Gartner

Licensed for Distribution

This research note is restricted to the personal use of Saurabh Gupta (saurabh.gupta@ge.com).

2019 Planning Guide for Data and Analytics

Published 5 October 2018 - ID G00361501 - 66 min read

By Analysts Carlton Sapp, Daren Brabham, Joseph Antelmi, Henry Cook, Thornton Craig, Soyeb Barot, Doreen Galli, Sumit Pal, Sanjeev Mohan, George Gilbert

Supporting Key Initiative is Data and Analytics Programs

New data and analytics strategies promise to accelerate digital transformation, but success will depend on the variety of complementary architectures. Technical professionals must shift from fixed, rigid architectures to flexible data and analytics portfolios to better adapt to future demand.

More on This Topic

This is part of an in-depth collection of research. See the collection:

- 2019 Planning Guide Overview: Architecting Your Digital Ecosystem

Overview

Key Findings

- Data and analytics will drive business operations through a variety of design patterns that combine multiple architectural styles. Technical professionals must use a “portfolio-based” approach to delivering an end-to-end data and analytics architecture.

- More advanced analytics will continue to spread to places where it never existed before — from mobile devices to an onslaught of endpoints. Meeting the demand will require a combination of integration styles that deliver at the optimal point of impact.

- Artificial intelligence and machine learning will generate new synergies in information management, and play greater roles in complementing sections of the data and analytics architecture to optimize information management strategies.

- Data-driven business trends continue to offer great opportunities to raise the profile of data professionals. However, they also bring the need for new skills and architectures, and continue to challenge traditional methods and processes.

Recommendations

To deliver an effective data and analytics program, technical professionals should:

- Combine architectural styles into a portfolio-based approach to building end-to-end data and analytics architectures. Use the architecture outlined in this document as a baseline.
- Shift your focus from collecting data to embedding analytical functionality in existing applications and integrating that functionality into custom product offerings.
- Use ML as a tool to solve information management challenges. Start with invoking ML within the logical data warehouse to augment data ingestion strategies.
- Embrace new roles driven by rising business demand for analytics. Develop technical and professional effectiveness skills to support the end-to-end architecture vision.

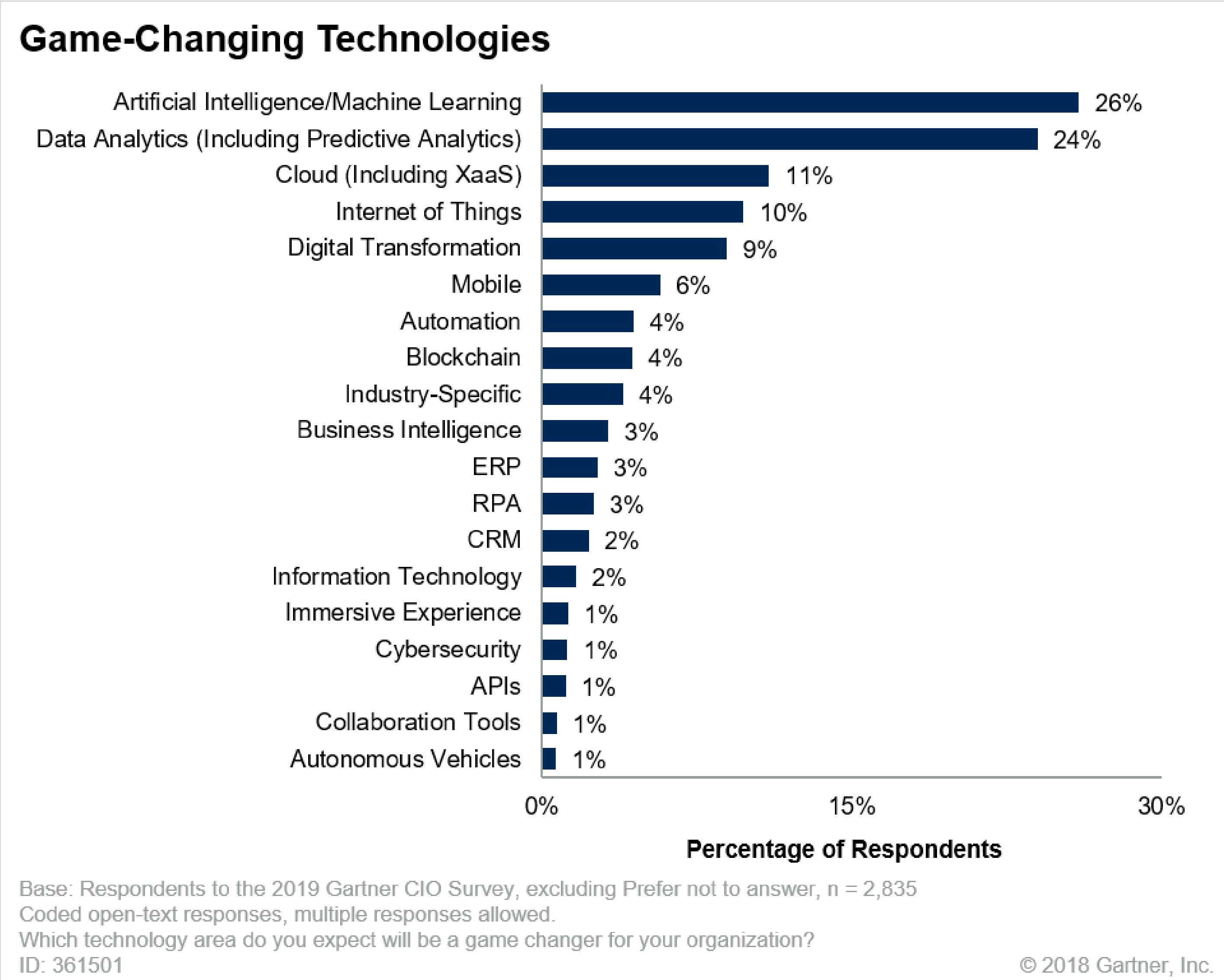
Data and Analytics Trends

More organizations are realizing the impact of data and analytics on business strategy and operations. However, fewer organizations are successful at exploring new ways to analyze, interpret and take advantage of data to drive competitive business advantages. The reason for this failure is simple: Organizations continue to drown in data without a plan to deal with the expected growth of data. This fruitless trend will continue unless technical professionals act now in preparing for future demand.

Data and analytics are showing no signs of slowing down. Data volume, variety and velocity continue to increase as business needs intensify. To help organizations capitalize on the opportunities that this information can reveal, data and analytics are taking on a more active and dynamic role in powering the activities of the entire organization, not just reflecting where it's been. More and more organizations are becoming truly "data-driven."

In a recent Gartner survey of CIOs, [1] artificial intelligence (AI) and machine learning (ML) together were identified as the technology category having the most potential to change the organization over the next five years (see Figure 1). Related categories, such as data analytics, also garnered significant attention. Taken together, AI/ML and data analytics represent a trend that can't be ignored: Analytics will drive significant innovation and disrupt established business models in the coming years.

Figure 1. Analytics' Potential to Drive Organizational Change



Source: Gartner (October 2018)

Many organizations claim that their business decisions are data-driven. But they often focus on reporting key performance metrics based on historical data — and on using analysis of these metrics to support and justify business decisions that will, hopefully, lead to desired business outcomes. While this approach is a good start, it is no longer enough.

> **Data and analytics are at the center of every competitive business, and are most effective when they are properly integrated into new or existing business processes.**

Data and analytics are no longer used just to support decision making; they are increasingly infused in places they haven't existed before. Today, data and analytics are:

- Shaping and molding external and internal customer experiences, based on predicted preferences for how each individual and group wants to interact with the organization
- Driving business processes, not only by recommending the next best action, but also by triggering those actions automatically
- Fueling AI and ML to better scale businesses through intelligent systems

Data and analytics continue to expand their role as the "brain" of the intelligent enterprise. They are becoming proactive as well as reactive, and coordinating a host of decisions, interactions and processes in support of business and IT outcomes.

To prepare for this trend, technical professionals must manage the end-to-end data and analytics process holistically. For several years, Gartner has recommended that organizations deploy a logical data warehouse (LDW) as a balance to dynamically integrating relevant data across heterogeneous platforms, rather than collecting all data in a monolithic warehouse. Key business benefits can be achieved by applying advanced analytics to these vast sources of data — and by providing business users with more self-service data access and analysis capabilities.

In 2019, we expect these trends to progress to the next level:

- Pervasive data and analytics will continue to demand a comprehensive end-to-end architecture using a portfolio-based approach.
- Organizations will invest to make analytics ubiquitous.
- AI and ML will generate new synergies in information management.
- Analytics services in the cloud will continue to accelerate to deliver greater performance at scale.
- Revolutionary changes in analytics will drive IT to adopt new technologies and roles.

In 2019, forward-thinking IT organizations will encourage "citizen" and specialist users by deploying self-service integration and analytics capabilities. They will also focus IT efforts on operationalizing and scaling analytics within the context of the organization's broader technology infrastructure. To enable their organizations' algorithmic potential, technical professionals must work to ensure that an arsenal of analytics is integrated into the fabric of autonomous processes, services and applications.

Pervasive Data and Analytics Will Continue to Demand a Comprehensive End-to-End Architecture Using a Portfolio-Based Approach

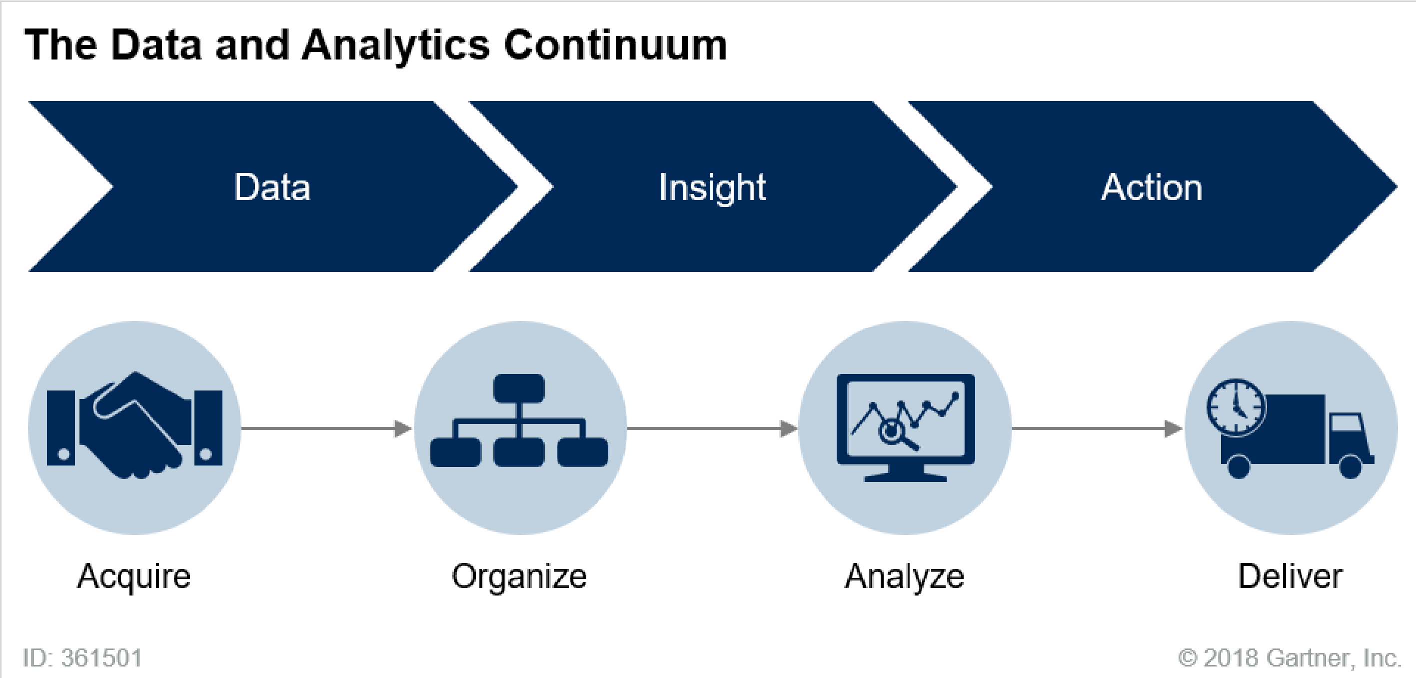
The volume, variety and velocity of data in today's business world continue to increase. Inexpensive computing at the edge of the enterprise enables a huge amount of information to be captured and processed. Be it video from closed-circuit cameras, temperature data from an Internet of Things (IoT) solution or RFID packets indicating product locations in a retailer's warehouse, the variety of data presents major challenges.

In addition, diverse sources of external, often cloud-based data are now being used to enrich customer, prospect and partner understanding. In some cases, a tipping point is reached where the gravity of data skews toward external, rather than internal, data.

These changes will force IT to envision a revitalized data and analytics continuum that incorporates diverse data and that can deliver "analytics everywhere" (see Figure 2). Some enterprises are capturing all data in hopes of uncovering new insights and spurring possible actions. Others are starting with the end goals in mind, after all the data has been generated for a specific purpose. This allows them to streamline the process and manage an end-to-end architecture that supports specific desired outcomes.

