

What is Data Management and Why is it Important?



In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

In this e-guide:

Effective data management is a crucial piece of deploying the IT systems that run business applications and provide analytical information to help drive operational decision-making and strategic planning by corporate executives, business managers and other end users.

It's more important than ever to understand the ins-and-outs of data management so your organization can realize the benefits of good data management and avoid data breaches and regulatory compliance problems. In this comprehensive guide, you'll explore:

- Key parts of the data management process
- Critical job skills and data management tasks
- Data warehouses vs. data lakes vs. DBMS types
- Benefits of a solid approach to data management
- Data management best practices
- And more

In this e-guide

▀ Importance of data management

▀ Types of data management functions

▀ Data management tools and techniques

▀ Data management best practices

▀ Data management risks and challenges

▀ Data management tasks and roles

▀ Benefits of good data management

▀ Data management history and evolution

▀ Further reading

▀ What is Data Management and Why is it important?

Margaret Rouse, [WhatIs.com](#), [SearchDataManagement.com](#)

Data management is the process of ingesting, storing, organizing and maintaining the [data](#) created and collected by an organization. Effective data management is a crucial piece of deploying the IT systems that run business applications and provide analytical information to help drive operational decision-making and strategic planning by corporate executives, business managers and other end users.

The data management process includes a combination of different functions that collectively aim to make sure that the data in corporate systems is accurate, available and accessible. Most of the required work is done by IT and data management teams, but business users typically also participate in some parts of the process to ensure that the data meets their needs and to get them on board with policies governing its use.

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

Importance of data management

Data increasingly is seen as a corporate asset that can be used to make more-informed business decisions, improve marketing campaigns, optimize business operations and reduce costs, all with the goal of increasing revenue and profits. But a lack of proper data management can saddle organizations with incompatible data silos, inconsistent data sets and [data quality](#) problems that limit their ability to run business intelligence (BI) and analytics applications -- or, worse, lead to faulty findings.

Data management has also grown in importance as businesses are subjected to an increasing number of regulatory compliance requirements, including data privacy and protection laws such as GDPR and the California Consumer Privacy Act. In addition, companies are capturing ever-larger volumes of data and a wider variety of data types, both hallmarks of the [big data](#) systems many have deployed. Without good data management, such environments can become unwieldy and hard to navigate.

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

Types of data management functions

The separate disciplines that are part of the overall data management process cover a series of steps, from data processing and storage to governance of how data is formatted and used in operational and analytical systems. Development of a data architecture is often the first step, particularly in large organizations with lots of data to manage. An architecture provides a blueprint for the databases and other data platforms that will be deployed, including specific technologies to fit individual applications.

Databases are the most common platform used to hold corporate data; they contain a collection of data that's organized so it can be accessed, updated and managed. They're used in both transaction processing systems that create operational data, such as customer records and sales orders, and [data warehouses](#), which store consolidated data sets from business systems for BI and analytics.

Database administration is a core data management function. Once databases have been set up, performance monitoring and tuning must be done to maintain acceptable response times on database queries that users run to get information from the data stored in them. Other administrative tasks include database design, configuration, installation and updates; data

In this e-guide

■ [Importance of data management](#)

■ [Types of data management functions](#)

■ [Data management tools and techniques](#)

■ [Data management best practices](#)

■ [Data management risks and challenges](#)

■ [Data management tasks and roles](#)

■ [Benefits of good data management](#)

■ [Data management history and evolution](#)

■ [Further reading](#)

security; database backup and recovery; and application of software upgrades and security patches.

The primary technology used to deploy and administer databases is a database management system ([DBMS](#)), which is software that acts as an interface between the databases it controls and the database administrators, end users and applications that access them. Alternative data platforms to databases include file systems and cloud object storage services; they store data in less structured ways than mainstream databases do, which offers more flexibility on the types of data that can be stored and how it's formatted. As a result, though, they aren't a good fit for transactional applications.

