Reference Architecture to Enable Self-Service Analytics

Published 4 April 2022 - ID G00765683 - 45 min read

By Analyst(s): Christopher Long, Shaurya Rana

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

Data and analytics technical professionals struggle to provide self-service capabilities that satisfy the diverse requirements of a growing analytics user base. This document provides a reference architecture to implement a modular, scalable and multitool A&BI environment.

Overview

Key Findings

- Self-service analytics programs require a balance between control and agility. Ungoverned development leads to duplication of effort, increased costs and analytical silos.
- To fill gaps in current analytic capabilities, many organizations find themselves implementing multiple platforms. Hence, a robust D&A architecture is needed to support multiple A&BI tools.
- The line between business analysts and citizen data scientists is blurred by the advanced and augmented capabilities provided by modern A&BI tools.

Recommendations

Data and analytics technical professionals responsible for implementing and managing a D&A architecture to support self-service users and multiple A&BI tools should:

- Use a federated implementation model, built on a shared data management foundation of a logical data warehouse, data governance and a semantic layer to deliver, manage and scale A&BI initiatives. Federation will grant autonomy to business users while protecting the organization from data risks.

- Institutionalize different technical capabilities and governance processes targeting two broad use cohorts: innovators and experts, along with explorers and consumers to support self-service analytics.
- Centralize the core business logic within a semantic layer to help maintain portability across multiple A&BI tools and provide uniformity to the information being consumed by applications, reports and dashboards.

Problem Statement

Data is the driver of today's competitive business environment. Data helps business leaders make fact-based decisions to plan for the future. Operating a company without analytics and business intelligence (A&BI) derived from data is like flying a plane without instrumentation. Given the importance of data in the modern enterprise, it is no wonder organizations are clamoring for new tools, reports and dashboards. However, the race for greater access to drive data-driven decisions has created a challenge for technical professionals to effectively store, manage, transform and deliver data to end users via tools of their choice. Many organizations begin with rolling out a single enterprise A&BI tool. But as new use cases exceed the capabilities of existing tools, companies implement new tools to support new capabilities. Many organizations today have embarked on self-service analytics programs (Gartner notes a multifold increase in related client inquiries), leveraging the business-user-friendly capabilities of modern A&BI platforms to democratize business intelligence throughout the enterprise. Indeed, self-service analytics programs are often key features of broader digital transformation and cloud initiatives. Based on client inquiries, Gartner finds that a growing number of organizations have implemented multiple A&BI tools to address the ever-demanding reporting, analytics and business intelligence needs of their users.

Traditionally, IT has played an important role in delivering analytics projects by providing tools and technologies that power business intelligence. Business users and stakeholders bring a set of requirements for their reporting needs to IT. The IT teams then look at existing tools to deliver the analytics solutions and, in the absence of certain capabilities, buy and deploy infrastructure to support a new A&BI tool. Analytics projects are generally driven by use cases coming from multiple lines of business (LOBs) and require the implementation of personalized end-to-end solutions. Many A&BI tools today have SaaS offerings, making it easier for businesses to implement solutions independent of IT and providing agility and faster go-to-market capabilities.

This ease of implementation can also result in multiple challenges to the business:

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

As organizations look for ways to simplify, scale and manage their multitool A&BI environments, they find two general approaches to tackle the issue:

1. Consolidate the number of A&BI tools.

2. Manage the multitool environment better.

Although maintaining an environment to support multiple A&BI tools comes with challenges, simply consolidating down the number of A&BI tools available in the organization can result in users that are underserved and/or underskilled with the resulting A&BI tools available. Effective methods for managing multiple platforms must also be implemented. Consolidating tools and effective platform management should not be viewed as separate resolution pathways. A sustainable solution will find both of these approaches working in concert to bring value to the business.

The primary goal of this document is to provide enterprise and solution architects with a guidance framework to design and build a reference architecture to support a multitool A&BI environment while addressing the following questions:

- How can we build, manage and scale an architecture composed of multiple A&BI tools?
- How do we integrate multiple A&BI tools to a common data platform?
- How can we provide features and capabilities consistently to all users within a multitool A&BI environment to meet the unique requirements of LOBs?

The target audience for this document includes enterprise architects, data and A&BI solution architects, and analysts. This research note can also be leveraged by chief data officers, chief analytics officers and project managers to better understand the feasibility and implementation of a multitool A&BI environment.

This document focuses on:

- Building the prerequisites to support a multitool A&BI environment
- Designing a federated reference architecture that supports the integration of multiple A&BI tools in order to derive maximum business value while reducing operational overhead and maintaining a scalable architecture
- Identifying A&BI tool capabilities desired by the various user groups within the organization

The Gartner Approach

We have observed that selecting the one ideal A&BI tool does not help address the problem of catering to the diverse requirements of multiple users within various lines of business of an enterprise. Implementing multiple A&BI tools in a decentralized model does not provide the scalability and consistency needed by organizations. And in many cases, organizations struggle with managing multiple A&BI tools due to lack of an implementation model, an integrated cohesive architecture and a well-defined analytics development workflow.

This guidance framework provides a systematic approach to evaluate users based on their expertise, align capabilities they look for within the various A&BI tools, decide on the implementation model, determine the prerequisites and build the integrated multitool A&BI solution. The document also provides a walk-through of an iterative analytics and A&BI project development cycle, providing autonomy to the users within a governed environment.

Gartner recommends that you follow the sequence laid out within the framework to identify capability gaps, build the necessary prerequisites and integrate the right set of A&BI tools to build a self-service analytics platform.

The Guidance Framework

The implementation of the federated A&BI architecture follows Gartner's familiar guidance framework to plan, build and run the platform (see Figure 1).

Download All Figures From This Material

Figure 1: Guidance Framework for a Federated A&BI Implementation Model

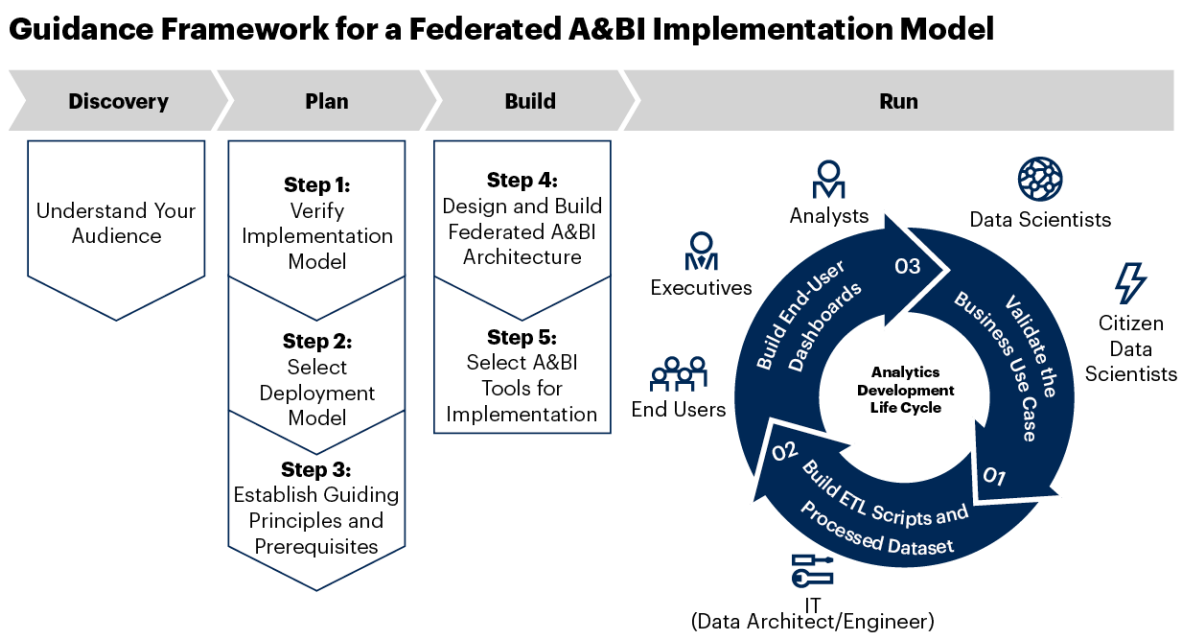


Figure 1 divides the process for implementing a federated A&BI architecture into four primary phases:

- **Discovery:** Evaluating the current environment
- **Plan:** Scoping the new architecture
- **Build:** Constructing the architecture and selecting tools
- **Run:** Operationalizing the new architecture

The discovery, plan and build phases represent projects taking a defined (but not insignificant) amount of time and effort. The run phase represents the cyclical nature of an ongoing, continual improvement process.

Discovery: Evaluating Your Current Users

An effective A&BI architecture is required to both meet the needs of business users and to scale efficiently to match growing demands of the business. The best place to start before defining a desired reference architecture is to identify the audience and its expectations of an A&BI solution.

