	Capabilities Within the Self-Service Environment
Consumer	Consumers may be frontline workers, executives or external customers who view analytic content periodically. They are “conversational” from a data literacy perspective.	Consume and interact with production content. No formal content creation rights, and typically do not request them. Unable to validate or nominate any new content to be promoted to pilots.
Explorer	Explorers may be business users who are looking to do more diagnostic-type analytics, or junior analysts who are looking to expand their understanding of specific domains. They are “literate” from a data literacy perspective.	Able to duplicate and then modify production content. Able to nominate their content to be promoted to pilots, but unable to self-validate their work or the work of others.
Innovator	Innovators may be more analytic-savvy business users, business analysts or citizen data scientists. They are “competent” from a data literacy perspective.	Can leverage any curated, approved or “sandbox” data source to build prototypes. May also bring in domain-specific data, flat files or third-party data to enhance analyses. Able to nominate their content to be promoted to pilots, but unable to validate the work of others.
Expert	Experts could be power users, experienced analysts, data engineers or data scientists. They	Similar to the innovator, with the added ability to bring new data into the data sandbox. Able to promote their content to pilots, self-validate and

	are “fluent” or “multilingual” from a data literacy perspective.	validate the work of others based on agreed-upon life cycle practices.

Source: Gartner (July 2023)