> **The complex and forever-changing requirements of analytics have created a scenario in which no one architectural style is sufficient to execute all required analytical use cases. To meet future demand, a more proactive and coordinated data and analytics strategy founded on a repository of architectural styles will be required.**

Figure 2. The Data and Analytics Continuum



Source: Gartner (October 2018)

Data, insight and action no longer represent separate disciplines, regardless of the approach. Many companies already combine them. The technical professional must fuse them into one architecture that encompasses the following:

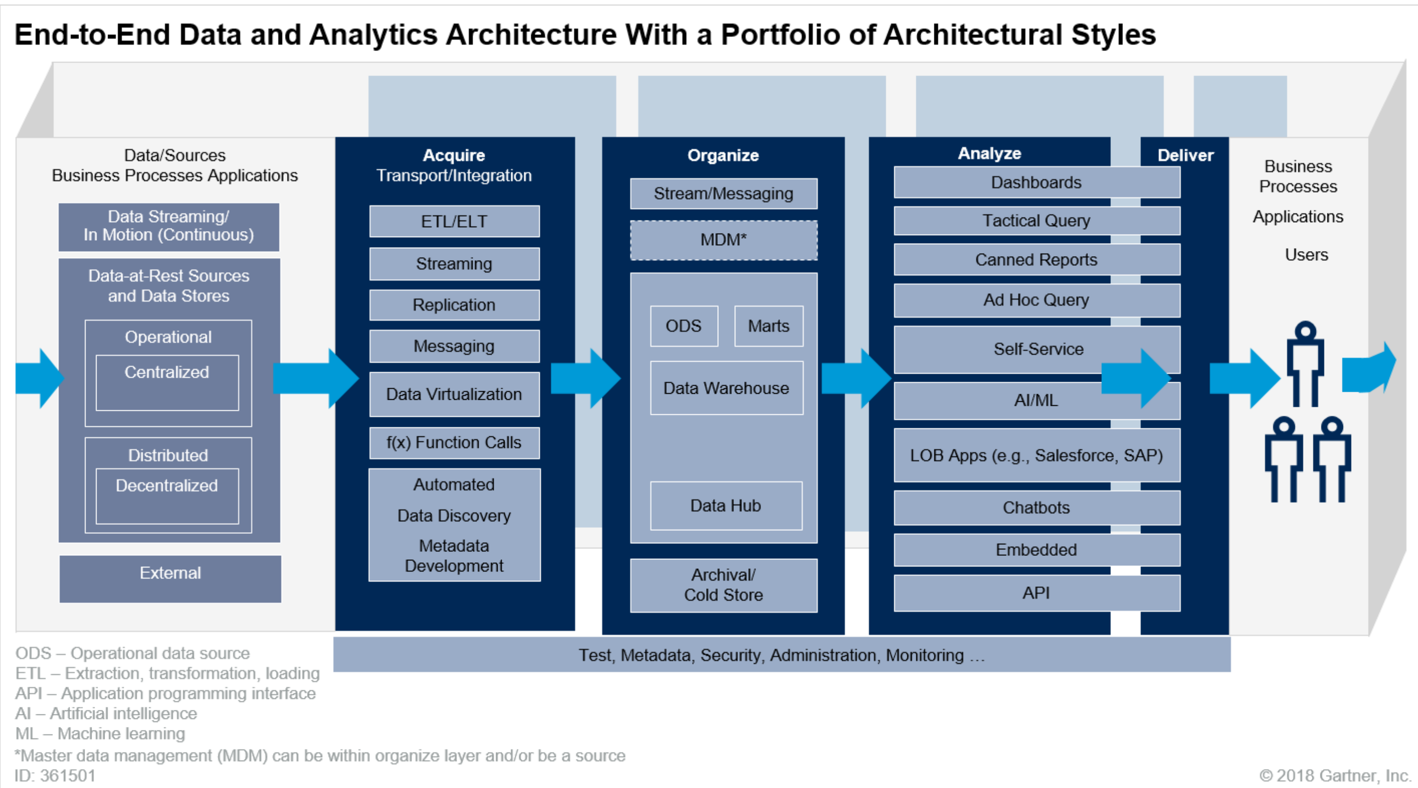
- **Data acquisition,** from anywhere the information is generated. An important aspect of acquisition is **integration** and combining the acquired data, regardless of type of data (structured versus unstructured), its velocity, or its veracity (especially for macroanalysis — the result of multiple data analytics results combined together).
- **Organization** of that data — for example, by using the multiengine LDW analytical architecture. The LDW can form the core to connect to data as needed, as well as collect it into efficient physical data servers. This theme of "collect and connect" is a major trend.
- **Analysis** of data when and where it makes most sense. This can include reporting, tactical and ad hoc querying, data visualization, ML, and more.
- **Delivery** of insights and data at the optimal point of impact to:
  - Support human activities with just-in-time insights.
  - Embed analysis into business processes that are capable of performing some action.
  - Analyze data as it streams into the enterprise and automatically take action based on the results.

This change doesn't mean that organizations should discard all of their traditional data and analytics techniques and approaches and replace them with new ones. The shift will be gradual and incremental — but also inevitable. Increasingly, analytics will drive business processes, not simply analyze them after the fact.

Planning Considerations

In 2019, technical professionals must build a data management and analytics architecture that can support changing and varied data and analysis needs. This architecture must accommodate both traditional data analysis and newer analytics techniques. It should be modular by design to accommodate mix-and-match configuration options as they arise. Figure 3 shows a high-level representation of such an architecture. This fits in with the major layers shown in Figure 2.

Figure 3. End-to-End Data and Analytics Architecture With a Portfolio of Architectural Styles



Source: Gartner (October 2018)

This is not intended to be prescriptive, as not every organization will have all the components. Architecting the system in layers helps to address numerous issues. Not all of the layers may be activated all of the time; however, to meet the demand of future requirements, technical professionals should use this architecture to build a portfolio of solutions to support various use cases.

Other, related planning considerations for technical professionals in 2019 include the need to:

- Extend the data architecture to acquire streaming and cloud-born external data.

- Modernize your data integration layer by enabling greater data delivery styles.
- Develop a virtualized data organization layer to connect to data as well as collect it.
- Develop a comprehensive analytics environment that spans from traditional reporting to prescriptive analytics.
- Deliver data and analytics at the optimal point of impact.

Extend the Data Architecture to Acquire Streaming and Cloud-Born External Data

The “acquire” stage in Figure 3 embraces all data, providing the raw materials needed to enable downstream business processes and analytics activities. For example, IoT requires data and analytics professionals to proactively manage, integrate and analyze real-time data. Internal log data often must be inspected in real time to protect against unauthorized intrusion, or to ensure the health of the technology backbone. Strategic IT involvement in sensor and log data management on the technology edge of the organization will bring many benefits, including enhanced analytics and improved operations. There should be a clear business purpose behind holding and processing this data. Other streaming sources, like social media feeds, bring more real-time data processing requirements while supporting additional business use cases. There should be a clear business purpose behind holding and processing any stream of data.

> **Data is the raw material for any decision.**

Organizations must shift their focus from getting data *in* and hoping someone uses it, to determining how to best get information *out* to the people and processes that will gain value from it.

The sheer volume of data can clog data repositories if technical professionals subscribe to a “store everything” philosophy. For example, ML algorithms can assess incoming streaming data at the edge and decide whether to store, summarize or discard it.

In-stream processing is emerging as a method for performing data quality, analytics and transformations as the data moves throughout the organizational pipeline. Certain use cases, like live analytics or real-time data cleansing, can be done in-stream and offloaded from target systems, allowing less persistence of data outside the stream.

The system architect should centralize most of the data quality. It is likely that data transformation and quality is a large part of the processing. Also, it is likely that a lot of your data is structured data that makes up a large amount of your reporting. In this case, it is simply more productive — and better from a governance point of view — to do this once. Then you can let everyone share the results. It makes no sense to force every analyst to do this for themselves.

A holistic understanding of *how* the data will be used is another key aspect of the end-to-end thinking required to determine whether and when to store data.

Beyond streaming data, considerable value-added content is available from third parties. Syndicated data comes in a variety of forms, from a variety of sources. Examples include:

- Consumer data from marketing and credit agencies
- Geolocation data for population and traffic information
- Weather data to enhance predictive algorithms, for purposes such as improving public safety or forecasting retail shopping patterns
- Risk management data for insurance

Businesses have been leveraging this type of data for years, often getting a fee-based periodic feed directly from the data provider. Today, however, increasing quantities of this data are available through cloud services — to be accessed whenever and wherever needed. A data and analytics architecture that can embrace these new forms of data in a dynamic manner is essential to providing the contextual information needed to support a data-driven digital business. For more information on the types of data available, see "Understand the Data Brokerage Market Before Choosing a Provider." (https://www.gartner.com/document/code/334439?ref=grbody&refval=3891182)

Modernize Your Data Integration Layer by Enabling Greater Data Delivery Styles

Data integration requirements are becoming increasingly diverse. They now demand real-time streaming, replication and virtualized capabilities, in addition to the more traditional bulky/batch data movement principles. To address the challenge of a disjointed data integration infrastructure, technical professionals should build a more "portfolio-based" approach to data integration.

The data integration discipline comprises the practices, architectural techniques and tools that ingest, transform, combine and provide data across the spectrum of information types (within enterprises and beyond) to meet the data consumption requirements of applications and business processes.

A mass proliferation of data associated with the rise of IoT, big data and digital business means that data integration teams, tools and architectures are under constant pressure to deliver integrated data:

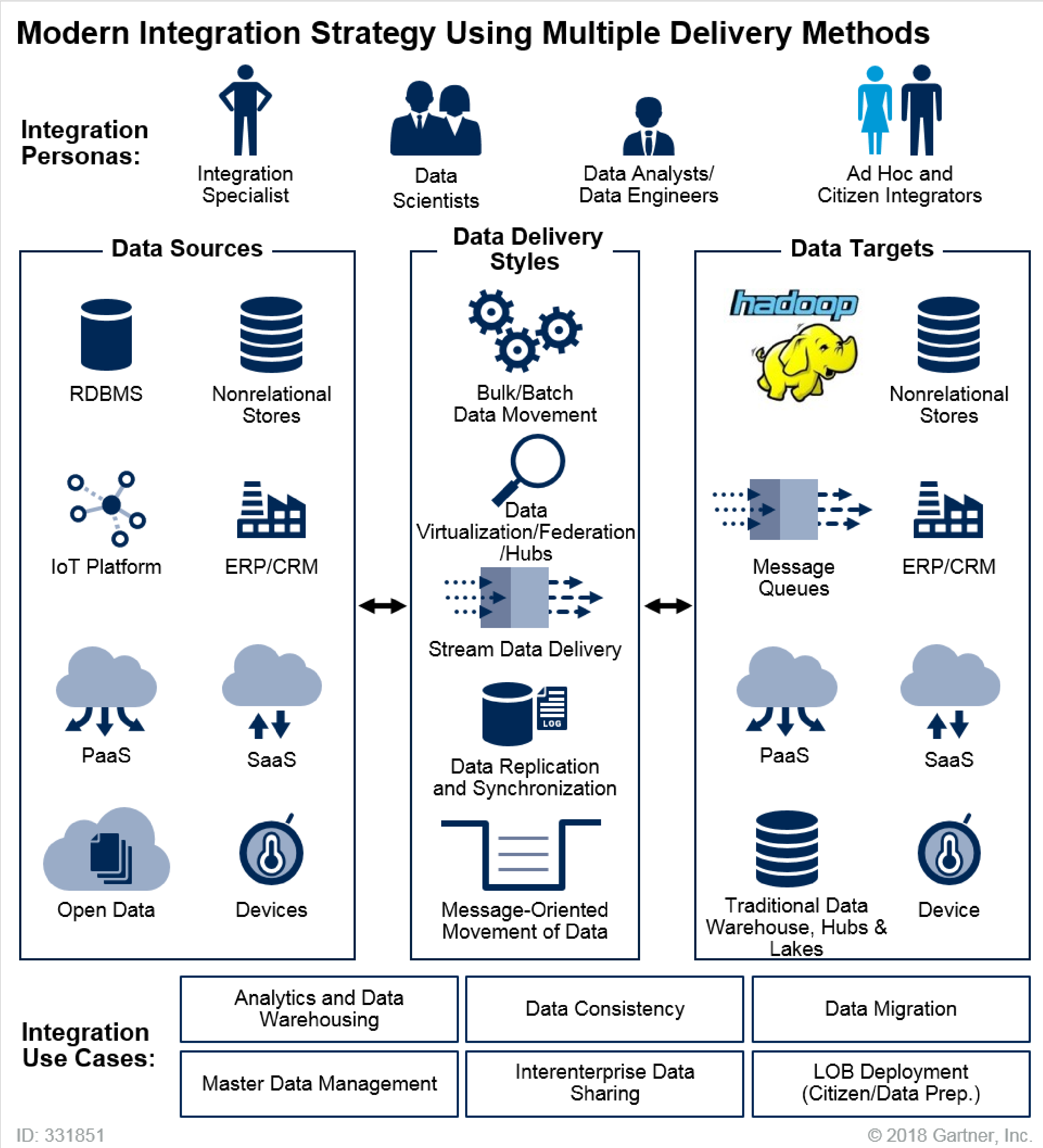
- At *variable latencies* (not simply fixed in batches or in real time)

- Across *all deployment scenarios* ("edge," IoT, on-premises, in the cloud or hybrids of all of these)
- For *all required data sources and types* (not just structured, but also multistructured)
- Across *all use-case scenarios* (not just for analytics, but also for operational requirements)
- For *all user personas* (not just integration specialists, but also business users and citizen integrators)

For more information, see "The State and Future of Data Integration: Optimizing Your Portfolio of Tools to Harness Market Shifts." (https://www.gartner.com/document/code/297000?ref=grbody&refval=3891182)

Figure 4 illustrates the point that a modern data integration strategy needs a combination of data delivery styles.