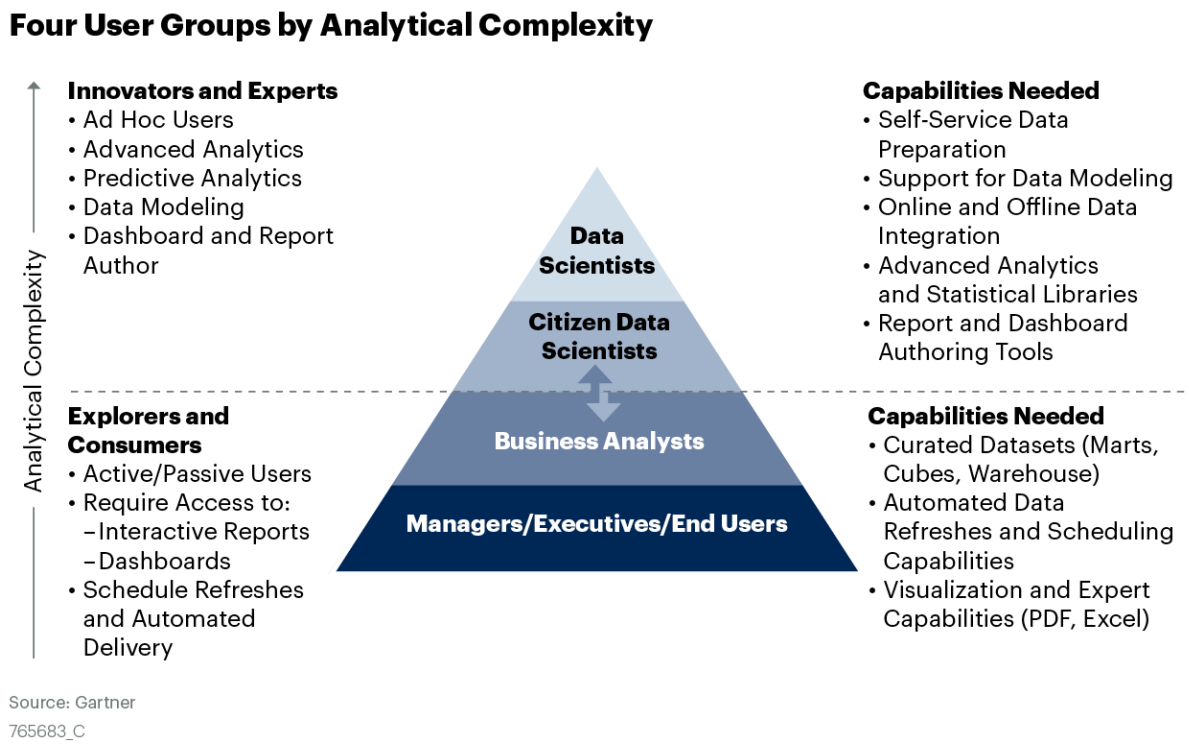
This discovery phase is a prerequisite designed to identify your organization's current needs and capabilities. Results gathered during this phase will shape the implementation of your A&BI architecture as outlined in this research.

Discovery

Understand Your Audience

Figure 2 describes the various business user groups identified by Gartner, organized by analytical complexity and skills.

Figure 2: Four User Groups by Analytical Complexity



Source: Gartner
765683_C

Gartner

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

Note the two-way arrows in Figure 2 between the business analyst and citizen data scientist user groups. These arrows suggest the blurred boundaries and range of skill sets across these user groups. The point is that there can be quite a vast spectrum of skills and analytical complexity between an advanced citizen data scientist and a novice business analyst. Further, users are likely to move between being innovators/experts and explorers/consumers in the course of analytical development. 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 1 for a summary of these four main user personas and their capabilities in a self-service data and analytics environment. Also refer to Essential Skills for Citizen Data Scientists for an examination of the unique skills scope of a citizen data scientist.

Table 1: 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 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

With this, the "discovery" phase is complete. We can proceed with the following steps laid out within the guidance framework to plan the A&BI architecture.

Plan

Step 1: Verify the Implementation Model

The three primary models for implementation that describe how A&BI tools connect to data sources are centralized, decentralized and federated. Each approach has its own strengths and weaknesses and appropriate use cases.

The Centralized and Decentralized Implementation Models

Figure 3 depicts a centralized A&BI implementation model.

Figure 3: Centralized A&BI Implementation Model

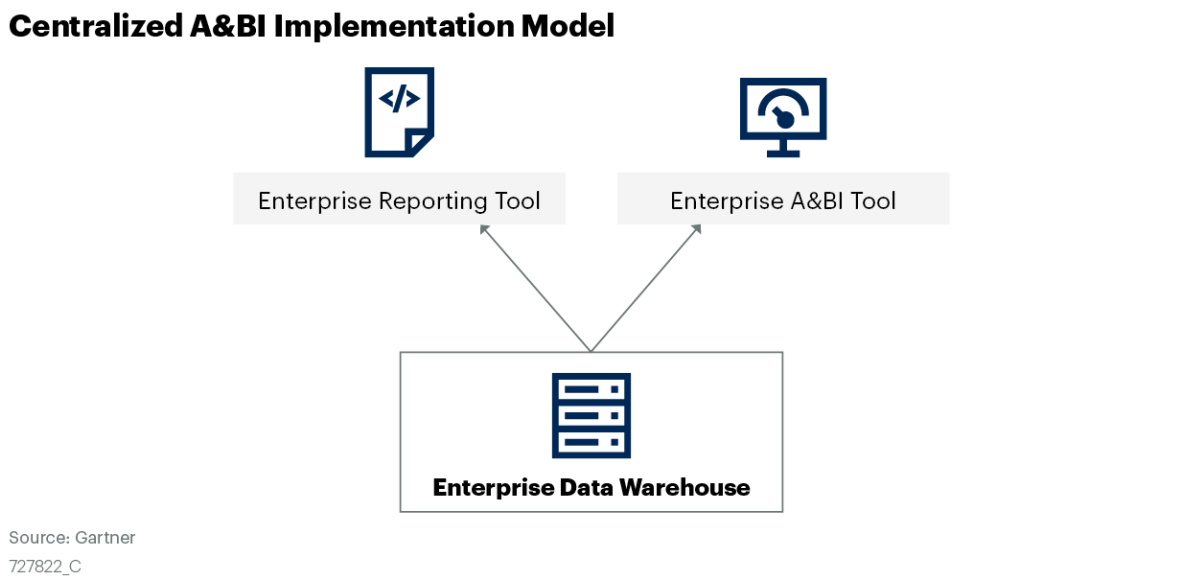


Figure 4 depicts a decentralized implementation model.

Figure 4: Decentralized A&BI Implementation Model

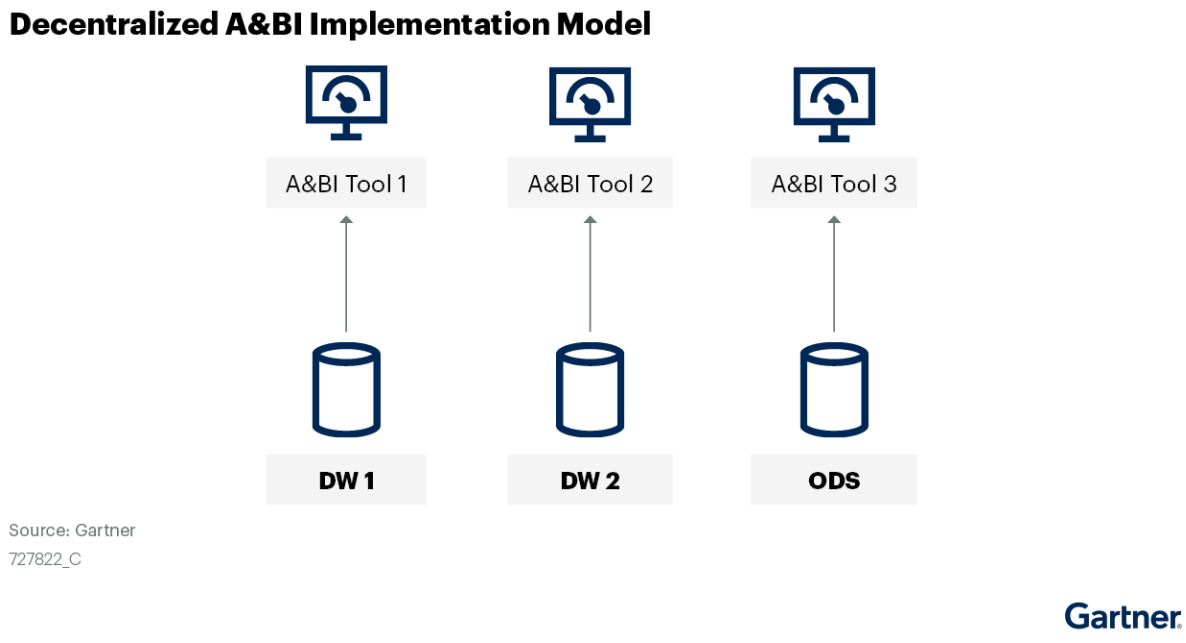


Table 2 summarizes the strengths and weaknesses of the centralized and decentralized A&BI implementation models.

Table 2: Strengths and Weaknesses of Centralized and Decentralized A&BI Implementation Models

(Enlarged table in Appendix)

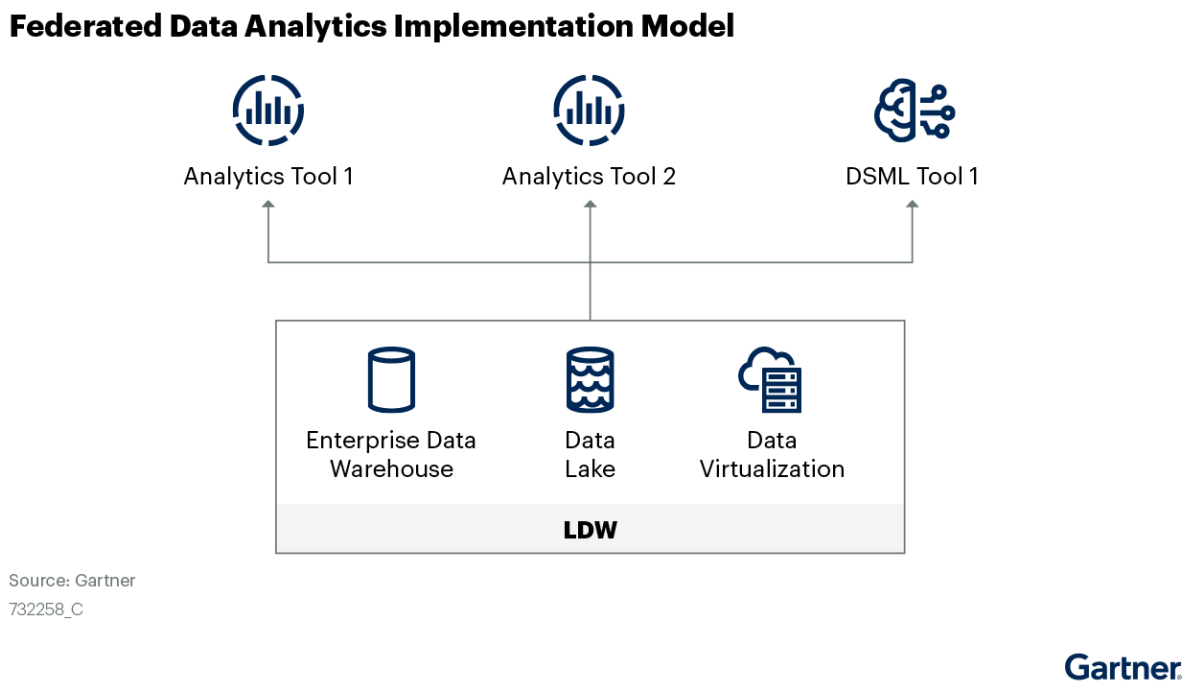
Implementation Model	Strengths	Weaknesses
Centralized A&BI: Information from across multiple source systems is brought together in a single location and then provides users access to the data via multiple tools and applications, including A&BI tools. This centralization approach helps improve efficiency in managing the data, improves the quality and helps govern it, since it is all in one place.	■ Eliminates redundant costs ■ Allows consistent data definitions ■ Presents an enterprise view of the data ■ Uses standardized, scalable architecture ■ Efficient management of data ■ Reinforces data quality and governance	■ Introduces bottlenecks ■ Uses a single data model ■ Restricts exploratory analysis ■ Creates significant backlogs
Decentralized A&BI: Each LOB provides its own set of A&BI requirements, and IT helps design and implement the solution, selecting the business users' tool of choice. The same data coming from the source systems gets processed multiple times, with its own set of business rules and logic, generating conflicting outputs for users.	■ Allows agile delivery ■ Increases flexibility ■ Enables customized views ■ Leverages domain-specific knowledge	■ Generates siloed A&BI solutions ■ Allows inconsistent A&BI architecture ■ Produces conflicting data definitions ■ Increases operational and licensing cost overhead ■ Fosters disparate set of capabilities ■ Allows inconsistency of report development ■ Creates redundant processes and potentially conflicting business logic