Other fundamental data management disciplines include [data modeling](#), which diagrams the relationships between data elements and how data flows through systems; [data integration](#), which combines data from different data sources for operational and analytical uses; [data governance](#), which sets policies and procedures to ensure data is consistent throughout an organization; and data quality management, which aims to fix data errors and inconsistencies. Another is [master data management](#) (MDM), which creates a common set of reference data on things like customers and products.

In this e-guide

- Importance of data management
- Types of data management functions
- Data management tools and techniques
- Data management best practices
- Data management risks and challenges
- Data management tasks and roles
- Benefits of good data management
- Data management history and evolution

Further reading

Key parts of the data management process

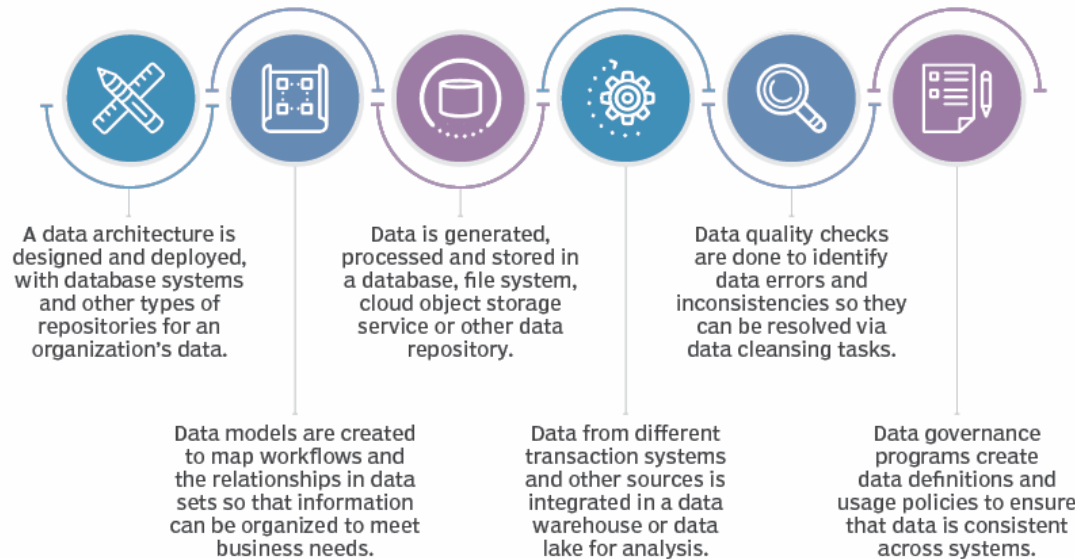


ILLUSTRATION: FIREOHEART/GETTY IMAGES; ALEXNDZ/A/OBE STOCK

©2019 TECHTARGET. ALL RIGHTS RESERVED TechTarget

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

Data management tools and techniques

A wide range of technologies, tools and techniques can be employed as part of the data management process. That includes the following available options for different aspects of managing data.

Database management systems. The most prevalent type of DBMS is the [relational database management system](#). Relational databases organize data into tables with rows and columns that contain database records; related records in different tables can be connected through the use of primary and foreign keys, avoiding the need to create duplicate data entries. Relational databases are built around the SQL programming language and a rigid data model best suited to structured transaction data. That and their support for the ACID transaction properties -- atomicity, consistency, isolation and durability -- have made them the top database choice for transaction processing applications.

However, other [types of DBMS technologies](#) have emerged as viable options for different kinds of data workloads. Most are categorized as NoSQL databases, which don't impose rigid requirements on data models and database schemas; as a result, they can store unstructured and semistructured data, such as sensor data, internet clickstream records and network, server and application logs.

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

There are four main types of NoSQL systems: document databases that store data elements in document-like structures, key-value databases that pair unique keys and associated values, wide column stores with tables that have a large number of columns, and graph databases that connect related data elements in a graph format. The NoSQL name has become something of a misnomer -- while NoSQL databases don't rely on SQL, many now support elements of it and offer some level of ACID compliance.

