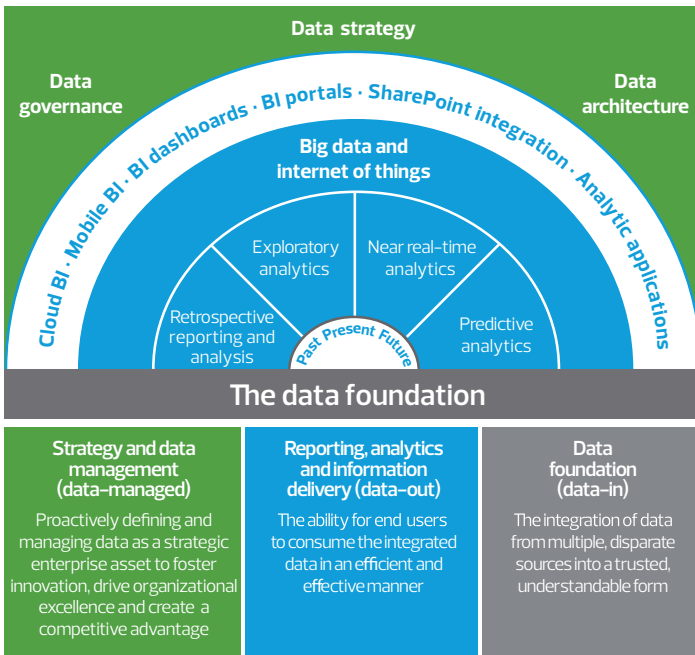


The importance of a solid data foundation

This is the first of a three-part series focused on designing a business intelligence (BI) solution. In order to design a complete solution, there are three main areas of focus:

- The **data foundation** encompasses the integration of data from multiple, disparate sources into a trusted, understandable form for use in reporting and analytics.
- **Reporting, analytics and information delivery** encompasses the ability for end users to efficiently consume the integrated data for proactive decision-making, creating a competitive advantage and driving organizational excellence.
- **Strategy and data management** encompasses proactively defining and managing data as an enterprise asset.



Criteria for a solid data foundation

The ultimate goal of business intelligence solutions is to provide information and insights that facilitate proactive decision-making, create a competitive advantage and drive organizational excellence. However, the data foundation is the basis for all types of reporting and analytics.

The goal of the data foundation is to provide integrated, trusted and timely data from which innovative reporting and analytics can be performed.

Integrated

Organizations have a wide range of data available to them. This includes both internal data from transaction systems and external data from industry groups, vendors and customers, and regulatory agencies. In addition, many organizations have a largely untapped source of information in unstructured data from social media, emails, etc. Integrating the organization's data for use in reporting and analytics not only saves a significant amount of time, it promotes enterprise consistency, collaboration and communication.

Trusted

Distrust in an organization's data is one of the primary reasons business intelligence initiatives fail. Data that is accurate, correct, consistent, complete and verifiable will facilitate people's trust and use of an integrated data foundation.

Timely

Effective decision-making requires all the necessary information be available when it is needed. By the time many organizations go through the process to manually collect and summarize the data they need, it is often too late to take action on it. Having an integrated data foundation that provides the necessary data when needed provides a significant boost to enterprise decision-making.

In essence, the data foundation is at the center of providing a trusted, single version of the truth.

Benefits of a robust data foundation

A robust data foundation provides an organization with tremendous benefits, both in terms of efficiency and effectiveness related to decision-making.

One-stop shopping for data

As noted earlier, one of the most significant uses of time in decision-making is getting the data into a usable format. This includes finding, collecting, summarizing and structuring data from multiple sources. In many cases, this can be a time-intensive, manual effort. A robust data foundation changes the paradigm from 80 percent of the time spent gathering the data to 80 percent of the time spent analyzing the data.

Single version of the truth

Getting different answers to the same question is a frustrating experience for decision-makers. This is a common occurrence

to organizations without an integrated data foundation. When left to individuals to find, collect and summarize data, it is inevitable that different answers will occur. A robust data foundation can provide the single version of the truth on which everyone in the organization can rely.

Drives common understanding across the enterprise

One of the key objectives of a data foundation is to integrate data from disparate sources. In order for this to be possible, there needs to be commonly understood business definitions for data, as well as agreed-upon dimensions of the business, hierarchies and reference data. A data foundation provides the structure and enforcement for these, resulting in everyone in the organization operating on the same page.