Figure 4. Modern Data Integration Strategy Using Multiple Delivery Methods



Source: Gartner (October 2018)

Data delivery styles for modern data integration challenges may include:

- Bulk/batch
- Message-oriented data movement

- Data replication
- Change data capture (CDC)
- Data synchronization
- Data virtualization
- Data hubs
- Streaming/event data delivery

Develop a Virtualized Data Organization Layer to Connect to Data and Collect It

To deal with the many different uses, varieties, velocities and volumes of data today, IT must employ multiple data stores across cloud and on-premises environments. However, IT cannot allow these multiple data stores to prevent the business from obtaining actionable intelligence. By employing an LDW approach, organizations can avoid creating a specialized infrastructure for unique use cases, such as big data. The LDW provides the flexibility to accommodate any number of use cases using a variety of data stores and processing frameworks. Big data is no longer a separate, siloed, tactical use case; it is simply one of many use cases that the architecture can accommodate to enable the digital enterprise.

> **The LDW integrates three analytics development styles: the classic data warehouse, an agile approach with data virtualization and the data lake.**

The core of the "organize" stage of the end-to-end architecture is the LDW (see "Solution Path for Planning and Implementing the Logical Data Warehouse" (https://www.gartner.com/document/code/320563?ref=grbody&refval=3891182)). An LDW:

- Provides a modern, scalable data management architecture that can support the data and analytics needs of the digital enterprise.
- Supports an incremental development approach that leverages the existing enterprise data warehouse architecture and techniques in the organization.
- Establishes a shared data access layer that logically relates data, regardless of source.

Building the LDW and the end-to-end analytics architecture will require technical professionals to combine technologies and components that provide a complete solution. This process requires significant data integration and an understanding of data inputs and existing data stores. The many technical choices available for building the LDW can be overwhelming. The key is to choose and integrate the technology that is most appropriate for the organization's needs. This work needs to be done by technical professionals who specialize in data integration. Hence, 2019 will see the continued rise of the data architect role.

Many clients still directly access many data sources using point-to-point integration. This means that any changes in data sources can have a disruptive impact. Although it's not always possible to stop all direct access to data, shared data access can minimize the proliferation of one-off direct access methods. This is especially true for use cases that require data from multiple data sources.

To increase the value of shared data access, organizations should:

- Define a business glossary, and enable traceability from data sources to the delivery/presentation layer.
- Use various levels of certification for data integration logic, thereby creating a healthy ecosystem that enables self-service data integration and analytics.
- Take an incremental approach to building this ecosystem to avoid the failures of past "big bang" approaches.

Although technical professionals can custom code the shared data access layer, commercial data virtualization tools provide many advantages over a custom approach. These advantages include comprehensive connectors, advanced performance techniques and improved sustainability. Gartner recommends that clients deploy these virtualization tools to create the data virtualization layer on top of the LDW. Providers that offer stand-alone data virtualization middleware tools include Cisco, Denodo, IBM, Informatica, Information Builders, Oracle and Red Hat.

Tools that can be leveraged for data virtualization may already exist in your organization. Business analytics (BA) tools typically offer embedded functions for this purpose. However, these tools are unsuitable as long-term, comprehensive, strategic solutions for providing a data access layer for analytics. They tend to couple the data access layer with specific analytical tools in a way that prevents the integration logic or assets from being leveraged by other tools in the organization.

To increase the value of shared data access, organizations should:

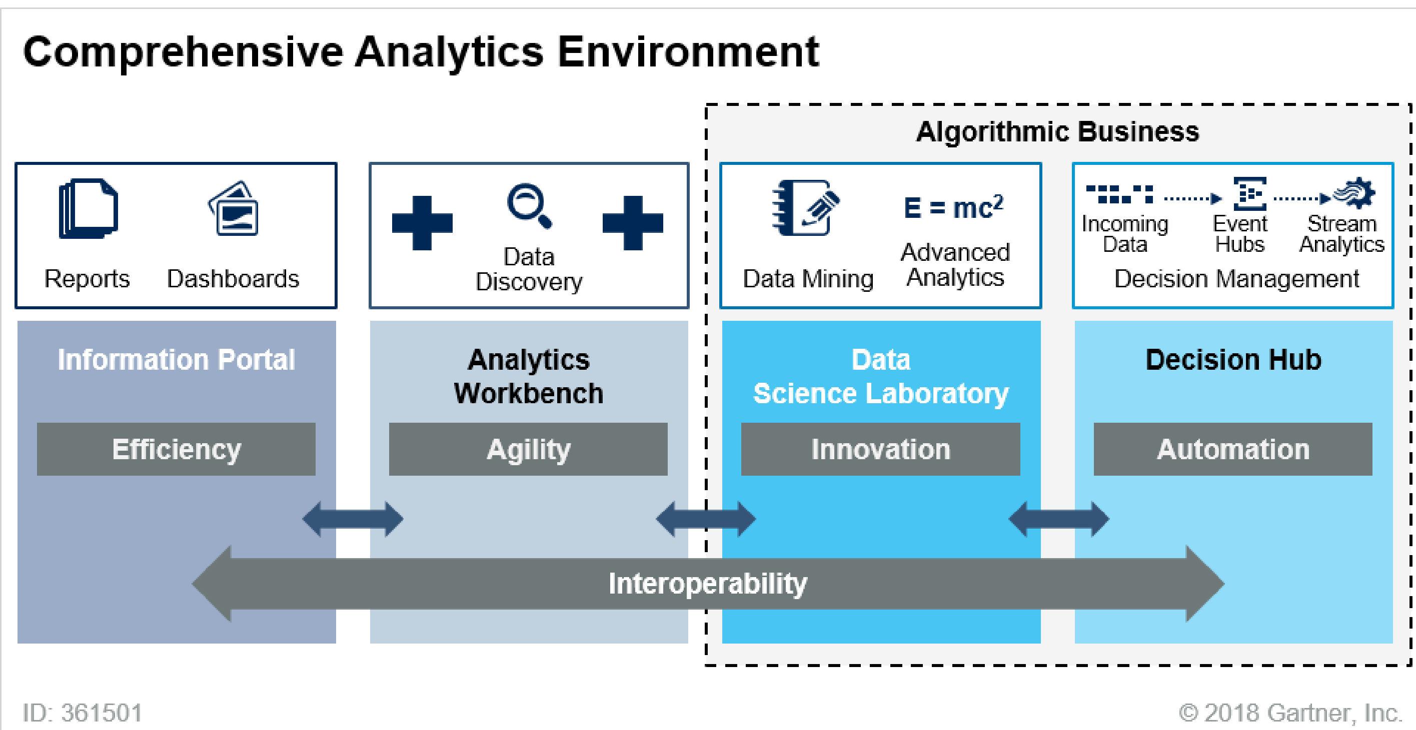
- Define a business glossary, and enable traceability from data sources to the delivery/presentation layer.

- Use various levels of certification for data integration logic, thereby creating a healthy ecosystem that enables self-service data integration and analytics.
- Take an incremental approach to building this ecosystem, to avoid the failures of past “big bang” approaches.
- Use automated data and metadata discovery tools to detect and manage copies of data across the organization.

Develop a More Comprehensive Analytics Environment

Comprehensive business analytics requires more than just providing tools that support analytics capabilities (see Figure 5). There are likely to be several types of users, and therefore several tools — possibly overlapping. However, this is preferable to trying to force-fit multiple user groups into using the same tool.

Figure 5. It Is Normal to Require a Mix of Analytical Tools



Source: Gartner (October 2018)

Many components are needed to build out an end-to-end data architecture that encompasses:

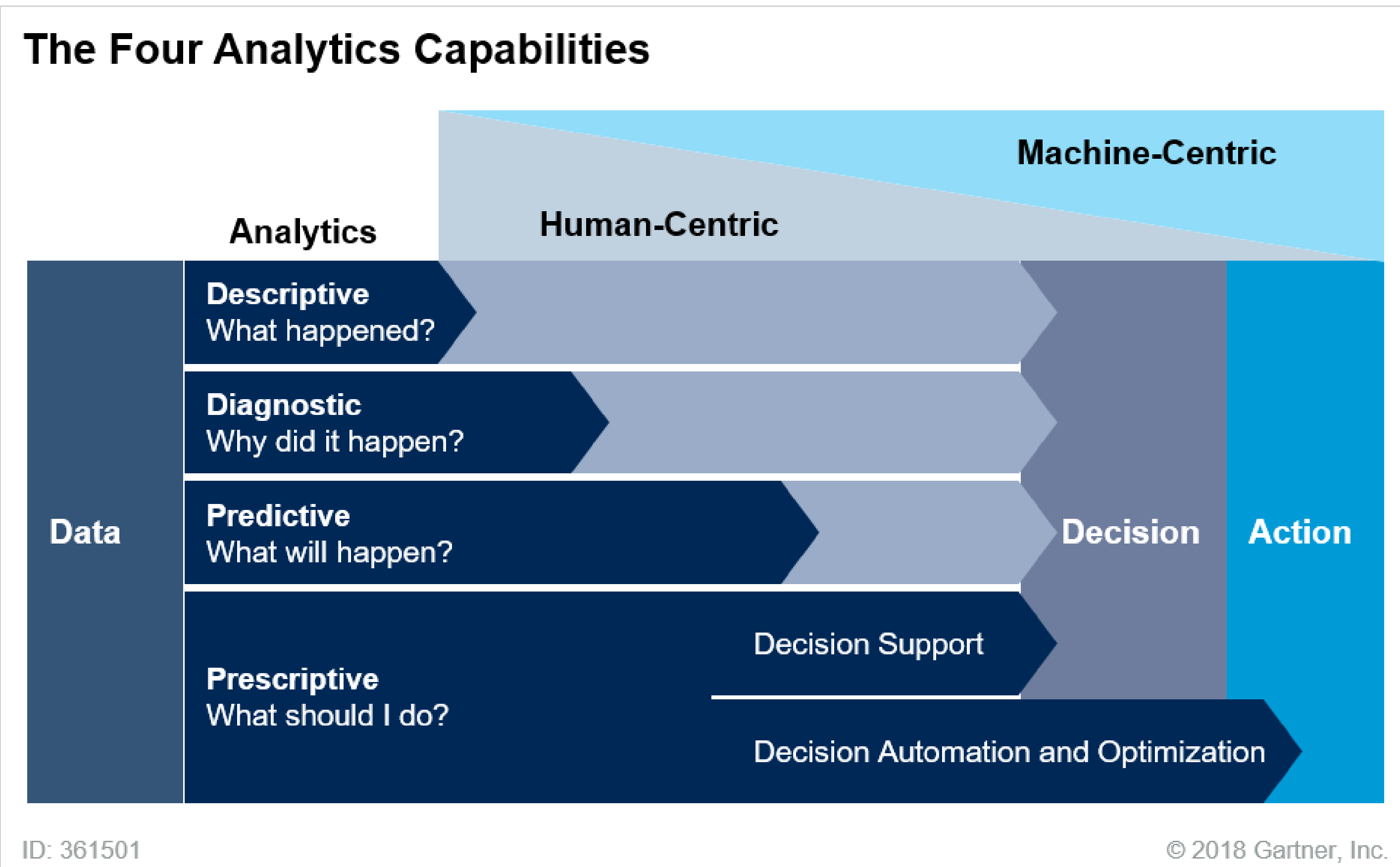
- Delivery and presentation of analyses
- Data ingestion and transformation
- Data stores

- Collaboration on results
- Integration with enterprise business applications

The "analyze" phase of the end-to-end architecture can be simple for some. However, as demand for predictions and real-time reactions grows, this phase can become increasingly multifaceted.

The range of analytics capabilities available goes beyond traditional data reporting and analysis (see Figure 6). Although Gartner estimates that a vast majority of organizations' analytics efforts (and budgets) are spent on descriptive and diagnostic analytics, a significant part of that work is now handled by business users doing their own analysis. This work often occurs outside the realm of the sanctioned IT data and analytics architecture. Predictive and prescriptive capabilities, on the other hand, are usually focused within individual business units and are not widely leveraged across the organization. That mix must change.

Figure 6. The Four Analytics Capabilities



Source: Gartner (October 2018)

Organizations will need to provide more business and IT institutional support for advanced analytics capabilities. In digital businesses, however, activities will be interactively guided by data, and processes will be automatically driven by analytics and algorithms. IT organizations must invest in ML, data science, AI and cognitive computing to automate their businesses. This automated decision making and process automation will represent a growing percentage of future investment and innovation.

Data and analytics professionals must embrace these advanced capabilities and be prepared to enable and integrate them for maximum impact. Programmatic use of advanced analytics (as opposed to a sandbox approach) is also on the rise, and must be managed as part of an end-to-end architecture.

Deliver Data and Analytics at the Optimal Point of Impact

The “deliver” phase of the end-to-end data and analytics architecture is often overlooked. For years, this activity has been traditionally equated with simply producing reports, interacting with visualizations or exploring datasets. But those actions involve only human-to-data interfaces and are managed by BA products and services. Analytics’ future will increasingly be partly human-interaction-based and partly machine-driven.

> **An expanding mesh of rich connections between devices, things, services, people and businesses demands a new approach to data delivery.**

Key considerations in the delivery of analyzed information include:

- **Devices and gateways.** Users can subscribe to content and have it delivered to the mobile device of their choice, such as a tablet or a smartphone. Having access to the right information, in the optimal form factor, increases adoption and value. For example, retail district managers may need to access information about store performance and customer demographics while they are in the field, without having to open a laptop, connect to a network and retrieve analysis.
- **Applications.** Organizations can embed in-context analytics within applications to enrich users’ experiences with just-in-time information that supports their activities. For example, a service technician could view a snapshot of a customer’s past service engagements and repairs while diagnosing the cause of a problem. Applications can also be automated using predictions generated by analytics processes running behind the scenes. One example is IoT-connected equipment: Diagnostics can be assessed in near real time to determine whether a given machine is at risk of failure and in need of maintenance.
- **Processes.** The output of an analytics activity — be it in real time or in aggregate — can recommend the next step to take. That result, coupled with rules as to what to do when specific conditions are met, can automate an operational process. For example, if a sensor in a refrigerated

storage area indicates that the temperature is rising, analytics can determine whether a problem exists and, if so, dispatch a technician to the site.