Additional database and DBMS options include in-memory databases that store data in a server's memory instead of on disk to accelerate I/O performance and columnar databases that are geared to analytics applications. Hierarchical databases that run on mainframes and predate the development of relational and NoSQL systems are also still available for use. Users can deploy databases in on-premises or cloud-based systems; in addition, various database vendors offer managed cloud database services, in which they handle database deployment, configuration and administration for users.

Big data management. NoSQL databases are often used in big data deployments because of their ability to store and manage various data types. Big data environments are also commonly built around open source technologies such as [Hadoop](#), a distributed processing framework with a file system that runs across clusters of commodity servers; its associated HBase database; the Spark processing engine; and the Kafka, Flink and

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

Storm stream processing platforms. Increasingly, big data systems are being deployed in the cloud, using object storage such as Amazon Simple Storage Service (S3).

Data warehouses and data lakes. Two alternative repositories for managing analytics data are [data warehouses and data lakes](#). Data warehousing is the more traditional method -- a data warehouse typically is based on a relational or columnar database, and it stores structured data pulled together from different operational systems and prepared for analysis. The primary [data warehouse use cases](#) are BI querying and enterprise reporting, which enable business analysts and executives to analyze sales, inventory management and other key performance indicators.

An enterprise data warehouse includes data from business systems across an organization. In large companies, individual subsidiaries and business units with management autonomy may build their own data warehouses. Data marts are another option -- they're smaller versions of data warehouses that contain subsets of an organization's data for specific departments or groups of users.

Data lakes, on the other hand, store pools of big data for use in predictive modeling, machine learning and other advanced analytics applications. They're most commonly built on Hadoop clusters, although [data lake deployments](#) are also done on NoSQL databases or cloud object storage; in addition, different platforms can be combined in a distributed data lake

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

environment. The data may be processed for analysis when it's ingested, but a data lake often contains raw data stored as is. In that case, data scientists and other analysts typically do their own data preparation work for specific analytical uses.

Data integration. The most widely used data integration technique is extract, transform and load (ETL), which pulls data from source systems, converts it into a consistent format and then loads the integrated data into a data warehouse or other target system. However, data integration platforms now also support a variety of other [integration methods](#). That includes extract, load and transform (ELT), a variation on ETL that leaves data in its original form when it's loaded into the target platform. ELT is a common choice for data integration jobs in data lakes and other big data systems.

ETL and ELT are batch integration processes that run at scheduled intervals. Data management teams can also do real-time data integration, using methods such as change data capture, which applies changes to the data in databases to a data warehouse or other repository, and streaming data integration, which integrates streams of real-time data on a continuous basis. Data virtualization is another integration option -- it uses an abstraction layer to create a virtual view of data from different systems for end users instead of physically loading the data into a data warehouse.

In this e-guide

■ [Importance of data management](#)

■ [Types of data management functions](#)

■ [Data management tools and techniques](#)

■ [Data management best practices](#)

■ [Data management risks and challenges](#)

■ [Data management tasks and roles](#)

■ [Benefits of good data management](#)

■ [Data management history and evolution](#)

■ [Further reading](#)

Data governance, data quality and MDM. Data governance is primarily an organizational process; software products that can help manage data governance programs are available, but they're an optional element. While governance programs may be managed by data management professionals, they usually include a data governance council made up of business executives who collectively make decisions on common data definitions and corporate standards for creating, formatting and using data.

Another key aspect of governance initiatives is [data stewardship](#), which involves overseeing data sets and ensuring that end users comply with the approved data policies. Data steward can be either a full- or part-time position, depending on the size of an organization and the scope of its governance program. Data stewards can also come from both business operations and the IT department; either way, a close knowledge of the data they oversee is normally a prerequisite.

Data governance is closely associated with [data quality improvement efforts](#); metrics that document improvements in the quality of an organization's data are central to demonstrating the business value of governance programs. Data quality techniques include data profiling, which scans data sets to identify outlier values that might be errors; data cleansing, also known as data scrubbing, which fixes data errors by modifying or deleting bad data; and data validation, which checks data against preset quality rules.