Source: Gartner (March 2022)

The Federated A&BI Implementation Model

The federated model provides the best of both worlds by addressing the weaknesses presented within the centralized and decentralized models, thereby providing the most value to both business and IT when designing an A&BI solution (see Figure 5). Because organizations wishing to enable self-service A&BI will end up needing to support multiple tools, Gartner recommends the federated A&BI implementation model to support multiple tools and serve the varied demands of different user groups in the organization.

Figure 5: Federated A&BI Implementation Model



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.

Within a federated implementation model, the corporate A&BI team would open up its data warehousing environment (i.e., LDW) to all of the divisions within the enterprise. Each division or line of business would have its own dedicated partition in the LDW or data lake to develop its personalized data marts, sets of reports and dashboards. IT teams provide self-service data preparation tools and train business users on how to blend local data with corporate data inside their EDW or data lake partitions. For groups that have limited or no A&BI expertise, the corporate A&BI team may continue to build custom data marts, reports and dashboards as before.

The federated model therefore provides a perfect blend of both a top-down and bottom-up approach:

- Standardized scalable architecture

- Support for an extensible physical and logical data model
- Centralized data repository, with dedicated partitions and zones for individual LOBs
- Consistent data definitions
- Agility to deliver multiple end-user analytics products and services
- Stronger partnership between IT and business
- Most important, the ability to truly support a self-service governed analytics platform

This model comes with its own unique set of challenges, but they are more process-dependent rather than being limited by the design of the architecture. This requires building a stronger communication line between the various business groups, users and IT.

A federated implementation model promotes self-service analytics within a secure, governed environment.

The success of the federated implementation model requires all constituents to adhere to some basic operating standards when using the enterprise solution. In order to succeed, enterprises need to establish a clear roadmap and a robust business-led governance program. The organization should also establish an A&BI center of excellence (COE) and develop a coordinated effort for IT to delegate control, where possible, back to business when it comes to data analytics.

See Essential Capabilities That Your Data and Analytics Center of Excellence May Be Lacking for more details on establishing an A&BI COE. For guidance on setting a clear data strategy and building a data and analytics governance program, refer to the following resources:

- Creating a Data Strategy
- Building a Comprehensive Governance Framework for Data and Analytics
- Data and Analytics Governance Approaches for the Technical Professional

Step 2: Select the Deployment Model

Another key architectural consideration for a self-service analytics program is whether to deploy it on-premises; in the cloud; or in a hybrid, multicloud or intercloud configuration. Because so many modern A&BI platforms featuring self-service capabilities exploit the benefits of cloud technology, self-service analytics is often part of broader cloud initiatives. There are many variables in play that influence an organization's choice of deployment options, including:

- **Cost:** The costs to maintain on-premises capabilities, the cost to migrate data and applications to the cloud, data egress charges levied by cloud service providers, the cost of labor and change management, and so on.
- **Governance and regulation:** Some data management and analytics deployment options are limited by geography, industry, law or corporate governance policies.
- **Performance and SLA requirements:** Some organizations require performance, speed or service at a level many cloud providers cannot guarantee, and so on-premises options are paramount.
- **Incumbent tools and data storage:** An organization's existing investment in a vendor's products may mean an alignment to the vendor's preferred deployment option, and where an organization's data already resides may influence a cloud strategy.
- **Technical capabilities of certain A&BI tools:** Some A&BI tools have limited deployment options in the first place, or the A&BI tools connect better to cloud or on-premises data sources.
- **Corporate culture, preferences and user skills:** In addition to corporate governance policies, existing familiarity or comfort with a particular deployment pattern often comes into play. Existing skills among an organization's staff — or the availability of those skills in the job market — may also influence a deployment option.

No matter the reasons, it is important to thoroughly consider the suitability of different on-premises, cloud and hybrid deployment options for the organization when planning the architecture. Further still, some A&BI tools offer a variety of licensing options layered atop deployment choices, which must also be considered. For example, an A&BI tool may only be available as a managed cloud SaaS offering in one particular vendor's cloud environment, or it may allow the organization to choose between a variety of deployment patterns.

The Gartner Cloud Decisions platform offers a number of interactive tools for assessing your organization's readiness and fit for the cloud. And the following resources examine the decisions around whether or not to move data, tools and workloads to the cloud from different perspectives:

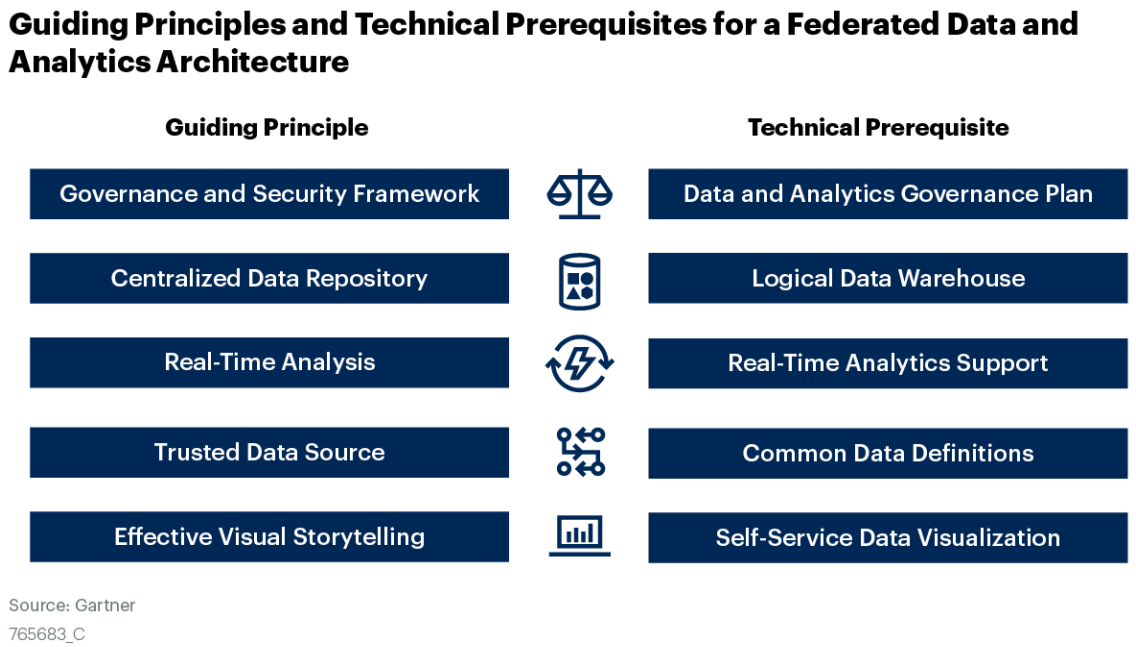
- Build the Right Justification for Moving to the Cloud
- Designing a Cloud Strategy Document
- The Impacts of Emerging Cloud Data Ecosystems: An Architectural Perspective
- Understanding Cloud Data Management Architectures: Hybrid Cloud, Multicloud and Intercloud
- Decision Point for Selecting Cloud Analytics Solution Architecture
- Solution Path for Modernizing Analytic Architectures

Step 3: Establish Guiding Principles and Technical Prerequisites

Before we start designing the federated A&BI architecture, we first need to establish a set of guiding principles and technical prerequisites to help us identify the set of core capabilities to be considered within the architecture.

There are five guiding principles identified in this framework to build an enterprise A&BI solution. The technical prerequisites function to operationalize each of the guiding principles (see Figure 6).

Figure 6: Guiding Principles and Technical Prerequisites for a Federated Data and Analytics Architecture



Guiding Principle 1: Governance and Security Framework

A modern self-service A&BI program is a sociotechnical system, meaning it is not merely a technical execution but one that involves people and processes throughout both IT and business units across the organization. Data and analytics governance is a four-way framework that helps ensure integrity, security, applicability and access to reliable data. 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. If the data is governed properly at the physical layer, it helps reduce the complexity of managing access to the users from within the A&BI tools.