- **Data stores.** Data generated from one analytics activity can be used in other analytics activities. That is, the output of one activity can be the input to another. This is often the case when the organization seeks to monetize its data to external audiences. Insights generated by acquire, organize or analyze activities are output to another data store for eventual access by third parties that need the data to support their decisions and actions. The emergence of connected business ecosystems will drive even more of these changes.
- **Interfaces.** By embedding AI into applications and designing a fit-for-digital architecture, one can create unprecedented system integration, developing digital twin models and deploying advanced human interfaces. These advanced interfaces may use natural language, be embedded as chatbots or even use human voice interaction.

The range of analytics options must be integrated into the fabric of how you work. We address the “how” aspect of this planning more fully in the next section.

Organizations Will Invest to Make Analytics Ubiquitous

The “data-driven” mantra is nothing new, but organizations still struggle to overcome established practices and supplant them with new analytics processes. Decisions should no longer be left to gut instinct. Instead, decisions and actions should be based on facts, with algorithms used to predict optimal outcomes.

> **Use analytics to proactively make decisions that drive action and influence the future course of the organization.**

In general, analytics are becoming more pervasive in business. More people want to engage with data, and more interactions and processes need analytics to automate and scale. Use cases are exploding in the core of the business, on the edges of the enterprise and beyond. This trend goes beyond traditional analytics, such as data visualization and reports. Analytics services and algorithms will be activated whenever and wherever they are needed. Whether to justify the next big strategic move or to optimize millions of transactions and interactions a bit at a time, analytics and the data that powers them are showing up in places where they rarely existed before. This is adding a whole new dimension to the concept of “analytics everywhere.”

Not long ago, IT systems' main purpose was to automate processes. Data was stored and then analyzed to assess what had already happened. That passive approach has given way to a more proactive, engaged model, where systems are architected and built around analytics. Today, analytics capabilities are:

- Embedded within applications (IoT, mobile and web) to assess data dynamically and enrich the application experience
- Just-in-time, personalizing the user experience in the context of what's occurring in the moment
- Running silently behind the scenes and orchestrating processes for efficiency and profitability

Massive amounts of data at rest have fueled innovative use cases for analytics. With the addition of data in motion — such as sensor data streaming within IoT solutions — new opportunities arise to use ML and AI, in real time, to assess, scrub and collect the most useful and meaningful information and insights.

These developments do not mean that traditional analytics activities will cease to be important. Business demand for self-service data preparation and analytics continues to accelerate, and IT should enable these capabilities. As data and analytics expand to incorporate ecosystem partners, this demand will also increase from outside the organization.

Planning Considerations

In 2019, technical professionals can expect even more emphasis on analytics as it is embraced throughout the enterprise. The expansion from human-centric interaction to machine-driven automation will have a profound impact on how analytics will be deployed.

Related planning considerations for technical professionals in 2019 include the need to:

- Incorporate enterprise information management (EIM) and governance for internal and external use cases.
- Integrate fragmented analytics initiatives to improve operations and scalability.
- Prepare for the onrush of machine learning.
- Enhance application integration skills to embed analytics everywhere.
- Adopt agile database development strategies.

Incorporate EIM and Governance for Internal and External Use Cases

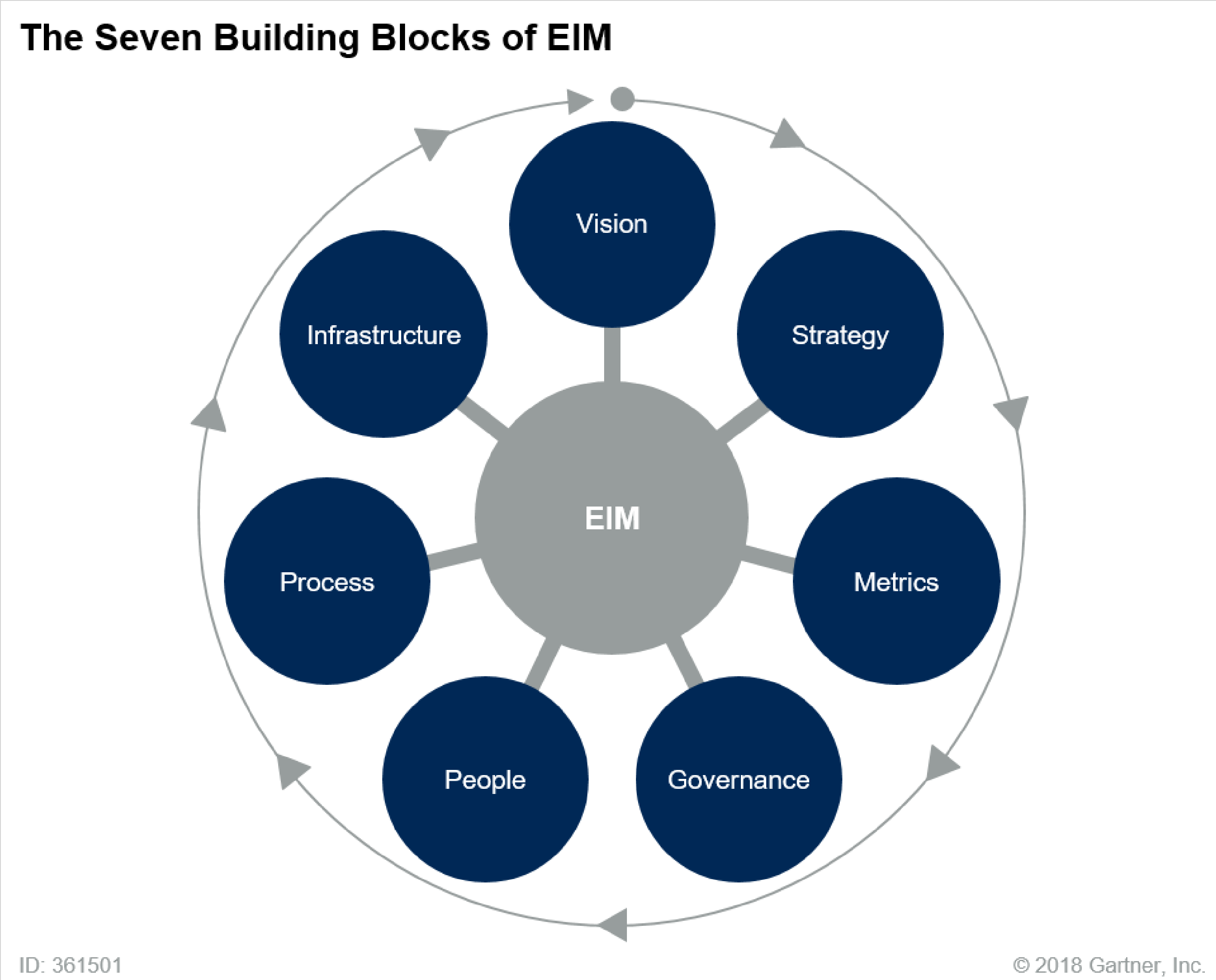
Enterprise information management is an integrative discipline for structuring, describing and governing information assets — regardless of organizational and technological boundaries — to improve operational efficiency, promote transparency and enable business insight. As the data sources for analytics increasingly reside outside of the analytics group's or the organization's control, it becomes more important to assert just enough governance over all sources of analytics data to enable new and existing use cases. All types of data must be included in an EIM architecture — including structured data in traditional relational databases, semistructured and unstructured data found in data lakes, and streams of data that have not yet landed.

> **Effective EIM synchronizes decisions between strategic, operational and technical stakeholders, coordinating efforts to improve the organization's analytics capabilities.**

An EIM program based on sound information governance principles is an effective tool for managing and controlling the ever-increasing volume, velocity and variety of enterprise data to improve business outcomes. EIM is increasingly needed in today's digital economy. It remains a struggle, however, to design and implement enterprisewide EIM and information governance programs that yield tangible results. In 2019, a key question for technical professionals and their business counterparts will be, "How do we successfully set up EIM and information governance?"

Most successful EIM programs start with one or more initial areas of focus, such as master data management (MDM), data quality, data integration or metadata management initiatives. All EIM efforts need to include the seven components of effective program management shown in Figure 7. For more information on EIM, see "Solution Path for Planning and Implementing a Comprehensive Architecture for Data and Analytics Strategies" (https://www.gartner.com/document/code/351281?ref=grbody&refval=3891182) and "EIM 1.0: Setting Up Enterprise Information Management and Governance." (https://www.gartner.com/document/code/342309?ref=grbody&refval=3891182)

Figure 7. The Seven Building Blocks of EIM



Source: Gartner (October 2018)

Metadata should be added to support EIM. Metadata management is a supportive function that helps to enable EIM and data governance programs. Organizations are leveraging toolsets to capture technical and operational metadata to build the basis for data lineage and consumption. ML algorithms deployed in these tools can crawl multiple types of data and build catalogs to support both business processes and data governance functions.

For more details on metadata, see: "Deploying Effective Metadata Management Solutions." (https://www.gartner.com/document/code/347645?ref=grbody&refval=3891182)

Integrate Fragmented Analytics Initiatives to Improve Operations and Scalability

The growing range of new analytics use cases across organizational boundaries will drive the need for fragmented analytics to improve operations and scalability.

Consider the following example: In a biotech firm, data scientists working in the genomics division conduct their own small-scale analytics project that reveals some potentially transformative insights for the company. However, they developed the algorithms used for their analysis on their laptops, employing their own unique operating environment and programming languages. In this case, every time an executive wants them to refresh that insight or analytics capability, the data scientists need to run some data science routines on their own laptops, and pull data down from their own local systems. This effort is labor-intensive and doesn't scale well. If this implementation grows, it will likely become increasingly unwieldy, and it will lack the organization's standard controls in areas such as data governance, privacy and security. If multiple efforts like this spring up throughout the organization, problems and inefficiencies can multiply exponentially.

Many organizations have decided that they cannot wait for IT to deliver the data and intelligence they need. They have instead forged ahead with their own initiatives — a situation that has led to "shadow analytics" stacks and a certain degree of anarchy. Too many shadow analytics efforts can cause issues such as data inconsistency, inefficiency and security breaches. To deal with this anarchy:

- First, data-related technical professionals must discover where ad hoc analytics efforts have sprung up in the enterprise. To do this, they need to reorient themselves to be more collaborative and socially aware of these fragmented initiatives.
- Technical professionals must then collaborate with business users to build the case for an infrastructure and environment that will help business users effectively, safely and quickly perform these analytics on their own. Additionally, by getting involved, technical professionals can not only form relationships that will help track analytics activities, but also support those analytics activities with data, resources, processes and frameworks.

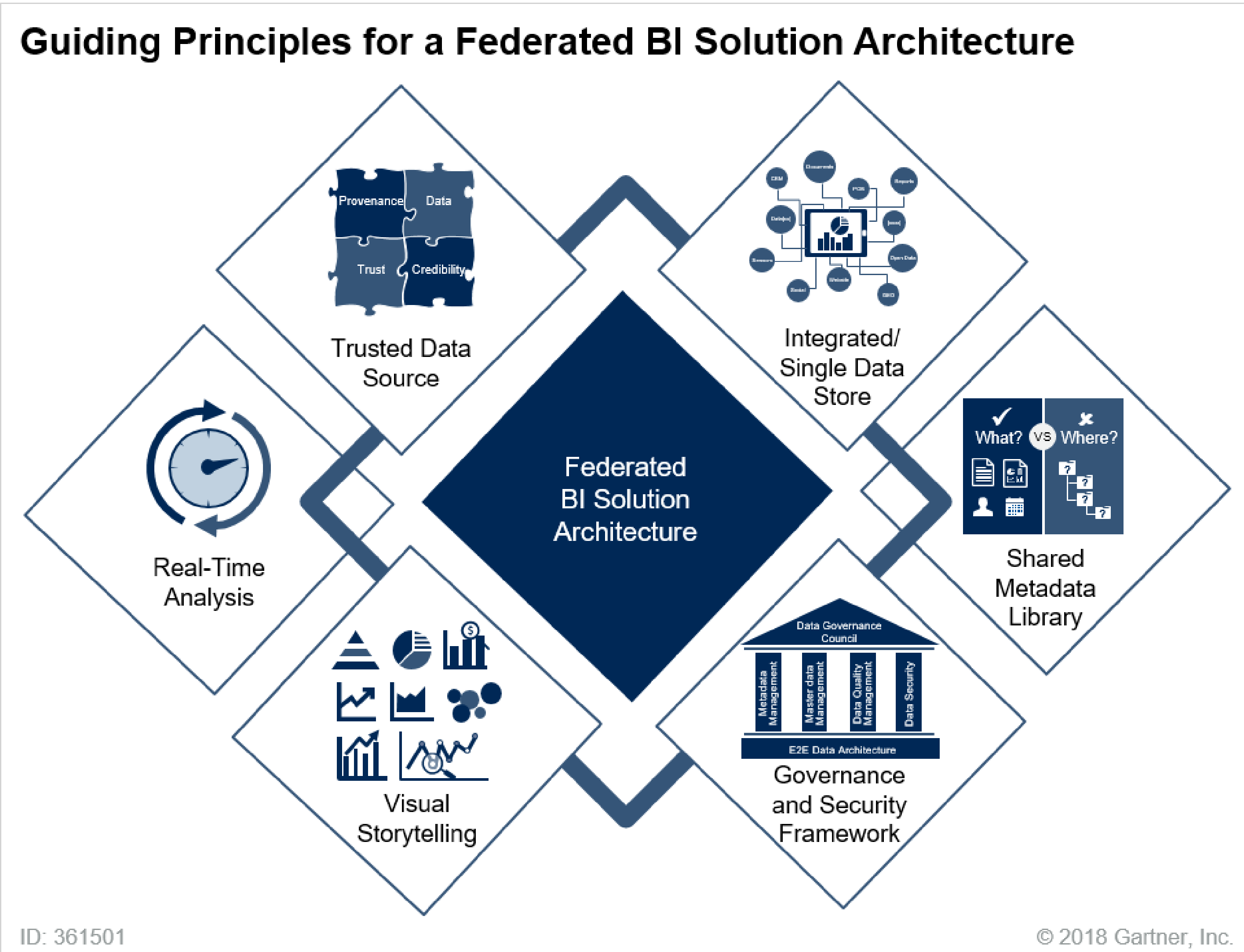
If IT does not act, then the IT organization will miss opportunities to identify — and eventually operationalize — shadow analytics.