In this e-guide

■ [Importance of data management](#)

■ [Types of data management functions](#)

■ [Data management tools and techniques](#)

■ [Data management best practices](#)

■ [Data management risks and challenges](#)

■ [Data management tasks and roles](#)

■ [Benefits of good data management](#)

■ [Data management history and evolution](#)

■ [Further reading](#)

Master data management is also affiliated with data governance and data quality, although MDM hasn't been adopted as widely as the other two data management functions. That's partly due to the complexity of MDM programs, which mostly limits them to large organizations. MDM creates a central registry of master data for selected data domains -- what's often called a *golden record*. The master data is stored in an MDM hub, which feeds the data to analytical systems for consistent enterprise reporting and analysis; if desired, the hub can also push updated master data back to source systems.

Data modeling. Data modelers create a series of conceptual, logical and physical data models that document data sets and workflows in a visual form and map them to business requirements for transaction processing and analytics. Common [techniques for modeling data](#) include the development of entity relationship diagrams, data mappings and schemas. In addition, data models must be updated when new data sources are added or an organization's information needs changes.

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

Data management best practices

A well-designed [data governance program](#) is a critical component of effective data management strategies, especially in organizations with distributed data environments that include a diverse set of systems. A strong focus on data quality is also a must. In both cases, though, IT and data management teams can't go it alone. Business executives and users have to be involved to make sure their data needs are met and data quality problems aren't perpetuated. The same applies to data modeling projects.

Also, the multitude of databases and other data platforms available to be deployed requires a careful approach when designing a data architecture and evaluating and selecting technologies. IT and data managers must be sure the systems they implement are fit for the intended purpose and will deliver the data processing capabilities and analytics information required by an organization's business operations.

DAMA International, the Data Governance Professionals Organization and other industry groups work to advance understanding of data management disciplines and offer best-practices guidance. For example, DAMA has published *DAMA-DMBOK: Data Management Body of Knowledge*, a reference book that attempts to define a standard view of data management

In this e-guide

Importance of data management

Types of data management functions

Data management tools and techniques

Data management best practices

Data management risks and challenges

Data management tasks and roles

Benefits of good data management

Data management history and evolution

Further reading

functions and methods. Commonly referred to as the DMBOK, the book was first published in 2009; a [DMBOK2 second edition](#) was released in 2017.

Data management risks and challenges

If an organization doesn't have a well-designed data architecture, it can end up with siloed systems that are hard to integrate and manage in a coordinated way. Even in better-planned environments, enabling data scientists and other analysts to find and access relevant data can be a challenge, especially when the data is spread across various databases and big data systems. To help make data more accessible, many data management teams are [creating data catalogs](#) that document what's available in systems and typically include business glossaries, metadata-driven data dictionaries and data lineage records.

The shift to the cloud can ease some aspects of data management work, but it also creates new challenges. For example, [migrating to cloud databases](#) and big data platforms can be complicated for organizations that need to move data and processing workloads from existing on-premises systems. Costs are another big issue in the cloud -- the use of cloud systems and managed services must be monitored closely to make sure data processing bills don't exceed the budgeted amounts.

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

Many data management teams are now among the employees who are accountable for protecting corporate data security and limiting potential legal liabilities for data breaches or misuse of data. Data managers need to help ensure compliance with both government and industry regulations on data security, privacy and usage. That has become a more pressing concern with the passage of GDPR, the European Union's data privacy law that took effect in May 2018, and the California Consumer Privacy Act, which was signed into law in 2018 and is scheduled to become effective at the start of 2020.

Data management tasks and roles

The data management process involves a wide range of tasks, duties and skills. In smaller organizations with limited resources, individual workers may handle multiple roles. But in general, [data management professionals](#) include data architects, data modelers, database administrators (DBAs), database developers, data quality analysts and engineers, data integration developers, data governance managers, data stewards and data engineers, who work with analytics teams to build data pipelines and prepare data for analysis.