Data as a service

When everyone agrees upon the dimensions of the business, business definitions for data and how data is to be aggregated or summarized, the focus turns to how to analyze the data and make decisions. A robust data foundation provides the framework to focus discussions on analysis and decision-making instead of arguing over how data should be defined. Additionally, the chief information officer's (CIO) job can shift from maintaining disparate sources of data, to focusing full-time equivalent (FTE) time on value-add capabilities to the organization.

Consequences of the lack of a robust data foundation

While it is possible to achieve quality reporting and analytics without investing the time and effort of establishing the data foundation, not doing so introduces a significant increase in risk and inefficiency to the organization.

The lack of a data foundation can result in:

Multiple answers to the same question

One of the primary benefits of the data foundation is to provide a single version of the truth for the organization. Integrated data with commonly understood business definitions structured around the way an organization operates provides a consistent base for all reporting and analytics. Without such a base, it is often left up to individuals to find, collect and summarize data for their own use. This also leaves the definition of data and

derived calculations open to individual interpretation. The result is often conflicting information and answers related to the same question. This is one of the primary reasons individuals have a distrust in an organization's data.

Making less than optimal business decisions

Without a trusted, integrated base of information from which to generate reports and perform analysis, finding information on which to make business decisions can be a time-consuming task that does not always provide what is needed. Often shortcuts are taken and concessions are made in order to make decisions in a timely manner. This can result in multiple answers to the same question or decisions being made on faulty analysis due to poor data.

Wasted time finding, collecting and summarizing data for use in reporting and analytics

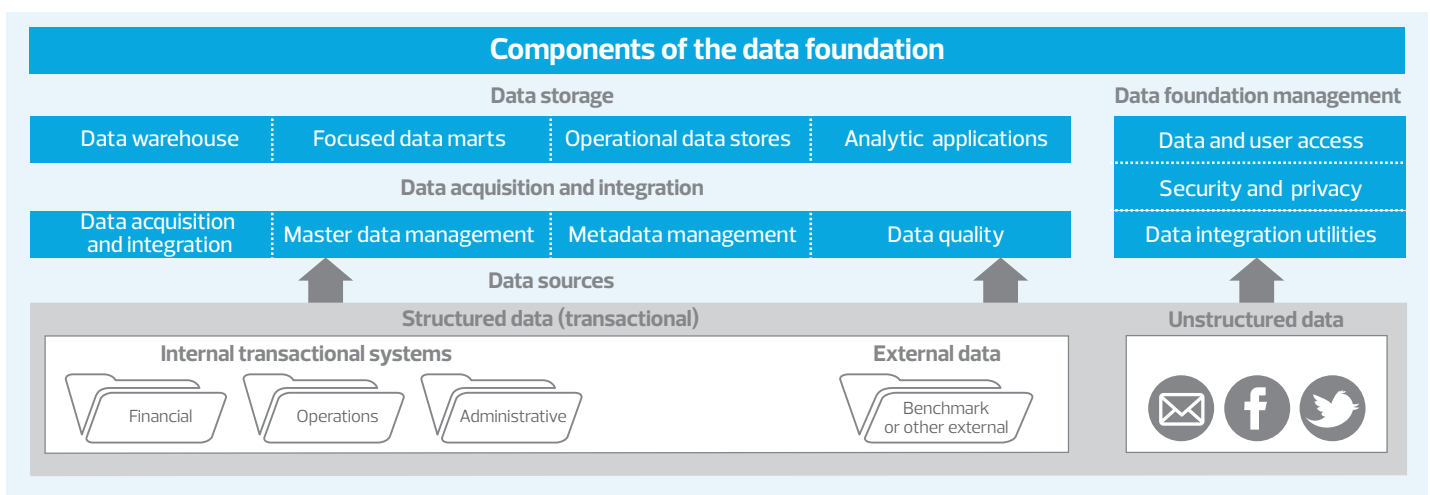
As noted earlier, many organizations waste an incredible amount of time manually finding, collecting, integrating and summarizing data to perform the analysis they need for decision-making. Not only is this a poor use of an organization's resources, the fact that this effort is often duplicated by many different people can result in different, and often erroneous, answers.