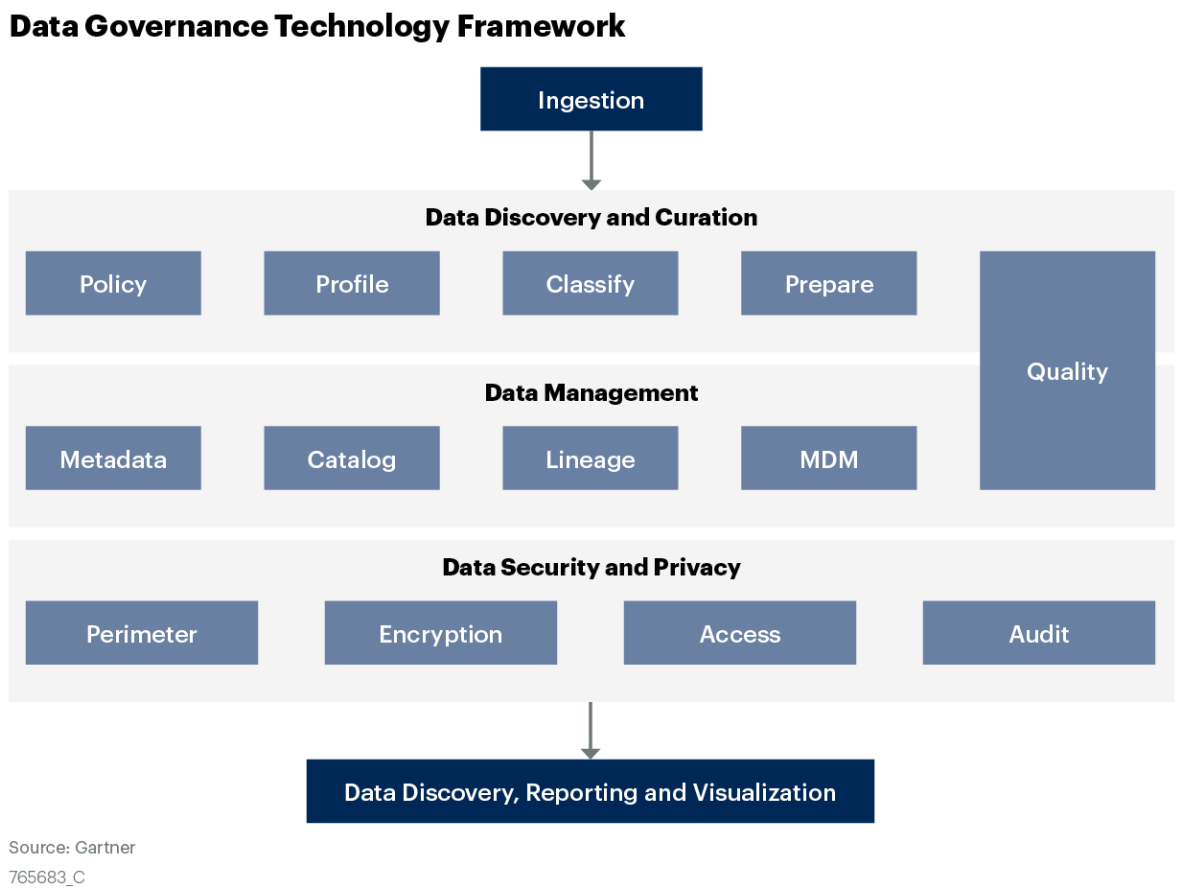
Technical Prerequisite 1: Data and Analytics Governance Plan

> 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.

Implementation of a successful data governance practice within an enterprise is 70% process and 30% technology. Typical data governance includes:

- Aligning and defining the business strategy and determining how to leverage data to solve the business problem.
- Policy and processes for managing data effectively. These are mapped to the roles and responsibilities of all the constituents (business and IT) within a data management life cycle.
- Definition of roles and responsibilities as part of the stewardship to help monitor and control the flow of data within the enterprise. Defining a clear set of roles and responsibilities helps implement a complex governance model involving various business and IT user groups.
- Tools and technologies selected to help operationalize a governance framework. Figure 7 displays how the governance technology stack fits together.

Figure 7: Data Governance Technology Framework



A robust data and analytics governance framework may be developed through tools that provide end-to-end data governance solutions or tools that support individual capabilities. Regardless of approach, conduct a comprehensive proof of concept based on your evaluation criteria before selecting the appropriate one. Note that the implementation of these tools may be part of an evolving D&A governance program, and investment is not a prerequisite to overall program development.

Vendors providing end-to-end data governance platforms include:

- Alation
- Cloudera
- Collibra
- Hitachi Vantara Lumada Data Catalog

- Precisely (Infogix)
- Informatica
- Qlik
- Talend
- Boomi (UniFi Software)
- Zaloni

Tools supporting specific capabilities are included in Table 3. These may be incorporated to manage targeted aspects of your governance framework.

Table 3: Governance Technologies

(Enlarged table in Appendix)

Tool Category	Description	Example Vendors
Metadata management	The definition of data governance implies effective management of data, which is impossible without managing its associated metadata. The selected tool should support manual and automated tagging of data to specific business domains, alongside an ability to search and steward the underlying data. See Market Guide for Active Metadata Management for further market analysis.	■ Adaptive ■ Alation ■ Alex Solutions ■ ASG Technologies ■ Collibra ■ Data Advantage Group ■ data.world ■ Quest (erwin) ■ Global IDs ■ IBM ■ Infogix ■ Informatica ■ Oracle ■ SAP ■ Semantic Web Co. ■ Smartlogic ■ Syniti
Data quality	Tools that support data quality are an absolute prerequisite to ensure successful implementation of the governance framework. The tool should provide profile, cleanse, merge and visualization capabilities and help create a workflow to maintain data in its pristine state. See Magic Quadrant for Data Quality Solutions for an assessment of this market and Data Quality Fundamentals for Data and Analytics Technical Professionals for guidance on data quality.	■ Ataccama ■ Datactics ■ Experian ■ IBM ■ Informatica ■ Innovative Systems ■ Melissa Data ■ MIOsoft ■ Precisely ■ Redpoint ■ SAP ■ SAS ■ Syniti ■ Talend
Data security	Data security tools can help locate sensitive data elements across multiple systems. They also help enforce regulatory, contractual and architectural compliance, support access management, and identify any security breaches (see Securing the Data and Advanced Analytics Pipeline).	■ Microsoft (BlueTalon) ■ PKWARE ■ Micro Focus ■ Protegrity ■ Thales (Vormetric)

Source: Gartner (March 2022)

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

- Creating a Data Strategy
- Building a Comprehensive Governance Framework for Data and Analytics
- Data and Analytics Governance Approaches for the Technical Professional

Guiding Principle 2: Centralized Data Repository

A centralized data repository unifies — physically or virtually — all of the organization's data in a single repository. This reduces data movement across multiple data stores and analytical systems, helps identify new relationships between data elements, and enables more effective governance and control over the data. It is a core architectural component of a federated A&BI solution, and the logical data warehouse (LDW) paradigm is ideal for realizing a centralized data repository.

Technical Prerequisite 2: Logical Data Warehouse (LDW)

The LDW is a growing data management architecture for analytics that combines the strengths of traditional repository warehouses with alternative data management and access strategy — specifically, federation and distributed processing. It establishes a centralized trusted data source for users to perform their analysis. The LDW will help store the active raw data, historical processed and raw datasets, semistructured and unstructured data, along with curated datasets in the form of cubes and data marts. The key here is to build a data management system that reduces the amount of data movement across multiple systems (see Adopting a Logical Data Warehouse).

The LDW, sometimes referred to as a "data hub," can be built using a set of RDBMSs, NoSQL data stores, data virtualization (see Assessing the Relevance of Data Virtualization in Modern Data Architectures) or distributed computing frameworks. Depending on the scale of the implementation, the LDW can be built using a single persistent data store (see Assessing the Optimal Data Stores for Modern Architectures).

Guiding Principle 3: Real-Time Analysis

Information about consumers and sales received in real time is more relevant than data from last month or even last week. Access to real-time data generated has become an important consideration to maintain an enterprise's competitive edge. Organizations should determine exactly what kind of real-time requirements they have by answering the question, "What is real time?" There are significant architectural differences between an organization needing data refreshed in near-real time (e.g., minutes, seconds, milliseconds, etc.) and an organization needing a constant stream of data as it is ingested.

Many organizations do not need the latter (yet), and thus, real-time streaming architectures are not critical. What is important for all organizations to know, however, is that real-time streaming analytics is possible.

The key questions to ask when planning for real-time analytics are:

- How quickly will the value of the decision degrade?
- How much better will a decision be if more time is spent?