In this e-guide

- ▣ Importance of data management
- ▣ Types of data management functions
- ▣ Data management tools and techniques
- ▣ Data management best practices
- ▣ Data management risks and challenges
- ▣ Data management tasks and roles
- ▣ Benefits of good data management
- ▣ Data management history and evolution
- ▣ Further reading

Data management role

CRITICAL JOB SKILLS

Data governance, data quality assurance, data stewardship, master data management and data security management; knowledge of data frameworks, SQL and forward-looking data management technologies

COLLEGE

Bachelor's degree or higher; focus on computer science or information technology

CERTIFICATIONS

Software certifications useful in career advancement; may be required by some employers

DATA MANAGER

AVERAGE SALARY (2017)

\$80,502- \$104,653*

*Salary based on 5 years' experience in Boston. Provided by Dice.com

ILLUSTRATION: NEVBO/DOBE STOCK ©2018 TECHTARGET. ALL RIGHTS RESERVED. TechTarget

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

Data scientists and other data analysts may also handle some data management tasks themselves, especially in big data systems with raw data that needs to be filtered and prepared for specific uses. Likewise, application developers often help deploy and manage big data environments, which require new skills overall compared to relational database systems. As a result, organizations may have to hire new workers or retrain traditional DBAs to meet their big data management needs.

Benefits of good data management

A well-executed data management strategy can help companies gain potential competitive advantages over their business rivals, both by improving operational effectiveness and enabling better decision-making. Organizations with well-managed data can also become more agile, making it possible to spot market trends and move to take advantage of new business opportunities more quickly.

Effective data management can also help companies avoid data breaches, data privacy issues and regulatory compliance problems that could damage their reputation, add unexpected costs and put them in legal jeopardy. Ultimately, the biggest benefit that a solid approach to data management can provide is better business performance.

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

Data management history and evolution

The first flowering of data management was largely driven by IT professionals who focused on solving the problem of garbage in, garbage out in the earliest computers after recognizing that the machines reached false conclusions because they were fed inaccurate or inadequate data.

Beginning in the 1960s, industry groups and professional associations promoted best practices for data management, especially in terms of professional training and data quality metrics. Mainframe-based hierarchical databases also became available that decade.

The relational database emerged in the 1970s and then cemented its place at the center of the data management process in the 1980s. The idea of the data warehouse was conceived in the late 1980s, and early adopters of the concept began deploying data warehouses in the mid-1990s. By the early 2000s, relational software was a dominant technology, with a virtual lock on database deployments.

In this e-guide

Importance of data management

Types of data management functions

Data management tools and techniques

Data management best practices

Data management risks and challenges

Data management tasks and roles

Benefits of good data management

Data management history and evolution

Further reading

But the initial release of Hadoop became available in 2006 and was followed by the Spark processing engine and various other big data technologies. A range of NoSQL databases also started to become available in the same time frame. While relational technology still has the largest share by far, the rise of big data and NoSQL alternatives has given organizations a broader set of data management choices.

Next Article

In this e-guide

■ Importance of data management

■ Types of data management functions

■ Data management tools and techniques

■ Data management best practices

■ Data management risks and challenges

■ Data management tasks and roles

■ Benefits of good data management

■ Data management history and evolution

■ Further reading

■ About SearchDataManagement

[SearchDataManagement.com](https://www.searchdatamanagement.com) is a knowledge guide for data management professionals and business intelligence leaders. With our offering of news articles, learning guides, expert advice, and customized research, we offer a rich collection of insight, tips, and advice on how you can efficiently manage your data supply chain.

With the ever-growing volume of internal and external data flooding into businesses, educated data management strategies are essential.

Our site helps DBAs, developers, programmers, and data scientists to process relevant data efficiently, integrate traditional warehouses with modern tools like Hadoop and Spark, select a compliant governance strategy, ensure data quality/trustworthiness, and better streamline techniques and practices to enable informed, knowledgeable business decisions.

For further reading, visit us at:
[SearchDataManagement.com](https://www.searchdatamanagement.com)

Images: Fotolia

©2019 TechTarget. No part of this publication may be transmitted or reproduced in any form or by any means without written permission from the publisher.