> **Technical professionals must shift and expand their role from being content creators and data access controllers to user and data enablers.**

IT professionals can empower business-led analytics with professional tooling and best practices, helping to maximize ROI. Likewise, business analysts can team with IT to ensure business focus.

A key part of this effort is to facilitate a self-service data and analytics approach. For most organizations, this will mean building an architecture to support different analytics and business intelligence (BI) platforms for different users, as shown in Figure 8.

Figure 8. Guiding Principles for a Federated BI Solution Architecture



Source: Gartner (October 2018)

The architecture shown above should not be solely centralized or decentralized, but federated. The guiding principles for a federated architecture are that it should have:

- **An integrated single data store:** An LDW can provide a centralized data hub to store all the data, help identify new relationships between data elements, and reduce data movement across multiple data stores and analytical systems.

- **A shared metadata library:** A centralized business glossary helps standardize data definitions. Metadata, defined as the set of data that describes and provides context about the data elements, can help support discovery and self-service analytics. Providing the ability to search and filter an enriched metadata library helps users locate relevant information to use for further analysis. This metadata library can be created via an automated process using profiling tools and ML-driven AI, or it can be created manually with the help of data stewards.

- **A governance and security framework:** Data governance, as discussed in the previous section, will provide a set of processes, implemented and used by stakeholders leveraging technology, to ensure that critical data is protected and well-managed. This becomes significantly more important when developing an LDW that is built upon a data lake. If the data is governed properly at the physical layer, it reduces the complexity of managing access for the users from within the BI tools.

- **A trusted data source:** The combination of the LDW with a metadata library that is both governed and secured presents a trusted data source for the users to do their analysis. This provides them with a unique set of capabilities and promotes self-service analytics across the enterprise with less reliance on IT. Users now have a single location to fetch pristine-quality data. This supports ad hoc analysis and standardized operational reporting, and even advanced analytics and machine learning in the future.

- **Real-time analysis:** Information about consumers and sales received in real time is more relevant than information from last month or even last week. It is more useful to your business users because it gives them a true understanding of what is happening within their business operations — as it is happening.

- **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. What better way to do that than by visually representing it 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 provide an interactive way to visualize and deliver the data to the end users. Having established the principles and the goal to turn shadow IT into managed self-service, it is also important to consider the capabilities that users will need to have in place. A typical organization contains the following four business user groups, which are further classified into two major categories, depending on the functions they perform:

  - **Information producers:**

    - Data scientists

    - Power users (report authors)

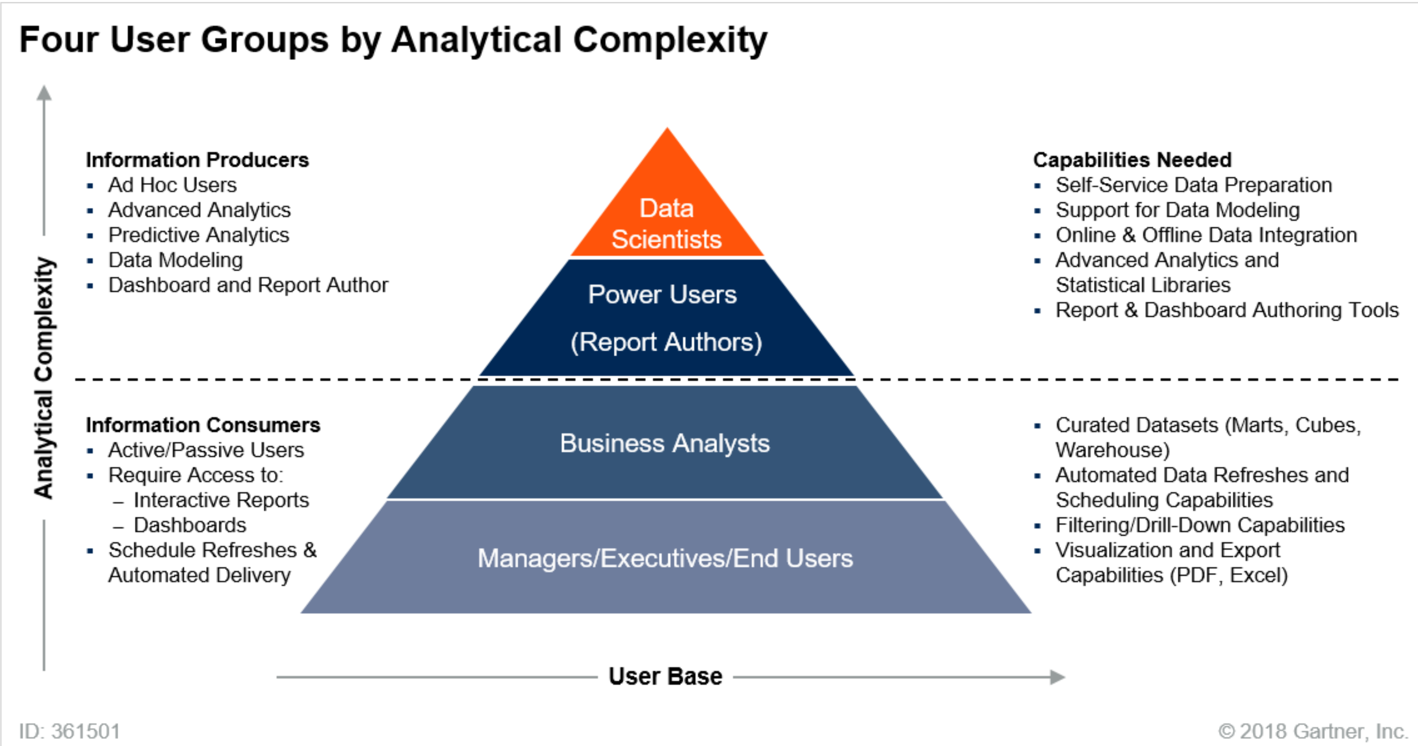
  - **Information consumers:**

    - Business analysts

    - Managers, executives and end users

Each business user group has different analytical needs. See Figure 9 for details.

Figure 9. Four Business User Groups by Analytical Complexity



Source: Gartner (October 2018)

Finally, with the principles, and the understanding of users, in place, it is time to implement federated BI.

> **A federated model is a pattern within [the enterprise architecture (https://en.wikipedia.org/wiki/Enterprise_architecture)](https://en.wikipedia.org/wiki/Enterprise_architecture) that allows interoperability and information sharing between semiautonomous, noncentrally organized lines of business (LOBs), IT systems and applications.**

Within a federated implementation model, the corporate BI team would open up its data warehousing environment to all of the divisions within the enterprise. Each division or LOB would have its own dedicated partition in the EDW or data lake to develop its own data marts, 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 BI expertise, the corporate BI team would continue to build custom data marts, reports and dashboards as before.

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

- Standardized scalable architecture
- Support for an extensible physical and logical data model
- A centralized data repository, with dedicated partitions and zones for individual LOBs
- Consistent data definitions
- The agility to deliver multiple end-user analytics products and services
- A stronger partnership between IT and business
- True support for a self-service-governed analytics platform

This model comes with a unique set of challenges —principally, building consensus within the organization to change user behavior, and gaining the support to fund and build the necessary capabilities for an architecture that centralizes data but decentralizes analytics. For more information about the federated BI architecture, see "Create a Data Reference Architecture to Enable Self-Service BI." (https://www.gartner.com/document/code/333398?ref=grbody&refval=3891182)

Key points to consider when addressing this priority include the following:

- Whenever analytics efforts are pulled under the umbrella of the IT organization, IT governance, "ownership" and management of the operationalized data systems and frameworks will need to be addressed. These issues will need to be resolved under the organization's established IT governance processes and frameworks.
- On an ongoing basis, data professionals in the IT organization should implement reviews to capture, automate and repurpose analytical insights from all sides of the organization. As part of this effort, they should maintain an inventory of people, projects and capabilities associated with ad hoc analytics initiatives underway in the enterprise.
- The emerging position of data engineer can have an important role to play in this area. To support analytics initiatives and use cases that occur outside of IT, this person works and collaborates across business boundaries to facilitate the extraction of information from systems. This role can also help facilitate the collaboration across business boundaries needed to identify and integrate fragmented analytics initiatives.

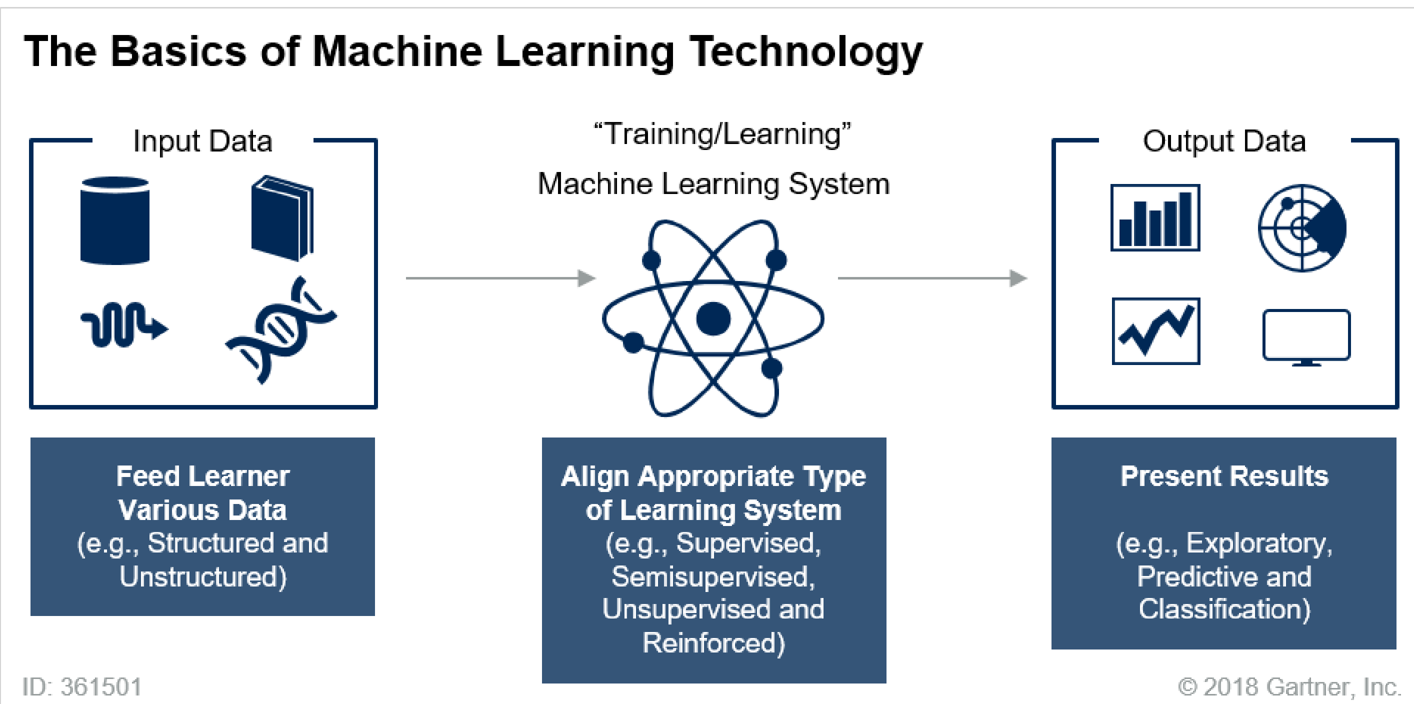
Prepare for the Onrush of Machine Learning

For many data and analytics technical professionals, advanced analytics and ML techniques are a mystery. However, the immense volume, variety and velocity of data available today are fueling new demands from the business. Thus, ML and algorithms must soon become part of the knowledge base of these professionals.

> **Machine learning is not just for data scientists; it is a tool of the trade for digital architects.**

The ML concept is a simple, data-driven one: Algorithms can be trained and learn from data without being explicitly programmed. ML techniques are based on statistics and mathematics, which are rarely part of traditional data analysis. Any type of data is input, learning occurs and results are output. In supervised learning, known sample outcomes are used for training to achieve desired results. Unsupervised learning relies on ML algorithms to determine the answers (see Figure 10).

Figure 10. The Basics of Machine Learning Technology



Source: Gartner (October 2018)

To prepare for the increasingly important role of ML in their future, data and analytics technical professionals should start with the basics, and learn by doing. Steps include:

- **Define a business challenge to solve:** The challenge may be either of the following:
  - *Exploratory* (for example, determining what factors contribute to a consumer's default on a bank loan)

- *Predictive* (for example, predicting when the next natural gas leak will occur and what factors will drive the next failure)

Start small, and build in stages.