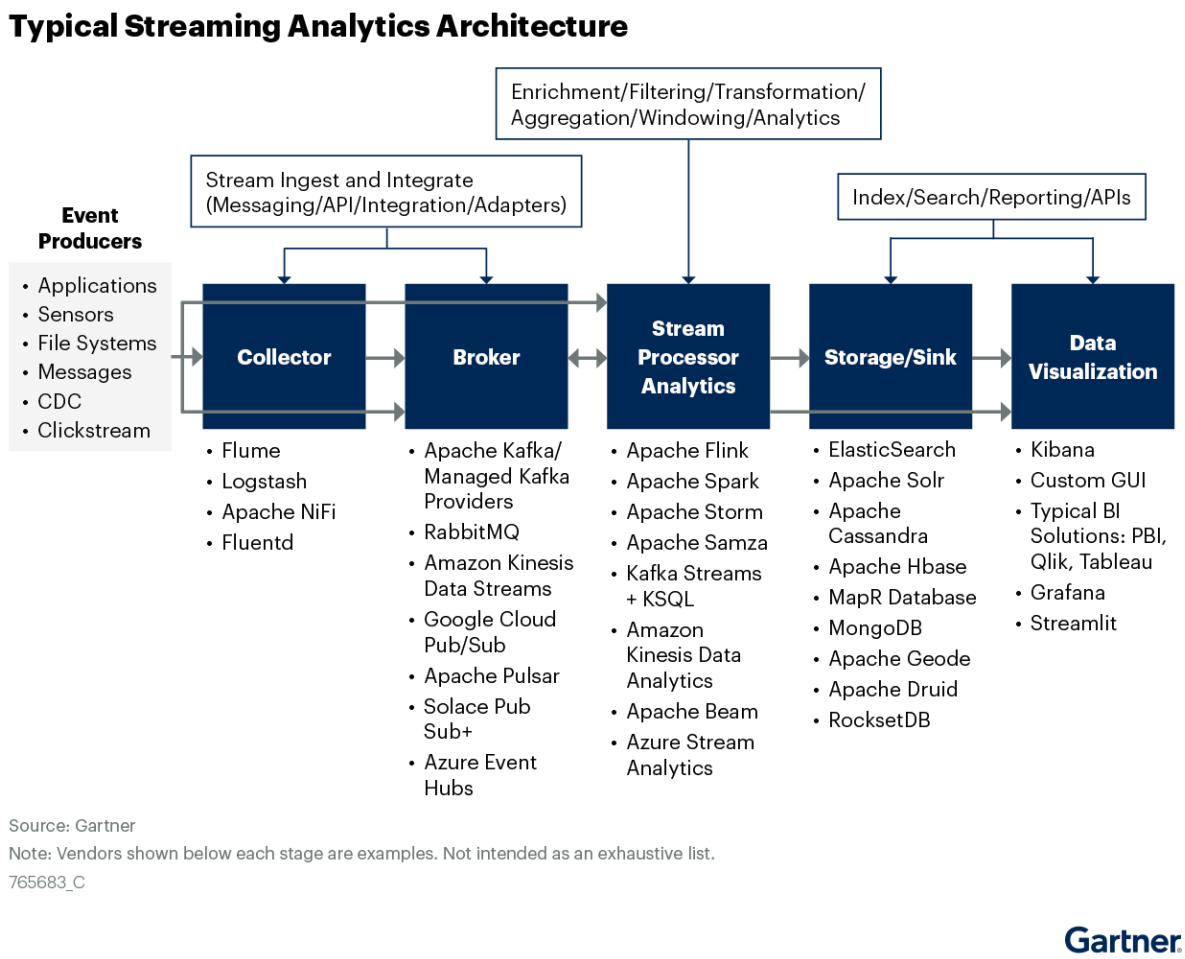
Typically these questions can help decipher which concept of “real time” is relevant to the business problem being addressed or the level of automated operational decision making being achieved.

Technical Prerequisite 3: Real-Time Analytics Support

If supporting real-time streaming analytics is a requirement for your implementation, you will need to incorporate components that can handle the stream of data coming in and even process and analyze that data in real time. Further, to achieve operational efficiencies and automation, you might need real-time downstream action to be triggered via the insights generated. Hence, the architectural considerations while designing real-time systems require them to preserve the “time value of data,” which causes a paradigm shift from how batch-based data systems are designed.

Stream processing is used to query continuous data streams and detect conditions, within a short time from receipt of the data. It can also be used for filtering, enriching and normalizing the data within the data packets. The stream broker provides support for integration to the source and ensures that data flows contiguously. With support for real-time analytics, users can receive alerts and notifications and facilitate real-time decision making with critical business-life operations. Figure 8 shows a typical streaming analytics architecture that facilitates data coming from streaming sources. To learn more about streaming ingestion and processing, refer to Stream Processing: The New Data Processing Paradigm and Streaming Analytics in the Cloud: A Comparative Analysis of Amazon, Microsoft and Google.

Figure 8: Typical Streaming Analytics Architecture



Source: Gartner
Note: Vendors shown below each stage are examples. Not intended as an exhaustive list.
765683_C

Gartner

Guiding Principle 4: Trusted and Accessible Data Source

The combination of the LDW and related tools for cataloging, metadata management and data quality that are governed and secured presents a trusted data source for users to analyze. It provides a unique set of capabilities to promote self-service analytics across the enterprise with less reliance on IT. Users now have one single location to fetch high-quality data supporting ad hoc analysis, standardized operational reporting, and even advanced analytics and machine learning in the future. When data can be trusted throughout the organization, users can begin to conduct more significant self-service data preparation activities.

Technical Prerequisite 4: Common Data Definitions

The primary goal of implementing a modern self-service analytics platform is to empower its business users to derive faster insights from the data. Through numerous surveys, Gartner has learned that the biggest bottleneck for any analytics project is the data preparation phase; it takes more than 65% of the overall project time to deliver data in a consumable format. The key ingredients of a self-service data preparation platform are to provide users the ability to combine, cleanse and transform relevant data for further analysis.

A couple tools that help create those centralized, governed sources for analytics developers include data virtualization, semantic layers and metric stores. Data virtualization layers serve to connect users to data rather than to collect the data. As an abstracted data layer, data virtualization provides a flexible, optimized source for self-service users. By mapping complex data into familiar business terms, semantic data layers provide a unified, consistent view of data in the organization (see Demystifying Semantic Layers for Self-Service Analytics). An emerging tool in providing consistent, reliable data to analytics users is the metric store. In these stores, metrics are built as code and stored as a reference point to aggregate data as queried by a user. By querying through the store, a calculation (for example, revenue) remains consistent regardless of the user or purpose. Both semantic layers and metric stores bring trust to the architecture by ensuring uniformity. To be effectively implemented, these tools should be implemented in the data layer. This enables the data to be capitalized on for a variety of purposes and consumed by multiple A&BI platforms without concerns for vendor lock.

Analytical developers and data engineers may look to the following vendors for these tools:

- Amazon
- AtScale
- Denodo
- Dremio
- Microsoft
- MicroStrategy
- SAP

Guiding Principle 5: Effective Visual Storytelling

Once you have provided the capabilities to process the data and extract value, it becomes extremely important for users to communicate insights back to the decision makers. The best way to do this is through visual representation in the form of graphs and charts. This is where the tools and architecture come into play. The architecture should support discovery, self-service data preparation and modeling, and it should provide an interactive way to visualize and deliver the data to end users.

Technical Prerequisite 5: Self-Service Visualization and Ongoing Literacy Training

Select tools that make the creation of compelling visualizations as easy as possible for users and those with self-service authoring and sharing capabilities that conform to your overall governance framework. Modern A&BI platforms like Microsoft Power BI, Tableau, Qlik Sense, ThoughtSpot and SAP Analytics Cloud and many others provide easy-to-use native functionality for creating visualizations and dashboards that bring data to life. Additionally, extended libraries are available for some tools that allow the innovators and experts to create custom visualizations.

Ongoing training in data literacy throughout the organization is critical for supporting widespread adoption of a self-service analytics initiative and upskilling business users into data steward roles and into more advanced user groups. Work with business users or an analytics COE to develop best practices and standards for the most effective ways to tell stories with data that drive business insights. See Communicate Insights Effectively With Augmented Data Visualization and Storytelling for guidance on this topic.

With the completion of Step 3, we have now come to the end of the plan phase of the guidance framework. Let's quickly summarize what we have discovered so far from the outcomes of the discovery and plan phases:

- Each organization has a set of four business user groups, classified within two categories: innovators and experts (data scientists and citizen data scientists) and explorers and consumers (business analysts and managers/executives/end users). Each group has its own set of capability requirements from an A&BI solution.
- IT teams will need to support a multitool A&BI environment to serve the needs of the various business user groups.
- A federated implementation model is an ideal approach to supporting self-service business analytics.
- Whether it be on-premises, in the cloud, or a hybrid or multicloud approach, the deployment model must be determined.

- The six guiding principles (along with related technical prerequisites) to be considered when designing the reference architecture are:
  - A governance and security framework
  - A centralized data repository
  - Real-time analysis
  - A shared metadata library
  - A trusted and accessible data source for users
  - Effective visual storytelling

Next begins the build stage of the guidance framework. This involves building the federated A&BI architecture in Step 4 and selecting the right A&BI tools in Step 5.

Build

Step 4: Design and Build the Federated A&BI Architecture

All of the analysis done until this point indicates that, to implement a scalable multitool A&BI architecture, we must align the capabilities of the individual tools to the requirements of the user groups identified within the pyramid in Figure 2. Hence, the A&BI solution architecture should be built on a capability model.

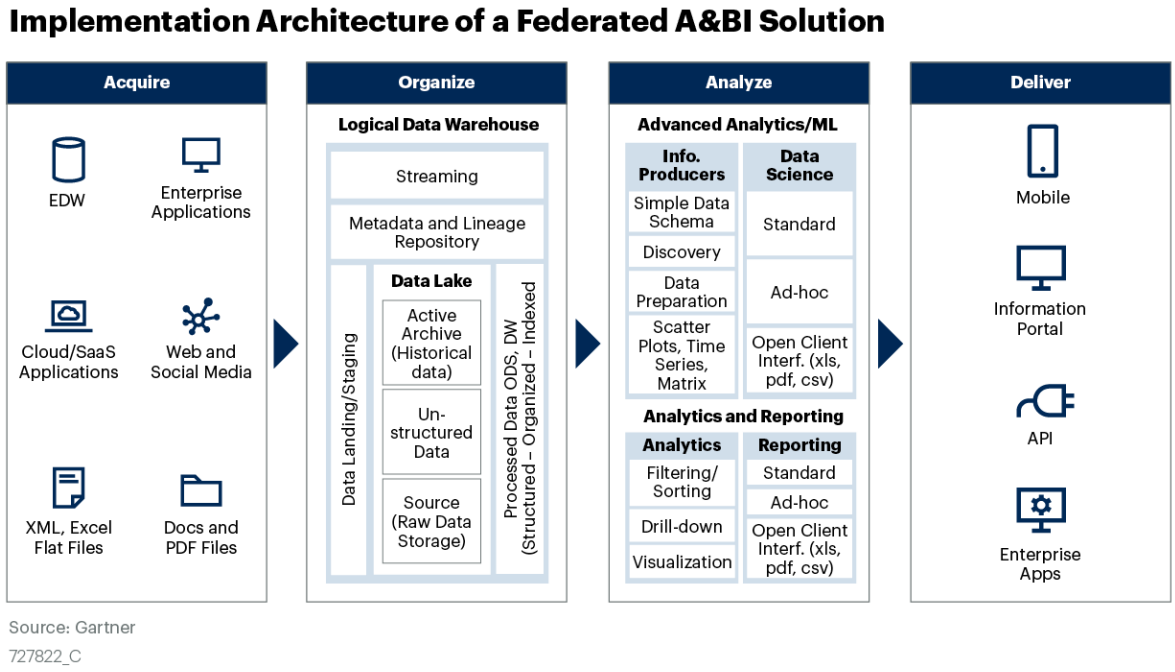
The capability model provides a comprehensive set of features that an enterprise requires via the integration of individual tools or a single platform in order to execute its business model. In designing an A&BI solution, think of it as a set of features and functionalities that users look for based on their level of expertise and technical skill sets within a given tool.