Implementing a robust data foundation

The core purpose of the data foundation is to provide trusted, integrated data that turns an organization's raw data into actionable information in a timely manner. In order to implement a robust data foundation, there are certain core components that should be considered, as illustrated below. Each of these components is discussed below:

Data sources

An organization has a wide range of internal and external data available for use in reporting and analytics. This includes structured data from transaction systems and external data sources, as well as unstructured data from social media, emails and spreadsheets. Integrating data from each of these types of data sources brings its own unique value and challenges. The data foundation needs to be able to address these challenges through its data acquisition and integration, and data storage frameworks.



Data acquisition and integration

Basically, data acquisition and integration take raw data from the disparate data sources available to an organization and integrate it into a single version of the truth. There are different approaches associated with data acquisition and integration, as follows:

▪ *Data acquisition*

Data acquisition encompasses the organizational constructs; processes, procedures and best practices; and tools and technologies used to gain access and secure the required data (both structured and unstructured) from the authoritative source systems to be used in reporting and analytics.

There are four basic methods in which organizations can acquire data for integration into an analytics environment:

- **Direct system access:** This method of data acquisition is focused on gaining access to and securing structured data from internal and external sources in a recurring batch or on an individual data element basis through access directly to the source system (normally done through application program interfaces, open database connectivity, etc.).
- **Batch data acquisition:** This grouping of capabilities is focused on gaining access to and securing structured data from internal and external sources in a recurring batch manner, usually through data extracts.
- **Near real-time (message-based data acquisition):** This grouping of capabilities is focused on providing continuous flow processing of data from source systems into the business intelligence environment, usually sourced from real-time messages.
- **Unstructured data acquisition:** This grouping of capabilities is focused on creating a metadata-based link to unstructured data that facilitates the integration with structured data.

▪ *Data integration*

Data integration encompasses the organizational constructs; processes, procedures and best practices; and tools and technologies to logically and physically integrate structured and unstructured data from disparate sources within the business intelligence environment.

There are several groupings of capabilities that should be considered when implementing data integration. These are largely aligned with the data acquisition methods outlined earlier:

- **Batch data integration:** This grouping of capabilities is focused on taking the batch data extracts noted earlier and processing them through a series of steps to load them into the target data store. Capabilities included in this method include standardization, data de-duplication, business rules application and transformation, data collision resolution, reference integrity enforcement and change data capture.

- **Near real-time integration:** This method is focused on processing individual data elements through continuous flow processing to its intended consumption. Under this method, the data needs to be delivered to its intended data consumer with little to no delay from the time it was acquired from the source.
- **Big data integration:** This method is focused on integrating data (both structured and unstructured) from disparate data sources. The need for big data integration is driven by an increase in the complexity of the data involved (as defined by the volume, variety and velocity of the data) to the point where traditional methods are not sufficient to process it in an efficient manner.

Master data, metadata and data quality

Within the data integration framework, there are certain core components that need to be in place to ensure the integrity of the integrated data. These include:

▪ *Master data management*

Master data provides the framework for the management of integrated enterprise data and includes:

- **Core concept definitions:** Centered on the way in which an organization views its data, and provides the pillars that connect enterprise data
- **Hierarchies:** Centered around the way organizations aggregate or roll up their data for use in reporting and analytics
- **Key reference data:** Centered around common codes used throughout the organization

▪ *Metadata management*

Metadata and master data are the glue that ties enterprise data together to ensure common understanding. Metadata is commonly defined as “data about data.” Metadata management encompasses the organizational constructs; processes, procedures and best practices; and tools and technologies used to provide common definitions of individual data elements, as well as transparency into its data lineage. Basically, metadata provides information about data from the time it is acquired from the source system to the time it is accessed and used by the data consumer.

While there is a wide variety of metadata available within a business intelligence environment, the two most critical are data lineage and business definitions:

- Data lineage metadata contains information as to what happened to the data from the time it entered the business intelligence environment until the time it was ultimately consumed.
- Business definitions ensure consistent understanding of what a given data element means.

- **Data quality**

Data quality encompasses the organizational constructs; processes, procedures and best practices; and tools and technologies used to facilitate the ability of an organization's data to serve its intended purposes and meet the quality standards of accuracy, correctness, availability and timeliness, completeness, relevance, consistency, and auditability and verifiability.