- **Partner with the data science team:** Work with this team to deliver a processing environment for the data needed to address the defined business challenge. Enable a platform that will scale to execute the required models and algorithms. This environment might be cloud-based. With the growing demand to support analytics by leveraging ML, there is a need to think outside of endpoint reporting solutions. Instead, an analytics development life cycle should be built, where the ML models are monitored and constantly fine-tuned for optimal performance.
- **Get trained now:** Before you act, you must learn. Several online courses offer good basic knowledge on the mechanics of ML. Two examples worth reviewing are Coursera's Machine Learning (https://www.coursera.org/learn/machine-learning) and Udacity's Intro to Machine Learning. (https://www.udacity.com/course/intro-to-machine-learning--ud120) For a primer on the benefits and pitfalls of ML, the requirements of its architecture, and the steps to get started, see "Preparing and Architecting for Machine Learning: 2018 Update." (https://www.gartner.com/document/code/365935?ref=grbody&refval=3891182)
- **Form a team of experts:** Think about putting together a team of experts to tackle data science and ML challenges. The business domain expert can work with a data scientist or citizen data scientist, who can leverage ML as a service (MLaaS) platforms to build and train ML models. The DevOps platform engineer can help support the underlying infrastructure, and the analytics and app developer can assist with integrating the models within an application or a BI platform to bring the value of ML to business.

Enhance Application Integration Skills to Embed Analytics Everywhere

Data and analytics systems are often architected and developed in parallel with systems that capture and process data. While these systems are logically connected, they are usually physically separated. For data and analytics to be delivered at the optimal point of impact, monolithic analytics systems must be architected and decomposed into callable services that can be integrated wherever they are needed.

In a mix-and-match world, components must be architected in a more modular way, using features such as:

- Standard data model and transport protocols to locate and retrieve the right data, be it on-premises or in the cloud

- ML algorithms that can be developed in the R environment, and then executed within a Python program or another analytics tool
- Visualization widgets (for example, components offered by D3.js) that deliver information in the optimal format based on the calling device (web or mobile)
- Data services to deliver raw data to analytics processes via RESTful APIs
- Data virtualization to provide standard access to a wide variety of data

Whether you integrate using a commercial BI and analytics platform or an open-source option, pay particular attention to the provider's API granularity. The finer-grained the services are, the more flexibility you will have.

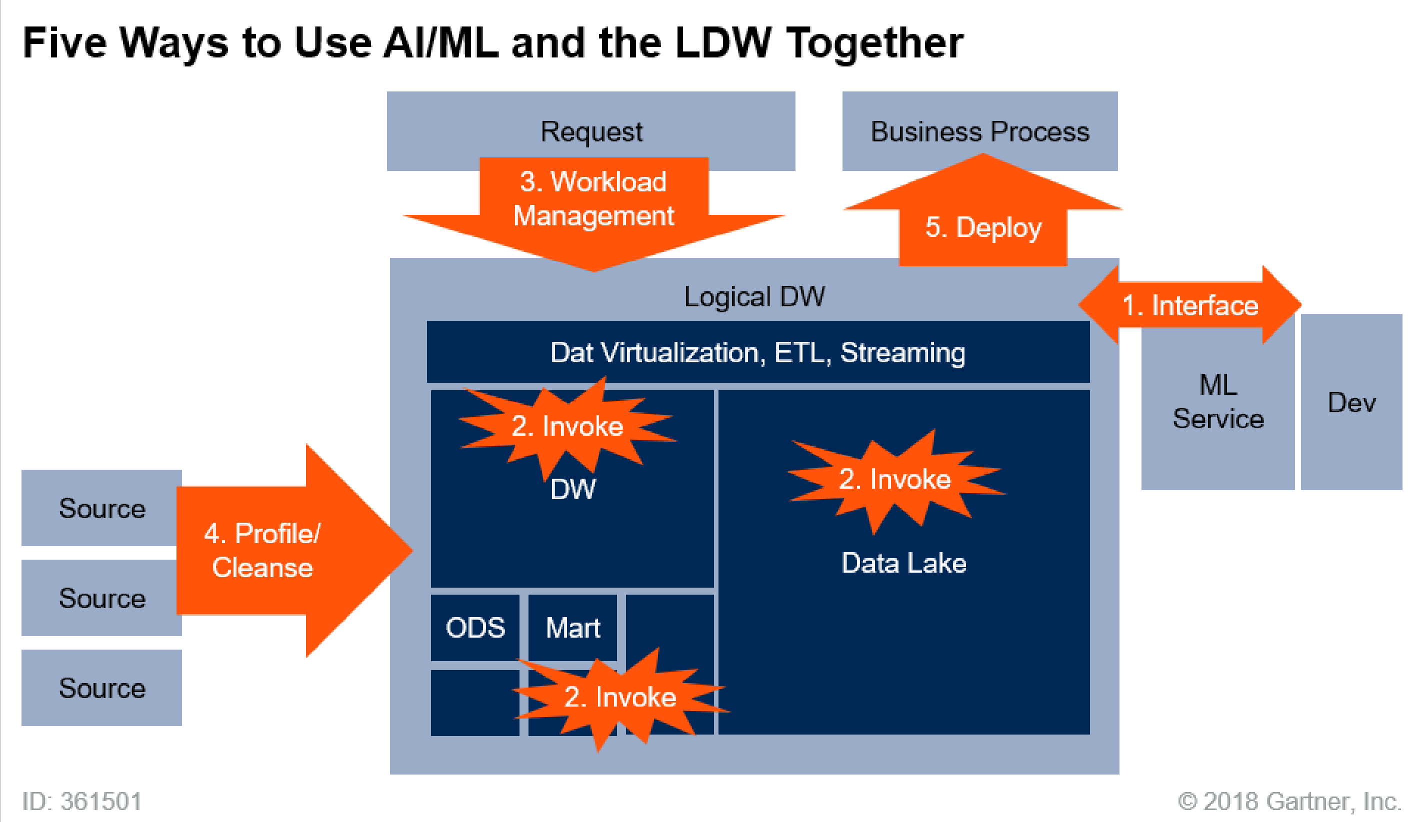
AI and ML Will Generate New Synergies in Information Management

AI and ML collectively is often thought of as a strategy for driving business decisions and rarely in the context of solving information management challenges. Going into 2019, there are several practical use cases where AI and ML can be applied to solve information management challenges. These use cases demonstrate the value of applying AI and ML to different components of the data and analytics architecture to improve overall operations. AI and ML should no longer be viewed as an independent initiative, but as a complementary strategy to improve information management strategies.

One useful example of this synergy is the use of AI/ML to complement the LDW. Technical professionals who design and develop analytics systems often see these two types of system as being entirely separate. Practitioners of one type of system may not be familiar with the other.

There are five ways in which the combination of the LDW and AI/ML continues to support modern information management requirements — such as to make available necessary data to the enterprise, as outlined in Figure 11.

Figure 11. The Five Ways in Which AI/ML and the LDW Can Help Each Other



Source: Gartner (October 2018)

AI/ML and the LDW have a symbiotic relationship in these five ways:

1. The LDW can be interfaced to an AI/ML service, where users query the LDW using AI/ML technologies.
2. AI/ML routines can be invoked directly from within the component engines of the LDW.
3. AI/ML can be used to help manage the complex workload of the LDW.
4. Understanding the structure and content of the data being input into the LDW is very important — and AI/ML can assist. This is one of the most exciting areas of the market today.
5. AI/ML can leverage LDW infrastructure to deploy its models into production.

Planning Considerations

Look for Opportunities to Use AI/ML and the LDW in Combination

The LDW can provide reliable data to AI/ML, large computing and data storage resources, and a reliable means of deploying models. Equally, AI/ML can inform the LDW in its data ingestion, workload management, in addition to adding to its portfolio of analytical techniques.

Technical professionals should take stock of all the ML routines available in their incumbent software. Most commercial database management system (DBMS) software has useful libraries of the most popular ML algorithms. Where new analytical requirements can be met by common ML algorithms, these incumbent libraries provide a simple and low-cost means of meeting analytics requirements.

Look for Synergies in Data Quality Work Between DW Staging and ML

In many development environments, practitioners treat data quality processing for the data warehouse, data lake or other components of the LDW separately from the AI/ML work. However, practitioners can do much of this work in common with data warehouse processing, using the industrial strength tools used for the major data platforms.

One of the main missions of the data warehouse is to provide quality-assured data to all its users. It does this by gaining consensus from its users on what data quality means, and then applies extraction, transformation and loading (ETL) and quality routines to check and enforce this. The result is that users can take data from the warehouse and be assured that it is well-understood and reliable.

AI/ML clearly needs good quality data as input. If the data cannot be relied on, either during training or execution, then the results themselves will be unreliable.

Therefore, technical professionals should look for opportunities to use the industrial strength data transformation and quality tools used for the LDW for machine learning. This is especially true where modern data crawling tools can automatically discover data content and create metadata.

Consider How Components Can Support Each Other

The aim is to have a wide enough variety of servers to meet any and all requirements. But it’s also to have the minimum number of servers to avoid unnecessary overhead. Technical professionals need to integrate those servers so that all of their resources can be coordinated in meeting new business requirements.

The aim is to have a small number of large servers that together can meet any and all requirements. If the architect needs to supplement the data in the core LDW, then data virtualization is a good way to do that.

Position LDW-Enabled AI/ML by Speed and Ease of Development, Scalability and Cost

Every requirement should have two estimates attached to it: cost and potential benefit. These should determine which requirement is highest priority and also ensure development is tracking maximum return on investment for the LDW.

This determines what data should be loaded and which new processing performed. This in turn determines how the system expands what parts of the analytics landscape. If we have an LDW, then this is simple; there is only a limited number of components, and their purpose is well-understood.

Assess Components for Ease of Integration

Different components may have different ML capabilities, and they also present different interfacing capabilities. Therefore, it is useful to explore which ML libraries are available in each component. Also, we can assess how easily the system interfaces one component with another.

If you can develop the model in a DBMS or data management solution for analytics (DMSA), and invoke it in the same component, then this is the simplest case.

Alternatively, if one component can do analysis and emit Predictive Model Markup Language (PMML), the technology-independent analytics description language, and another can consume it, then that is ideal.

Alternatively, components may be sources and sinks of data for each other. This might not be the deciding factor for choosing components, but it is a factor that may well influence which components make up your analytical landscape.

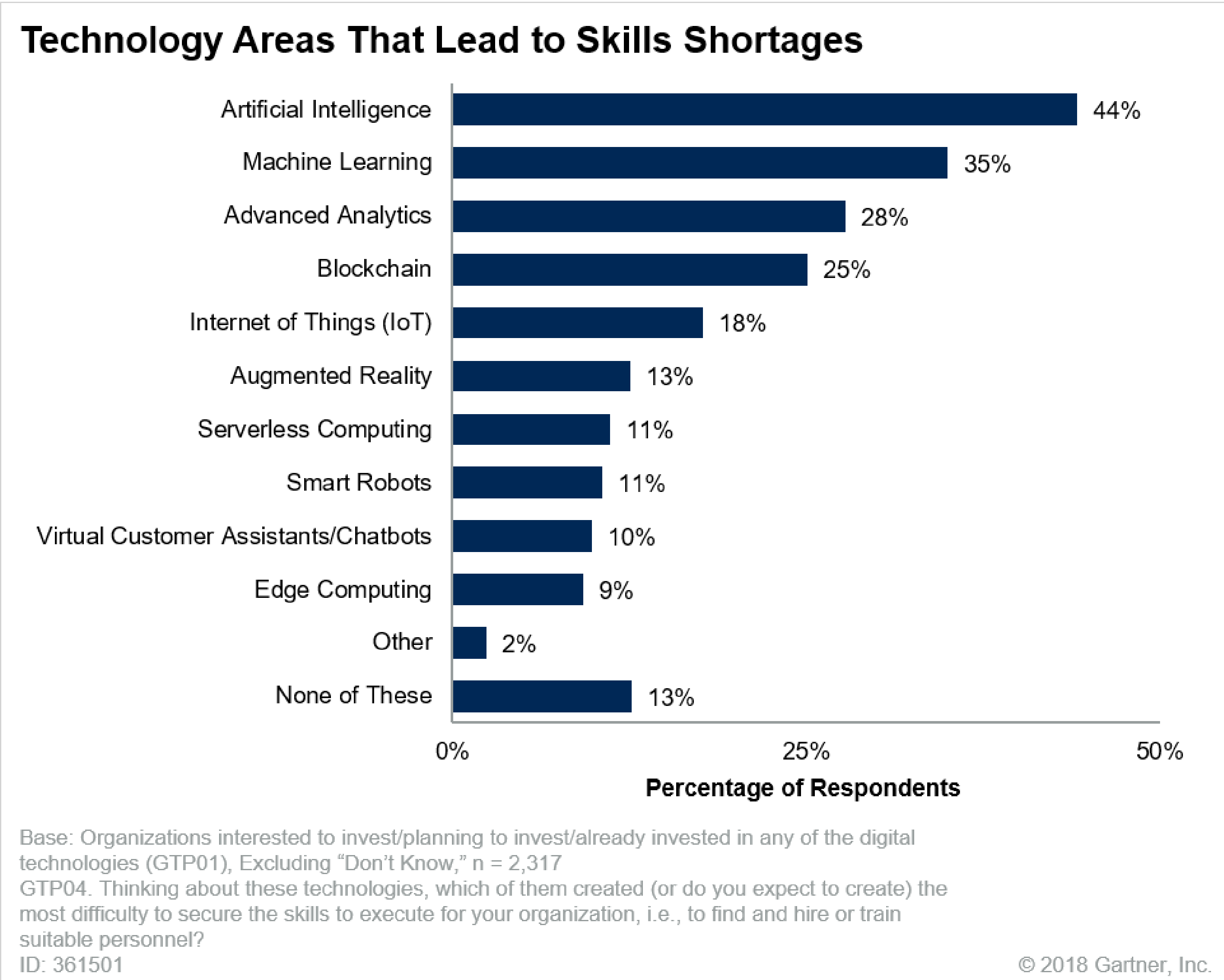
Thinking in advance about how easily (or not) data sharing and ML capabilities can be distributed over the different system components can make major differences to implementation effectiveness and cost.