The architecture design should also factor in the federated implementation model and take into consideration the guiding principles identified as part of the initial analysis within the plan stage.

To build the reference architecture, we are making the assumption that the organization has selected two A&BI tools for implementation. Each tool individually caters to the user groups within the two major categories: producers and consumers. However, this architecture is extensible and may support more than two A&BI tools.

Figure 9 shows the federated A&BI architecture supporting two A&BI tools.

Figure 9: Implementation Architecture of a Federated A&BI Solution



A modern architecture consists of four main layers: acquire, organize, analyze and deliver. Data from all sources needs to be brought into the LDW (organize layer) for batch processing, while the streaming data can be supported via a stream broker and streaming analytics vendor products (see Figure 8). The LDW, or data hub, can be implemented using a traditional relational database management system (RDBMS), NoSQL databases or Hadoop, or a combination, depending on the type of data that needs to be stored (structured, semistructured or unstructured). This centralized hub should hold data in its most granular format, as well as the preprocessed data in the form of data marts and/or cubes.

The LDW will help reduce the amount of data movement between various systems for analysis by business user groups.

The LDW should be divided into five high-level zones:

- **Landing/staging:** This is where data from the source systems is pulled in and stored as is. This process can be implemented using the enterprise data integration tool. No transformation should be performed on the data during the process of moving it from source to sink, but you need to ensure all of the metadata is collected and stored in the centralized metadata repository to provide lineage.
- **Data lake:** Gartner defines a data lake as a collection of storage instances of various data assets in addition to the originating data sources. Data is stored in a near-exact format as its source. The data lake can be further partitioned to hold structured raw source data, semistructured and unstructured data, and historical archives. Most organizations have to abide by compliance regulations set by government or industry, and thus it is critical to secure their data. Hence, when moving data from its landing area to either of the partitions within the data lake, sensitive data elements should be encrypted, masked or tokenized.
- **Processed zone:** This zone holds the data marts for the individual LOBs, use cases and even golden copies of business domain reference data, like the client and product master. Data stored in this zone can be highly indexed to support faster reads via enterprise applications and multiple reporting and A&BI tools.
- **Metadata repository:** The metadata repository can act as a business glossary or catalog for users to access relevant data within the data lake, so it should be stored within the LDW. Metadata should be collected from the data's originating systems, and as it flows through the various layers, getting transformed along the way. This repository is a critical component within the architecture to help support self-service analytics due to the search, discovery and lineage capabilities it provides to the end users.

- **Streaming:** The streaming zone is used to surface real-time data for downstream applications. This could be in the form of dashboards for real-time visualization, automation processes or operational data manipulation and analytics. Based on the streaming architectures involved, the streaming zone could be coupled or decoupled with the static data repositories like the data warehouse and data lake. Operational data stores or other real-time data stores can also be leveraged in this zone. Further, this zone can be used to achieve operational efficiencies of the entire LDW in the form of real-time monitoring and ingestion of data.

Step 5: Select the Right A&BI Tools for Implementation

The final step prior to operationalizing the new federated architecture is putting in place the A&BI tools used by your user groups (innovators and experts, and explorers and consumers; see Figure 2). Most organizations already have a set of A&BI tools, licenses and infrastructure in place. There is no need to bring in a set of new tools as long as the prevailing A&BI tools provide all of the capabilities these business users are looking for.

A&BI Tool for Innovators and Experts

The A&BI tool for innovators and experts should provide features and capabilities that those stakeholders look for, such as data discovery, ad hoc data modeling and self-service data preparation. Additionally, these users may look for extensibility features that allow for advanced coding in languages such as Python, R and Java.

Individuals within these categories will typically connect with more granular data within the LDW and work to validate hypotheses and use cases for businesses and executives. Using their preferred A&BI tool of choice, they can provide valuable insights based on trends and patterns they see when combining data from various sources in the LDW. As part of their analysis, these individuals can help create the business and ETL logic for the resultant analytical models, reports and dashboards as they are operationalized.

The resultant data marts, cubes and models should be built in the LDW with the help of data engineers, rather than within the A&BI tool's semantic layer. This facilitates portability and reusability of the data by another A&BI tool or even an enterprise application in the long run.

A&BI Tool for Explorers and Consumers

Explorers and consumers may leverage the A&BI platform to access the processed, curated data developed by the innovators and expert groups, along with data engineers. These datasets may also be consumed by enterprise web and mobile applications alongside the explorer and consumer-focused A&BI tool. The level of drill-down and access to granular data can be set within this A&BI tool, providing complete lineage if required by the end users.

These users, typically business analysts, managers, executives and end users, will look for tools that provide features and capabilities to filter and pivot data and that provide visualization capabilities in the form of charts and graphs. Visualization and data storytelling will be a principal focus for these users to provide clarity to the message being delivered by the underlying data.

Can We Support More Than Two A&BI Tools?

It is possible to implement and support more than two A&BI tools using the same reference architecture. In fact, the ability to adapt to and support multiple A&BI tools as necessary is one goal of a federated A&BI reference architecture. Because the federated architecture separates your data layer from the A&BI tools, businesses are able to scale out and optimize analytics platforms to meet the evolving needs of stakeholders.

When it comes to selecting the right A&BI tools for implementation, we need to look at the key technical features for each of the tools. Table 4 describes key A&BI platform features that form the basis for the Solution Criteria for Analytics and Business Intelligence Platforms, which can be used to assess the technical capabilities of specific tools.

Table 4: Key Analytics and Business Intelligence Platform Features

(Enlarged table in Appendix)

Feature	Description
Analysis and Content Creation	
Advanced Analytics	Enables citizen data scientists and other power users to easily access statistical and analytical functions and machine learning models that are self-contained within the A&BI platform or are available for import and integration from externally developed models.
Automated Insights, NLG, NLQ	Enables business people and citizen data scientists to use machine-learning-based services to automatically find, visualize and narrate relevant findings such as correlations, exceptions, clusters, links and predictions without having to build models or write algorithms. Natural language query (NLQ) gives users the ability to query data using terms that are either typed into a search box or spoken. Natural language generation (NLG) entails the automatic creation of linguistically rich descriptions of insights found in data.
Data Visualization	The ability to create highly interactive dashboards and content, with visual exploration and embedded advanced and geospatial analytics, to be consumed by others.
Interactive Visual Exploration	Enables the exploration of data via manipulation of chart images, with the color, brightness, size, shape and motion of visual objects representing aspects of the dataset being analyzed.
Mobile Exploration and Authoring	Enables development and delivery of content to mobile devices in a publishing and/or interactive mode, and takes advantage of the native capabilities of mobile devices, such as touchscreens, cameras, location awareness and natural language query.
Data Management	
Catalog and Governance	Availability of functionality that enables the sharing of semantic model and metadata objects, including the ability to automatically generate and curate a searchable catalog of the artifacts created and used by the platform and their dependencies.
Data Preparation	Drag-and-drop, user-driven data combination of different sources and the creation of analytics models (such as user-defined measures, sets, groups and hierarchies).
Model Complexity and ETL	Enables support for complex data models, including the ability to handle multiple fact tables, interoperate with other analytic platforms and support knowledge graph deployments. Extract, transform, load (ETL) capabilities enable the access, integration, transformation and loading of data. ETL features include the ability to index data and manage data loads and refresh scheduling.
Infrastructure	
Cloud	The ability to support building, deploying and managing analytics and analytic applications in the cloud, based on data both in the cloud and on-premises, as well as across multicloud deployments.
Data Source Connectivity	Capabilities that allow users to connect to and ingest structured and unstructured data contained in various types of storage platforms, both on-premises and in the cloud.
Manageability	Enables a variety of A&BI platform administration and infrastructure functions, including capabilities for scaling the A&BI platform to optimize performance and ensure high availability and disaster recovery.
Security and User Admin	Enables platform security, administering users and auditing of platform access and utilization.
Share Findings	
Collaboration	Enables users to share and discuss information, analyses, analytic content and decisions via discussion threads, chat and annotation.
Embedded Analytics	Capabilities include providing a software developer kit (SDK) with APIs and support for open standards for creating and modifying analytic content, visualizations and applications. The capabilities also include analytic content embedding into a business process, application or portal.
Reporting	Capabilities allow users to publish and deploy analytic content delivery through various output types and distribution methods, with support for content search, storytelling, scheduling and alerts.