There are two main types of data quality to be considered in a business intelligence environment:

- **Business data quality** – Focused on ensuring that the data is in alignment with quality expectations of the business
- **Technical data quality** – Focused on ensuring the data is in alignment with the technical standards of the business intelligence environment

Implementing data quality processing within the data integration framework can take two forms:

- **Preventive data quality** – Focused on identifying potential data quality issues during the integration process to prevent potentially erroneous data from entering the business intelligence environment
- **Detective data quality** – Focused on identifying data quality issues that currently exist, or are likely to exist in the future, in the business intelligence environment

Data storage

Data storage encompasses the organizational constructs; processes, procedures and best practices; and tools and technologies in which integrated data is organized and stored in a manner that is optimized for both efficiency and flexibility, and end-user access and consumption.

Within an analytics environment, there are several methods in which to organize and store the data to be consumed by reporting and analytics. While each offers benefits and drawbacks, the key is to provide the flexibility necessary to meet current and future reporting and analytics needs of the business.

The main methods to organize and store data for consumption through reporting and analytics include the data warehouse, focused analytic data marts, an operational data store and analytic applications.

- **The enterprise data warehouse (EDW)** – The goal of the EDW is to provide a comprehensive store of integrated organizational data for use in reporting and analytics and is often organized by main subject areas of the business (e.g., financial, operational, administrative, etc.). An EDW usually takes the form of second or third normal form data structures or conformed dimensional data marts.
- **Focused analytical data marts** – Often, organizations do not need a comprehensive EDW to satisfy their reporting and analytics needs. Focused analytical data marts allow for quick realization of reporting and analytic capabilities around a given business challenge

or domain. However, if implemented correctly, focused analytical marts can evolve into an enterprise reporting and analytic platform.

- **Operational data store (ODS)** – An ODS tends to be used primarily in one of two ways. First, ODSs are often used to integrate data for use in operational reporting. ODSs are often more volatile in nature and used for one-off reporting and analysis scenarios than EDWs. Second, ODSs are sometimes used as a staging area for integrating data into an EDW or data mart. Under either scenario, ODSs can be a valuable component to an analytics platform and introduce added flexibility to address certain unique reporting and analytics needs.

Data foundation management

The final components of a data foundation are global in nature and can apply to and support any of the other components. Primarily, they are focused on providing, securing, protecting or processing the data. These components include:

- **Data and user access** – A data foundation can contain a significant volume and variety of data. Not all people need access to all the information. In fact, making all the data available to everyone often results in information overload and can often have the opposite effect than intended. Implementing proper access rights based on a person's role within the organization and the group to which they belong will allow for the targeted delivery of information they need to perform their job function.
- **Security and privacy** – With an analytics environment, there will often be data of a sensitive nature. This can include such things as salary information, personally identifiable information and, in health care, personal health information. Protecting this information against unauthorized access and data breaches is critical. The security approach must balance protecting the data with the need to making it available to certain people for use in reporting and analytics.
- **Data integration utilities** – Data integration utilities are those items that are reused throughout the business intelligence environment. They include such things as common error handling, restartability of data integration processes and automated data integration job processing.

Conclusion

Reporting and analytics can be transformational for an organization. However, having the proper data foundation that provides trusted, well-integrated and well-managed data is essential to realize the desired reporting and analytical capabilities. Mapping out a strategy and plan to establish the data foundation is time well spent and will provide a return many times over.

+1 800 274 3978
rsmus.com

This document contains general information, may be based on authorities that are subject to change, and is not a substitute for professional advice or services. This document does not constitute audit, tax, consulting, business, financial, investment, legal or other professional advice, and you should consult a qualified professional advisor before taking any action based on the information herein. RSM US LLP, its affiliates and related entities are not responsible for any loss resulting from or relating to reliance on this document by any person. Internal Revenue Service rules require us to inform you that this communication may be deemed a solicitation to provide tax services. This communication is being sent to individuals who have subscribed to receive it or who we believe would have an interest in the topics discussed.

RSM US LLP is a limited liability partnership and the U.S. member firm of RSM International, a global network of independent audit, tax and consulting firms. The member firms of RSM International collaborate to provide services to global clients, but are separate and distinct legal entities that cannot obligate each other. Each member firm is responsible only for its own acts and omissions, and not those of any other party. Visit rsmus.com/aboutus for more information regarding RSM US LLP and RSM International.

RSM® and the RSM logo are registered trademarks of RSM International Association. *The power of being understood*® is a registered trademark of RSM US LLP.