Analytics Services in the Cloud Will Continue to Accelerate to Deliver Greater Performance at Scale

Over the past four years, Gartner has seen a steady increase in adoption of and inquiries about cloud computing and data storage — both for operational and analytical data. Much of this interest can be attributed to cloud-native applications (such as Salesforce and Workday), emerging IoT platforms, AI and ML services, and externally generated data born in the cloud. However, an increasing number of organizations are making a strategic push to incorporate the cloud into *all* aspects of their IT compute and storage infrastructure.

The scale and capacity of the public cloud — coupled with increasing business demand to gather as much data as possible, from as many different sources as possible — are forcing the cloud into the middle of many data and analytics architectures. The data “center of gravity” is rapidly shifting toward the cloud. As more data moves to the cloud, analytics has already followed. Reflecting this trend, both cloud computing and analytics are front and center in the minds of architects and other technical professionals. In a Gartner survey of IT professionals, [2] artificial intelligence and ML were the top technology areas cited by respondents as talent gaps they needed to fill (see Figure 12). Those skill gaps will also need to consider cloud principles, as more AI and ML services are being offered in the cloud. Increasingly, this will result in the need for more core cloud skills, in addition to AI and ML skills, to be adopted by technical professionals.

Figure 12. Top Technology Talent Gaps Identified by Technical Professionals



Maximum of three responses allowed.

Source: Gartner (October 2018)

Cloud is already fundamentally impacting the end-to-end architecture for data and analytics. Technology related to each stage of the data and analytics continuum — acquire, organize, analyze and deliver — can be deployed in the cloud or on-premises. Data and analytics can also be deployed using "hybrid" combinations of both cloud and on-premises technologies and data stores.

Three foundational cloud competencies include:

- **Integration:** Integration involves bringing multiple cloud services together with on-premises infrastructure, and making them work together to deliver an integrated result. Such capabilities will include integration of cloud endpoints, governance, community management and migration skills to and from public and private clouds, colocation facilities, and on-premises, distributed infrastructure.

- **Customization:** — Customization is altering or adding to the capabilities of a cloud or on-premises service to perform its function and deliver a business-facing service. This may be incorporating new data and process functions, visibility and analytics, or generating a new look and feel to the service. Customization will be required as IT organizations change the people, processes and technologies to make hybrid clouds work for IT customers.

- **Aggregation:** Multiple services come together at a cloud scale. These may include provisioning, single sign-on (SSO), simplified billing, unification of disparate management platforms, facilitating access to cloud services, customer support and SLA management.

Data and analytics technical professionals will increasingly need to develop these core cloud technical competencies, as they also grapple with the technical challenges inherent in using the building blocks supplied by cloud providers to build cloud data architectures. They will then need to deliver a seamless experience to end users. At the same time, they may find themselves drawn into the other aspects of cloud provider management, which include activities such as cost management, vendor management, risk management and consensus building with stakeholders.

Gartner expects such hybrid IT approaches and deployments to be a reality of most IT environments in 2019 and beyond. Even with rapid adoption of cloud databases, integration services and analytics tools, enterprises will have to maintain traditional, on-premises databases. The key to success will be to manage all of the integrations and interdependencies while adopting cloud databases to deliver new capabilities for the business. This will make for a potentially complex architecture in the near term, as data and analytics continue their inexorable march into the cloud.

Planning Considerations

As they incorporate the cloud into data and analytics, technical professionals need to focus on long-term objectives, coupled with near-term actions, to flesh out the right approach for their organization. Planning considerations for 2019 should include the following actions:

- Start developing a cloud-first strategy for data, followed by analytics.
- Determine the right database services for your needs.
- Adopt a use-case-driven approach to cloud business analytics.
- Model cloud data and analytics costs carefully based on anticipated workloads.

Start Developing a Cloud-First Strategy for Data, Followed by Analytics

Public cloud services, such as Amazon Web Services (AWS), Google Cloud Platform and Microsoft Azure, are innovation juggernauts that offer highly operating-cost-competitive alternatives to traditional, on-premises hosting environments. Cloud databases are now essential for emerging digital business use cases, next-generation applications and initiatives such as IoT. Gartner recommends that enterprises make cloud databases the preferred deployment model for new business processes, workloads and applications. As such, architects and other technical professionals should start building a cloud-first data strategy now, if they haven't done so already.

This team should also develop a strategy for how the cloud will be used in analytics deployments. Data gravity, latency and governance are the major determinants that will influence when to consider deploying analytics to the cloud, and analytics-focused database services for the cloud are numerous. For example, if streaming data is processed in the cloud, it makes sense to deploy analytics capabilities there as well. If application data is resident in the cloud, you should strongly consider deploying BI and analytics as close to the data as possible. Additionally, cloud-born data sources from outside the enterprise will take on an increasingly important role in any data and analytics architecture.

Determine the Right Database Services for Your Needs

Depending on which cloud service provider you choose, many database options may be available to you. AWS and Microsoft Azure offer comprehensive suites of cloud-based analytics databases. Determining which database services to use is a key priority. It is important to understand the ideal usage patterns of each possible option. Matching the right technology to a specific use case is critical to success when using these products. This may lead you to choose different database services for unique workloads.

For example, AWS offers several standard services that are broadly characterized as operational or analytical for structured or unstructured data. (For more information, see "Evaluating the Operational Databases on Amazon Web Services" (https://www.gartner.com/document/code/346633?ref=grbody&refval=3891182)) You may use one service for transaction processing and another for analytics. One service does not have to fit all use cases.

This same model holds true for other cloud providers. For example, Microsoft offers Azure SQL Database for operational needs and Azure SQL Data Warehouse for analytics, among other offerings. (For more information, see "Evaluating the Operational Databases on Microsoft Azure." (https://www.gartner.com/document/code/346634?ref=grbody&refval=3891182)) In addition, a database service from an independent vendor can be run in the cloud, either by licensing it through a marketplace or by bringing your own license.

Adopt a Use-Case-Driven Approach to Cloud Business Analytics

As the data and analytics environment moves into the cloud, it is reasonable to expect the business analytics environment to follow. Cloud analytics and BI applications and deployments are growing. Cloud BI continues to grow in adoption. Data gravity, latency and governance, as well as the use cases supported, are important factors in determining when and how to deploy BI and analytics in the cloud. Another factor also weighs heavily, however — the reuse of existing functionality. In fact, this is the No. 1 concern raised in Gartner client inquiries about business analytics. There continue to be areas, such as enterprise reporting, where cloud analytics BI platforms are not building enterprise-ready capabilities.

For this reason, be wary of initiatives that seek to “standardize” on as few business analytics tools as possible. An optimal federated BI architecture should be able to support many different business analytics tools by centralizing the datasets at the layer of the data warehouse as much as possible, so that new visualization tools can be deployed easily.

Gartner has identified seven criteria that should be evaluated to help determine whether analytics use cases should be deployed to the cloud (see Table 1).

Table 1: Seven Criteria for Determining a Cloud Analytics Architecture

Criteria ↓	Essential Questions ↓	Decision Support ↓
Data Gravity	Where is the current center of gravity for data?	If the answer is “in the cloud,” a cloud BI product is most likely going to be a good fit for the organization.
Data Latency	How fresh does the data need to be?	In scenarios where extremely low latency is desired, the speed of light dictates that closer is better, which may drive investments in edge, IoT and on-premises analytics, rather than a cloud-only model.
Governance	How much governance is required based on domains and use cases?	Cloud BI platforms lag in governance capabilities in comparison to mature on-premises analytics and BI platforms, although there is improvement in this space.
Skills	What skills, tools and platforms are available in your organization?	Cloud BI platforms encourage end-user adoption by minimizing the barriers to entry and enabling a large amount of flexibility, as well as, in some cases, licensing flexibility. On the flip side, cloud BI integration can be extremely difficult due to the inflexibility of proprietary and aging data warehouse and data analytics architectures that most organizations continue to have to maintain.

Criteria	Essential Questions	Decision Support
Agility	How quickly must new requirements/components be added/updated?	Cloud BI adds agility, as new features, but because the infrastructure is controlled by the cloud provider, this agility is limited to the capabilities they let you have, compared to on-premises BI platforms, which often provide more customizability (albeit with a hefty development effort).
Functionality	Are certain functions available only in the cloud or only on-premises?	It is extremely common for cloud BI products to have the best features available only in the cloud, with a pared-back set of features on-premises, leading to difficult compromises for hybrid deployments.
Reuse	How much existing investment do you want to carry forward from your on-premises analytics platform?	Just as with on-premises BI products, there is usually no easy way to migrate existing dashboards and reports to any different BI platform, cloud or on-premises. Starting anew is usually the approach that is taken.

Source: Gartner (October 2018)

Model Cloud Data and Analytics Costs Carefully Based on Anticipated Workloads

The cost model for cloud data and analytics is completely different from on-premises chargeback models. Pricing constructs vary considerably among analytics vendors, with several offering cloud services both directly and through major marketplaces. Factors such as data volumes, transfer rates, processing power and service uptime will impact monthly charges. Use-case evaluations should include the goal of avoiding unexpected costs in the future.

Tools are available to help track and manage cloud costs. For more information, see "Comparing Tools to Track Spend and Control Costs in the Public Cloud." (https://www.gartner.com/document/code/323083?ref=grbody&refval=3891182)

Revolutionary Changes in Analytics Will Drive IT to Adopt New Technologies and Roles

The data and analytics domain is rapidly expanding, and new technologies are challenging established practices. The convergence of several factors is driving a "perfect storm" for technical professionals tasked with managing data and analytics. These factors include:

- Higher volumes of data from an ever-expanding variety of data sources

- Data ingestion and processing performed on-premises and in multiple cloud vendors
- Access to cloud-based, hyperscale compute, processing and storage capabilities
- Advances in computer vision, natural language processing and pattern recognition
- Greater embrace of crowdsourcing techniques to leverage human intelligence in automation, such as for data quality, ML and other use cases
- Real-time streaming and analytics-processing frameworks
- The evolution of hybrid analytical and transactional architectures
- Expanding IoT use cases, such as ML at the edge, and integration with third-party data, historical data and metadata
- Managed services, serverless architectures and infrastructure as code
- The emergence of the trust economy and supporting technologies such as blockchain
- Advances in bots and robotic process automation, making it possible to embed analytics into, or to supplant, human interactions and processes
- Increased oversight from governance and compliance bodies

The mandate to deploy AI and ML technologies often starts at the board level and filters down the organization, often without identifying business objectives. Technical professionals can help make these deployments more effective by developing the foundational components needed to support AI and ML in the enterprise. For more information, see "Laying the Foundation for Artificial Intelligence and Machine Learning: A Gartner Trend Insight Report." (https://www.gartner.com/document/code/373110?ref=grbody&refval=3891182)

Technical professionals should focus to centralize the scarcest skills — those of the data scientists — into a center of excellence in order to have a critical mass. Data scientists should have dotted-line reporting to lines of business. By being closer to the consumers of analytics, the data scientists can understand requirements better. They would also be in a better position to ensure a smooth hand-off of their models.

An alternate source of analytics platforms and models is emerging. Cloud vendors, starting with Amazon, Microsoft, Google and IBM, are building pretrained models that developers can customize without having any training in data science. Today, those models mostly focus on so-called "cognitive" processing, such as image processing and natural language processing. These vendors all have plans to add similarly customizable models for operational processes such as demand forecasting, fraud detection and predictive maintenance, among others. Gartner sees this class of AI and ML technology growing rapidly as vendors and customers look to empower the tens of millions of traditional developers.

The rapid advancements in data and analytics act as both a boon and a challenge to technical professionals. These trends offer tremendous promise and raise awareness of the important work that technical professionals do. However, they also necessitate new skills, architectures and patterns, and challenge traditional methods and processes.

Consider AI and ML. While there's no question that these technologies will play a significant role in the future, the growing hype surrounding them is driving an insatiable appetite for analytics in the enterprise and expanding notions about what's possible today. Business leaders are besieged with seductive catch phrases like "kinetic business" and "real-time decisioning." The promise of AI and ML is putting technical professionals in a difficult position. They must manage lofty expectations while simultaneously expanding their technologies and skills to prepare for the future.

Enterprises are having a difficult time strategizing about how to approach new AI projects — partly because of the complexity of the technology, talent gaps, and the lack of proven use cases and standards.

Within the broad scope of AI, technical professionals should familiarize themselves with the most common disciplines:

- Language processing, including translation, speech recognition, sentiment analysis and conversational platforms
- Computer vision, including facial recognition, gesture recognition and optical pattern recognition
- Machine learning, including IoT, sensors, deep learning and model algorithms

This will help identify tangible use cases within your domain. We also suggest that companies look for quick wins with existing data and analytics capabilities. Start with the data you have, and begin to build on that to create small differentiations. And remember to add domain-specific knowledge to cement the win.

Similar to ML, we see a similar team structure evolving with AI initiatives where there is a dearth of AI professionals and expertise. Leverage existing roles and skill sets of an application developer, a platform engineer and perhaps a data scientist to explore business process automation processes by building AI-based systems.

With the increasing use of crowdsourcing techniques for data management and analytics tasks, as well, technical professionals are interfacing with crowds of human workers integrated into the data workflow. Knowing how to design appropriate and clear microtasks, as well as knowing some psychological and sociological basics for motivating crowds and managing community dynamics, becomes new skills technical professionals must master.