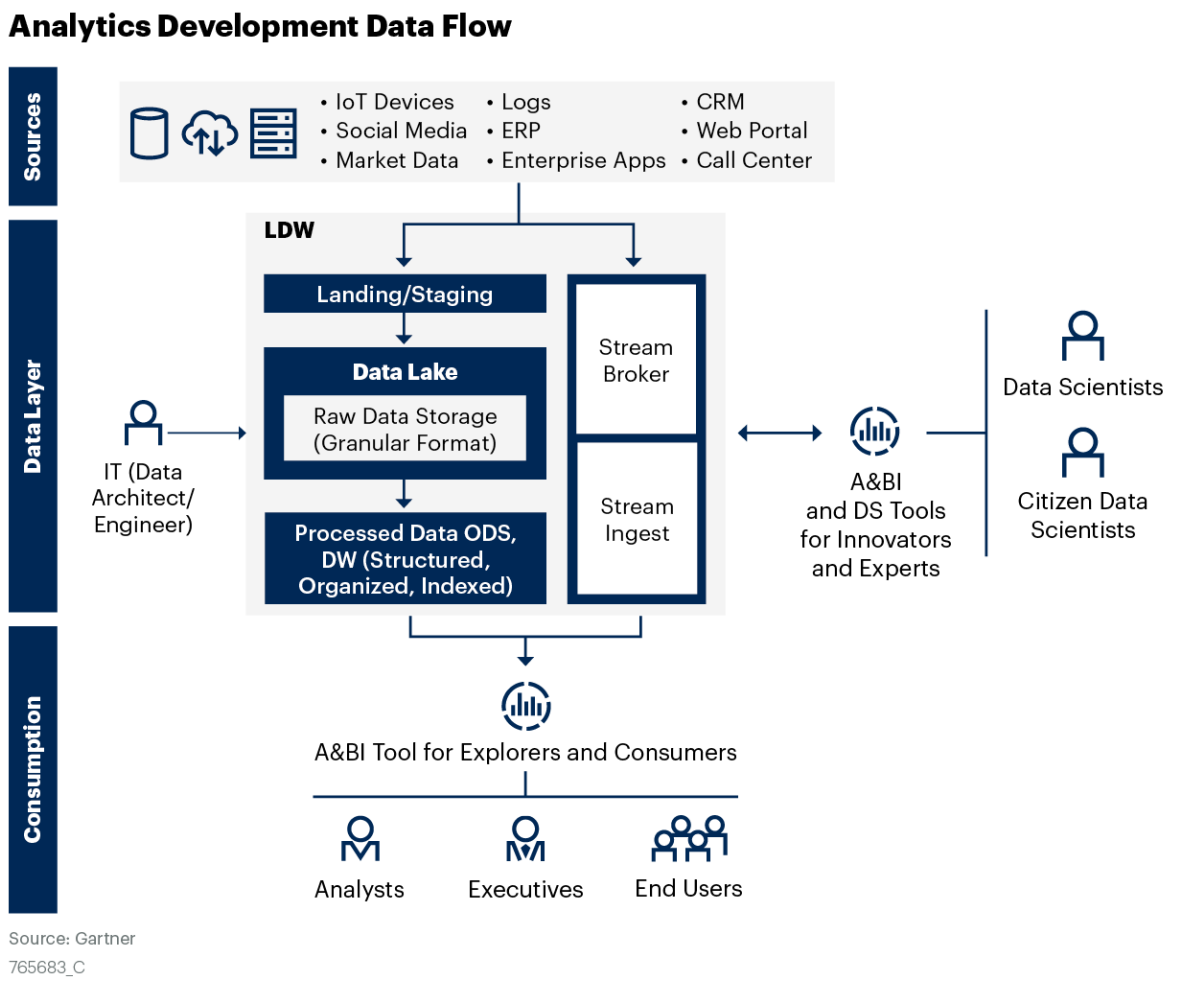
Source: Gartner (March 2022)

In the end, the selection of A&BI tools, and how many different platforms are placed in service, should be based on the requirements and criteria identified for your users' groups during the "discovery" phase of the guidance framework.

Analytics Development Life Cycle (Run)

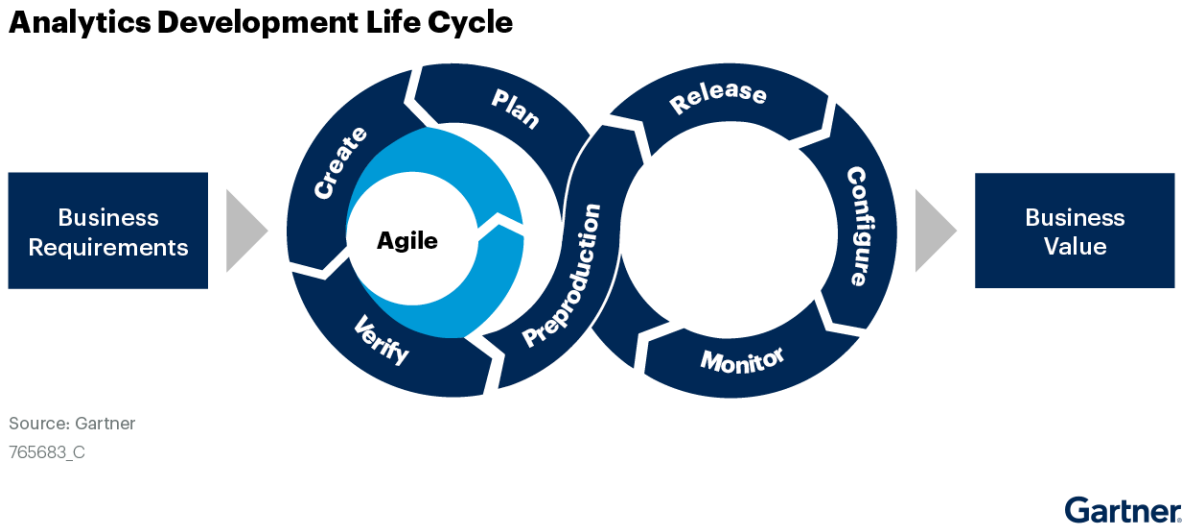
Having designed and built a federated A&BI architecture to support multiple tools, let's look at the recommended workflow for executing an analytics project using this newly federated A&BI architecture. Figure 10 depicts the flow of information through the various layers within an analytics development life cycle and its interaction points with various users.

Figure 10: Analytics Development Data Flow



When operationalizing analytics within an organization at scale, a DataOps practice creates predictable delivery and change management for analytical artifacts (see Data and Analytics Essentials: DataOps). Figure 11 depicts the life cycle flow for a DataOps practice:

Figure 11: Analytics Development Life Cycle



Analytics projects will be generally instantiated by the business. As these business requirements are defined, the DataOps team (a collaborative group of business analysts, data scientists, data stewards and data engineers) work to develop and build the desired analytical artifacts to be operationalized. The DataOps practice is designed for continual monitoring, feedback and improvement as needed in today's organizations.

The success of the newly designed, federated analytics architecture and extended DataOps programs will ensure:

- Alignment of tool capabilities to user group requirements, leading to faster time to insights.
- Data portability across A&BI tools by pushing logic to the data layer, preventing analytical silos.
- Governed, secured self-service analytics platform.

Follow-Up

After the implementation of the federated A&BI solution, it is important to assess whether the underlying objectives for building a scalable analytics platform were achieved. The following checklist can facilitate the follow-up process after the implementation:

- Introduction of new capabilities for both business and IT to scale a multitool A&BI platform.

- Transitioning of A&BI and analytical practice to business from IT. IT takes on a supporting role to manage the infrastructure of the tools and data management systems alongside providing access to data.
- Clear definition of roles and responsibilities of data engineers, architects (building data pipelines), data stewards (governance) and power users (data scientists and citizen data scientists).
- Establishment of a self-service analytics platform, catering to all user groups within the enterprise.
- Promotion of a franchise model for execution of future analytical programs by providing the business with tools and technologies that give them a personalized, specialized and simplified solution.
- Reduction of technical debt by reconciling processes and tools, which are now much more aligned to the capabilities and use cases within the enterprise architecture.

Organizations benefit by the introduction of new and innovative tools and features being rolled out by vendors that are under constant pressure to deliver more value. This results in a market that has a plethora of tools and technologies and that is constantly evolving. Therefore, it is important to document all of the findings throughout the implementation process and outcomes of proofs of concept (POCs) conducted while establishing or rearchitecting existing or new A&BI tools. The key findings gathered along the planning, build and execution stages should be revisited to look for opportunities to further optimize the architecture.

Risks and Pitfalls

When starting with a clean slate or introducing a second A&BI tool, enterprises have limited risks. That is not the case when they already have an implementation of multiple A&BI tools that would now need to be rearchitected. Following are potential risks, along with mitigation strategies.

Risk: Lack of clarity on the business use case and business buy-in. A lack of both a defined business use case and a clear definition of the problem that users are facing with the current implementation could lead to misalignment of the resulting solution with larger business priorities. Also, the lack of clear buy-in and support from other stakeholders in the business when designing the solution can result in creating an analytics platform that does not meet business requirements.

How to mitigate: Identify the enterprise's roadmap and capabilities that will be required to achieve both short-term and long-term business goals. Data, which is a key component driving the modern-day digital revolution, needs to be acquired, organized, analyzed and delivered to provide value. The business should articulate a minimum of four or five strategic use cases to help design the future-state architecture effectively. Use an agile implementation model and have someone from the business group be the product owner. This helps create transparency throughout the design and implementation stages. Business users should be actively involved in the tool selection process, mitigating any risks of selecting the wrong tool and thereby inhibiting users from effectively working with their data.

Risk: "Lift-and-shift" approach from legacy A&BI environment. When moving away from an A&BI tool that has been used for a significant amount of time, enterprises end up with a large number of reports and dashboards that need to be transitioned to the new platform. This can be cumbersome and lead to unnecessary effort and resources to migrate disused reports and dashboards.

How to mitigate: As part of implementing new A&BI tools, organizations should evaluate the current analytical artifacts (reports and dashboards) that should be migrated. Over time, organizations build reports that become disused and may no longer be relevant to the business. Technical professionals can support a review process by using current A&BI tools to inventory artifacts currently in production. Where available, also provide recent activity data to aid the business in making decisions on what elements to migrate into the new environment. Often, migration into a new environment will entail redevelopment efforts on the new platform. Therefore, it should be an active business decision on which elements to build in the new environment. This process also has the added benefit of identifying areas to improve legacy reports and dashboards that may no longer fully represent current business requirements.

Risk: "All-or-Nothing" launch approach to new tools. Some organizations and users may claim that new platforms are not useful if not all users have adopted the new tools or if all the needed reports and dashboards are not available on launch day. (Note: This risk may be tightly coupled with the "lift-and-shift" risk in practice.)

How to mitigate: The migration of existing reports and dashboards should be done in phases, where users can start leveraging capabilities of the new A&BI tool and then slowly transition from the older one. This approach helps prevent any operational impact on critical business processes while providing a parallel test environment. Additionally, not all users or teams will jump into the self-service environment straight away. A phased approach to migration supports adoption over time. A phased approach also allows IT professionals room to collaborate with business users to evaluate successes and challenges during an implementation and course-correct where needed.