Planning Considerations

IT organizations must adapt their methods, roles and skills to demonstrate agility and exert influence over their organization's analytics strategy. The role of the data management and analytics professional has never been more crucial.

Planning considerations for 2019 include the need to:

- Focus on new and emerging architectural, technical and product management roles.
- Devote time to enhancing technical and professional effectiveness skills.

Focus on New and Emerging Architectural, Technical and Product Management Roles

Opportunities will emerge for technical professionals to play new roles. These roles will help their enterprises exploit data and analytics technologies to improve and transform their businesses. Some roles, such as data architects and analytics architects, may already exist in the organization. These roles will have significant input in designing and developing the end-to-end data and analytics architecture discussed earlier, and will become more vital in 2019.

The data engineer — a role often linked with data science — designs, builds and integrates data stores from diverse sources to smooth the path for ever-more-complex analysis. It is a natural progression from the data integration specialist, and will become an essential part of any data science effort that furthers predictive, prescriptive and ML analytics efforts. Data engineer responsibilities include:

- Preparing data for use in data science projects
- Assisting with initial data exploration steps (binning, pivoting, summarizing and finding correlations, for example)
- Cataloging existing data sources and enabling access to resident and external data sources
- Supporting data stewards to establish and enforce guidelines for data collection, integration and processes

The IoT architect is another key role — one that will become critical for every IoT initiative. IoT encompasses a broad set of technical and nontechnical topics that include embedded systems, cloud computing, software development, security, data management and system engineering. The IoT solution architect must collaborate with business leaders and partner with IT management to address each of these topics (see "Solution Path for Developing an Internet of Things Technical Strategy" (https://www.gartner.com/document/code/354613?ref=grbody&refval=3891182)).

The AI and ML architect and engineer are rapidly emerging as new roles within teams pursuing AI and ML initiatives. These new roles focus on integrating and implementing AI and ML systems into new or existing business systems. The ML engineer is a key role in operationalizing and optimizing ML products and services — a responsibility that will continue to rise in demand as more AI/ML projects move away from experimentation and into operations.

We also expect new teams to appear, most likely in the form of transformation teams or centers of excellence. These teams will emphasize refinement, efficiency and ongoing improvement as data and analytics activities work their way into the fabric of the organization's processes and capabilities.

In addition, as more data and analytics services become outward-facing to connect ecosystems and to monetize data to external constituents, architects or other technical professional functions may also take on the role of "product manager." This role sits at the intersection of business, technology and user experience. Although this is a long-established role in the software vendor and OEM marketplaces, product managers are starting to appear with greater frequency in many other organizations. This position occupies a unique role in an organization, with responsibilities that include:

- Researching market needs and customer preferences
- Setting the vision for the product, and selling that vision to the rest of the organization
- Defining and prioritizing the business outcomes required to attain the vision
- Obtaining the resources needed to build and sustain the product
- Working with development teams to translate the targeted business outcomes into features
- Working across the organization — with stakeholders, users, development teams and operations — to ensure product success
- Working with sales, marketing, ecosystem partners and customers on products aimed at external customers

Any data products created for external consumption should have a product manager. This role is needed to ensure that the organization delivers the right products to the right markets at the right time. This role is not limited to external data products, however. The product management discipline is also a great addition for internally facing data and analytics products. See "Moving From Project to Products Requires a Product Manager" (https://www.gartner.com/document/code/289817?ref=grbody&refval=3891182) for more information.

> **Existing roles, such as the project manager or the scrum product owner, are neither appropriate nor sufficient for managing a significant product that requires many teams to build.** [3]

Devote Time to Enhancing Technical and Professional Effectiveness Skills

To capitalize on emerging opportunities, it is important to develop a broad range of technical and professional effectiveness skills. Although technical skills are a minimum requirement, professional effectiveness skills can make or break your success in any project or program you work on. Gartner has long advocated that technical professionals supplement their technical capabilities with additional "soft skills," such as the ability to:

- Better understand business goals and scenarios to help build business cases
- Critically think through problem resolution
- Articulate points of view in the language of the business audience

With the emergence of new, increasingly business-related and customer-facing roles in IT, communication skills and business acumen are more important than ever. When Gartner asked nearly 950 technical professionals where they saw skills gaps, three of the top 10 responses were related to professional effectiveness skills (critical thinking/problem solving, business acumen/knowledge and communication skills). [4]

Effectiveness skills without requisite technical prowess are only half of the story. New trends in data and analytics will require technical professionals to enhance their technical expertise in:

- Cloud technology
- Advanced analytics and ML
- Data virtualization and the LDW
- Streaming ingestion
- Real-time data movement and processing
- Integration capabilities to incorporate data and analytics everywhere

Take the following steps to improve your technical and professional effectiveness skill sets:

- Identify the skills you need to improve. Ask others you work with for their opinions.
- Research whether employee development and technical training programs are in place in your organization. HR often has relevant courses and programs available.
- Look to external resources if programs aren't available internally. For communication training, turn to vendors such as Toastmasters International. Explore course work at local universities or online (for example, Coursera). Depending on cost, determine whether your organization will assist you in these efforts.
- Spend time putting what you learn into practice. Make this new knowledge part of your standard operating procedure.
- Take personal responsibility for this improvement. It will not only benefit your company, but will also serve you well in any future endeavor.

Setting Priorities

Data and analytics technical professionals must focus on the following areas as they plan and prioritize their activities in 2019:

- **Design for "hybrid multicloud."** An increasing number of organizations are finding that they have to support applications that are on-premises as well as in one or more cloud service providers. This adds new sets of design challenges and also introduces a variety of integration options. Technical professionals should carefully architect data management aspects to minimize data latency as well as costs associated with unnecessary data ingress and egress.
- **Adhere to data governance and compliance requirements.** A host of existing and new compliance guidelines are forcing technical professionals to improve data hygiene as well as enhance data security and protection. In 2018, the EU General Data Protection Regulation (GDPR) came into enforcement. Also in 2018, the California government released the California Consumer Privacy Act (CCPA). This may be the beginning of many state regulations for which technical professionals need to plan. For instance, the EU GDPR's Article 25 mandates that data protections must be designed in the core architecture. New technologies such as data catalogs allow technical professionals to detect sensitive information and apply corporate policies.
- **Enable greater self-service and automation.** Technical professionals should prepare for deploying solutions that will simplify design, management, operations and the use of the architecture and applications. For example, technical professionals are increasingly deploying container technologies, and serverless and microservices architectures. Their goal is to increase agility and flexibility while enabling the business users to deploy self-service applications quickly across multiple on-premises and cloud infrastructures.

- **Design and build a comprehensive end-to-end architecture.** IT and the business should work together to design an end-to-end architecture for data and analytics. Technical professionals should start with the business goals in mind and holistically manage an architecture to support those outcomes. The four phases — acquire, organize, analyze and deliver — must be planned together, with each feeding off the others. Data, analysis and action can no longer represent separate disciplines; they must be fused into a cohesive plan of attack. Organizations are starting to add additional workloads and users to the systems. While in the past, the goal may have been to support a small number of data scientists, now the goal is support an army of business and data analysts.

- **Enable analytics to become pervasive within and outside the enterprise.** With more people wanting to engage with data, demand for analytics will continue to expand. It's critical to be prepared for more business user enablement by fostering a pragmatic approach to better self-service, coupled with processes to prioritize, facilitate and manage the proliferation. ML is rising quickly, and technical professionals need to understand the concepts, experiment with the technologies and integrate analytics wherever they are needed for optimal impact.

- **Incorporate the cloud as a core element of the organization's data and analytics architecture.** The cloud needs to become part — or most likely the centerpiece — of the organization's data and analytics architecture. Developing a cloud-first strategy for data, followed by analytics, is an essential first step. Choosing the right cloud service providers and technologies should follow. With many possible services available, technical professionals may select a mix-and-match approach for data and analytics as they gradually migrate data storage and computing capabilities to the cloud. In addition, technical professionals should exploit data marketplaces as much as possible to procure their cloud data and analytics services.

- **Expand roles and skill sets to deliver data service products for internal and external business ecosystems.** With chief data officers striving to increase business value, data and analytics products are being evaluated and designed for internal and external business ecosystem consumption. As internal projects turn into external products, new roles for technical professionals will emerge. Because solid technical and professional effectiveness skills are important components of these architect, engineer and product manager roles, it's important to devote time and effort to improving these capabilities.

- **Adopt AI and ML to enable process improvement, operational efficiency and automated actions.** AI and ML technologies will play an increasingly significant role in the future, driven by an insatiable appetite for analytics in the enterprise and by expanding notions about what's possible today. Technical professionals must manage lofty expectations regarding what's possible with AI and ML today, while simultaneously expanding their technologies and skills to prepare for the future.

Data and analytics technical professionals should begin by taking an inventory of their existing environments. All of the planning considerations discussed in this report should be approached as part of an evolution to a strategic end state, not as a rip-and-replace strategy. Some actions will move faster than others.

Technical professionals must ultimately keep the end goal in mind. It is easy to become enamored of new technology choices, but business value must be front and center in every decision. Maintain open channels of communication with constituents — both internal and external — and explain any technical actions or concepts in terms they can understand, support and champion. In this exciting time for data and analytics, technical professionals can play an increasingly critical role in helping their organizations achieve business success.

Evidence

[1] 2019 Gartner CIO Survey. The 2019 Gartner CIO Survey was conducted online from 17 April through 22 June 2018 among Gartner Executive Programs members and other CIOs. Qualified respondents are each the most senior IT leader (CIO) for their overall organization or a part of their organization (for example, a business unit or region). The total sample is 3,102, with representation from all geographies and industry sectors (public and private). The survey was developed collaboratively by a team of Gartner analysts, and was reviewed, tested and administered by Gartner's Research Data and Analytics team.

[2] The Gartner Technical Professionals Study was conducted online from 30 January 2018 to 2 March 2018 among 2,468 respondents in North America, EMEA, Asia/Pacific and Latin America. A subset of *Gartner for Technical Professionals* seatholders were invited to participate. In addition, Gartner *IT Leaders* seatholders with the job level of "associate" were invited to participate. Respondents were required to be a member of their organization's IT staff or department (or serve in an IT function). Furthermore, they could not serve as a member of the board, president or in an executive-level or IT leadership position. The survey was developed collaboratively by a team of Gartner analysts who follow technical professionals and was reviewed, tested and administered by Gartner's Research Data Analytics team.

[3] "Moving From Project to Products Requires a Product Manager" (https://www.gartner.com/document/code/289817?ref=grbody&refval=3891182)

[4] "Top Skills for IT's Future: Cloud, Analytics, Mobility and Security" (https://www.gartner.com/document/code/297698?ref=grbody&refval=3891182)

Document Revision History

2018 Planning Guide for Data and Analytics - 29 September 2017 (https://www.gartner.com/document/code/331851?ref=ddrec)

2017 Planning Guide for Data and Analytics - 13 October 2016
(https://www.gartner.com/document/code/311517?ref=ddrec)

2016 Planning Guide for Data Management and Analytics - 2 October 2015
(https://www.gartner.com/document/code/290775?ref=ddrec)

Recommended by the Authors

Solution Path for Implementing a Comprehensive Architecture for Data and Analytics Strategies
(https://www.gartner.com/document/3880568?ref=ddrec&refval=3891182)

Solution Path for Planning and Implementing the Logical Data Warehouse
(https://www.gartner.com/document/3719217?ref=ddrec&refval=3891182)

Comparing Cloud Data Warehouses: Amazon Redshift and Microsoft Azure SQL Data Warehouse
(https://www.gartner.com/document/3507817?ref=ddrec&refval=3891182)

Migrating Enterprise Databases and Data to the Cloud
(https://www.gartner.com/document/3671418?ref=ddrec&refval=3891182)

A Comparison of Master Data Management Implementation Styles
(https://www.gartner.com/document/3824466?ref=ddrec&refval=3891182)

Implement Agile Database Development to Support Your Continuous Delivery Initiative
(https://www.gartner.com/document/3559517?ref=ddrec&refval=3891182)

Top Skills for IT's Future: Cloud, Analytics, Mobility and Security
(https://www.gartner.com/document/3353917?ref=ddrec&refval=3891182)

EIM 1.0: Setting Up Enterprise Information Management and Governance
(https://www.gartner.com/document/3831267?ref=ddrec&refval=3891182)

Use Design Patterns to Increase the Value of Your Data Lake
(https://www.gartner.com/document/3876783?ref=ddrec&refval=3891182)

Deploying Effective iPaaS Solutions for Data Integration
(https://www.gartner.com/document/3714226?ref=ddrec&refval=3891182)

Evaluating the Operational Databases on Microsoft Azure
(https://www.gartner.com/document/3878365?ref=ddrec&refval=3891182)

Evaluating the Operational Databases on Amazon Web Services
(https://www.gartner.com/document/3878322?ref=ddrec&refval=3891182)

Recommended For You

Solution Path for Implementing a Comprehensive Architecture for Data and Analytics Strategies
(https://www.gartner.com/document/3880568?ref=ddrec&refval=3891182)

Design Your IoT Data Architecture for Streaming Edge Analytics and Platform Advanced Analytics (https://www.gartner.com/document/3885163?ref=ddrec&refval=3891182)

Optimize Digital Customer Experiences With Machine-Learning-Enhanced Analytics (https://www.gartner.com/document/3880102?ref=ddrec&refval=3891182)

Assessing Microsoft Power BI for Three Critical Use Cases (https://www.gartner.com/document/3890106?ref=ddrec&refval=3891182)

Evaluation Criteria for Analytics and Business Intelligence Platforms (https://www.gartner.com/document/3890122?ref=ddrec&refval=3891182)

© 2018 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."