Risk: A lack of training, patience with implementation and adherence to the workflow. Introduction of a new process (e.g., governance, data prep and self-service) can at times be overwhelming for end users. This will hinder the success and adoption of the newly designed A&BI platform. It is also tempting to hurry to market, and the business loves it when IT can deliver faster tactical solutions. However, you run the risk of not delivering on the enterprise's long-term vision.

How to mitigate: Communication and training play key roles in mitigating this risk. Build strategic partnerships within the business, and communicate clearly the roadmap and changes to the existing process. Educating and training users on new processes and tools will ensure successful onboarding and adoption of the new platform (see How to Use Influencers Within Communities to Increase Data and Analytics Adoption). This can be done by partnering with vendors and conducting workshops to train the users. After considering strategic business use cases to design the federated A&BI architecture, it is imperative to conduct a POC of the newly designed architecture before proceeding with the implementation. For example, you should validate the data flow, required analytical capabilities and user interaction points. Organizations conduct POCs when selecting products, but fail to do the same with business requirements. Focus the POC on the user requirements and capabilities, rather than just focusing on the underlying tools. Strictly enforce policies and processes to adhere to the recommended workflow. This can be achieved by a central A&BI or data governance council. Conduct periodic reviews of analytics project implementations to prevent any deviations from the recommended approach, but constantly be on the lookout for new approaches that decentralized groups may use that are worth implementing organizationwide.

Risk: Creation of analytic silos. As new A&BI platforms are introduced to the business and an increased number of dispersed and disparate users perform their own data modeling, there is an increased risk of creating analytic data silos. These silos are represented by the data pipelines, datasets and dashboards designed for analytic and BI purposes, either locked within the specific platform or within the business function having created the solution (or both). The challenges created by these silos include: inconsistent metric calculations, redundant efforts to build similar datasets and related increase in maintenance efforts. In addition to these challenges, the creation of A&BI silos prevent collaboration, portability of data and sharing of insights, which leads to the inconsistent interpretation of the data.

How to mitigate: Reducing the chance for and development of analytical silos may take multiple forms, including the use of technical solutions, policy and education. Through policy and education, technical professionals may focus on two primary areas: governance and data literacy. As part of the data and analytics governance team (see Data and Analytics Governance Approaches for the Technical Professional), technical professionals assist with creating and operationalizing governance policies. By working closely with analytics stewards, technical professionals guide in enforcing development and usage practices throughout the organization. Some analytics developer users may see governance as a bureaucratic hurdle that impedes the development process and time to deliver value to the business. However, by coaching users through governance and data literacy programs, technical professionals can also provide the foundation to these users of how and why common data language and collaboration across the business leads to value for the collective.

In addition to governance policy, organizations must implement procedures to integrate new data required by self-service developers. Procedures to onboard new data and present it as available within the data access layer will reduce the temptation of users to work outside prescribed analytical development guidelines. Then, to support shared usage of common data and build the environment where continuity is maintained, technical professionals may implement tools for data virtualization, semantic layers and metric stores. The use of these tools provides a common point where shared data and metric definitions reside, reducing (or eliminating) the need for varying business units to prep substantial amounts of data on their own. With the investment in these tools spread across the business, there is lower risk of redundant costs and maintenance efforts.

Document Revision History

Reference Architecture to Enable Real-Time Self-Service Analytics - 21 October 2020

Reference Architecture to Enable Real-Time Self-Service Analytics - 9 July 2019

Recommended by the Authors

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

Demystifying the Analytics and BI Space

How to Balance Control and Agility in Your Self-Service Analytics

Solution Path for Modernizing Analytic Architectures

Self-Service Analytics Governance With Microsoft Power BI

Critical Capabilities for Analytics and Business Intelligence Platforms

Comparing Platforms and Capabilities for Data Science and AI

Solution Path for Building a Holistic Data Management and Analytics Architecture

Adopting a Logical Data Warehouse

© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity." Gartner research may not be used as input into or for the training or development of generative artificial intelligence, machine learning, algorithms, software, or related technologies.

Table 1: 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 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 validate the work of others based on agreed upon

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

Source: Gartner (March 2022)

Table 2: Strengths and Weaknesses of Centralized and Decentralized A&BI Implementation Models

Implementation Model	Strengths	Weaknesses
Centralized A&BI: Information from across multiple source systems is brought together in a single location and then provides users access to the data via multiple tools and applications, including A&BI tools. This centralization approach helps improve efficiency in managing the data, improves the quality and helps govern it, since it is all in one place.	■ Eliminates redundant costs ■ Allows consistent data definitions ■ Presents an enterprise view of the data ■ Uses standardized, scalable architecture ■ Efficient management of data ■ Reinforces data quality and governance	■ Introduces bottlenecks ■ Uses a single data model ■ Restricts exploratory analysis ■ Creates significant backlogs

Implementation Model	Strengths	Weaknesses
Decentralized A&BI: Each LOB provides its own set of A&BI requirements, and IT helps design and implement the solution, selecting the business users' tool of choice. The same data coming from the source systems gets processed multiple times, with its own set of business rules and logic, generating conflicting outputs for users.	■ Allows agile delivery ■ Increases flexibility ■ Enables customized views ■ Leverages domain-specific knowledge	■ Generates siloed A&BI solutions ■ Allows inconsistent A&BI architecture ■ Produces conflicting data definitions ■ Increases operational and licensing cost overhead ■ Fosters disparate set of capabilities ■ Allows inconsistency of report development ■ Creates redundant processes and potentially conflicting business logic

Source: Gartner (March 2022)

Table 3: Governance Technologies

Tool Category	Description	Example Vendors
Metadata management	The definition of data governance implies effective management of data, which is impossible without managing its associated metadata. The selected tool should support manual and automated tagging of data to specific business domains, alongside an ability to search and steward the underlying data. See Market Guide for Active Metadata Management for further market analysis.	■ Adaptive ■ Alation ■ Alex Solutions ■ ASG Technologies ■ Collibra ■ Data Advantage Group ■ data.world ■ Quest (erwin) ■ Global IDs ■ IBM ■ Infogix ■ Informatica ■ Oracle ■ SAP ■ Semantic Web Co. ■ Smartlogic ■ Syniti

Data quality	Tools that support data quality are an absolute prerequisite to ensure successful implementation of the governance framework. The tool should provide profile, cleanse, merge and visualization capabilities and help create a workflow to maintain data in its pristine state. See Magic Quadrant for Data Quality Solutions for an assessment of this market and Data Quality Fundamentals for Data and Analytics Technical Professionals for guidance on data quality.	■ Ataccama ■ Datactics ■ Experian ■ IBM ■ Informatica ■ Innovative Systems ■ Melissa Data ■ MIOsoft ■ Precisely ■ Redpoint ■ SAP ■ SAS ■ Syniti ■ Talend
Data security	Data security tools can help locate sensitive data elements across multiple systems. They also help enforce regulatory, contractual and architectural compliance, support access management, and	■ Microsoft (BlueTalon) ■ PKWARE ■ Micro Focus Protegrity

	identify any security breaches (see Securing the Data and Advanced Analytics Pipeline).	■ Thales (Vormetric)

Source: Gartner (March 2022)

Table 4: Key Analytics and Business Intelligence Platform Features

Feature	Description
Analysis and Content Creation	
Advanced Analytics	Enables citizen data scientists and other power users to easily access statistical and analytical functions and machine learning models that are self-contained within the A&BI platform or are available for import and integration from externally developed models.
Automated Insights, NLG, NLQ	Enables business people and citizen data scientists to use machine-learning-based services to automatically find, visualize and narrate relevant findings such as correlations, exceptions, clusters, links and predictions without having to build models or write algorithms. Natural language query (NLQ) gives users the ability to query data using terms that are either typed into a search box or spoken. Natural language generation (NLG) entails the automatic creation of linguistically rich descriptions of insights found in data.
Data Visualization	The ability to create highly interactive dashboards and content, with visual exploration and embedded advanced and geospatial analytics, to be consumed by others.
Interactive Visual Exploration	Enables the exploration of data via manipulation of chart images, with the color, brightness, size, shape and motion of visual objects representing aspects of the dataset being analyzed.
Mobile Exploration and Authoring	Enables development and delivery of content to mobile devices in a publishing and/or interactive mode, and takes advantage of the native capabilities of

	mobile devices, such as touchscreens, cameras, location awareness and natural language query.
Data Management	
Catalog and Governance	Availability of functionality that enables the sharing of semantic model and metadata objects, including the ability to automatically generate and curate a searchable catalog of the artifacts created and used by the platform and their dependencies.
Data Preparation	Drag-and-drop, user-driven data combination of different sources and the creation of analytics models (such as user-defined measures, sets, groups and hierarchies).
Model Complexity and ETL	Enables support for complex data models, including the ability to handle multiple fact tables, interoperate with other analytic platforms and support knowledge graph deployments. Extract, transform, load (ETL) capabilities enable the access, integration, transformation and loading of data. ETL features include the ability to index data and manage data loads and refresh scheduling.
Infrastructure	
Cloud	The ability to support building, deploying and managing analytics and analytic applications in the cloud, based on data both in the cloud and on-premises, as well as across multicloud deployments.
Data Source Connectivity	Capabilities that allow users to connect to and ingest structured and unstructured data contained in various types of storage platforms, both on-premises and in the cloud.

Manageability	Enables a variety of A&BI platform administration and infrastructure functions, including capabilities for scaling the A&BI platform to optimize performance and ensure high availability and disaster recovery.
Security and User Admin	Enables platform security, administering users and auditing of platform access and utilization.
Share Findings	
Collaboration	Enables users to share and discuss information, analyses, analytic content and decisions via discussion threads, chat and annotation.
Embedded Analytics	Capabilities include providing a software developer kit (SDK) with APIs and support for open standards for creating and modifying analytic content, visualizations and applications. The capabilities also include analytic content embedding into a business process, application or portal.
Reporting	Capabilities allow users to publish and deploy analytic content delivery through various output types and distribution methods, with support for content search, storytelling, scheduling and alerts.

Source: Gartner (March 2022)