

# Data Management Concepts

Dave E. Marcial, Ph.D.



# What is Data Management?

- the practice of managing data as a valuable resource to unlock its potential for an organization.
- Managing data effectively requires having a data strategy and reliable methods to access, integrate, cleanse, govern, store and prepare data for analytics.
- In our digital world, data pours into organizations from many sources – operational and transactional systems, scanners, sensors, smart devices, social media, video and text. But the value of data is not based on its source, quality or format.
- Its value depends on what you do with it.

[https://www.sas.com/en\\_us/insights/data-management/data-management.html](https://www.sas.com/en_us/insights/data-management/data-management.html)



# What is Data Management?

- a comprehensive collection of practices, concepts, procedures, processes, and a wide range of accompanying systems that allow for an organization to gain control of its data resources.
- as an overall practice is involved with the entire lifecycle of a given data asset from its original creation point to its final retirement, how it progresses and changes throughout its lifetime through the internal (and external) data streams of an enterprise.

<https://www.dataversity.net/what-is-data-management/>



# What is Data Management?

- “The development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles.”



<https://dama.org/content/body-knowledge>

WWW.SU.EDU.PH



# What kind of data is managed via data management?

- Any kind of business data is subject to data management principles and procedures, but it is particularly useful in rectifying conflict among data from duplicative sources.
- Organizations that use cloud-based applications in particular find it hard to keep data orchestrated across systems.
- Data management practices can prevent ambiguity and make sure that data conforms to organizational best practices for access, storage, back up, and retirement, among other things.



# History of Data Management?

- Data management systems as we know them today weren't common until the 1970s. These data management systems were strictly operational. They provided records (reports) of business operations at a given point in time, pulled from a relational database that stored information in rows and columns (typically a data warehouse).

[https://www.sas.com/en\\_us/insights/data-management/data-management.html](https://www.sas.com/en_us/insights/data-management/data-management.html)

- Batch processing and extract, transform, load (ETL).
- Structured query language (SQL) and relational database management systems (RDBMSs).
- Not-only SQL (NoSQL) and nonrelational databases.
- Enterprise data warehouses, data lakes and data fabrics.
- Data federation and virtualization.
- Data catalogs, metadata management and data lineage.
- Cloud computing and event stream processing (data streaming).



# History of Data Management?

- is the organization of data, the steps used to achieve efficiency, and gather intelligence from that data.
- **Data Management**, as a concept, began in the 1960s, with ADAPSO (the Association of Data Processing Service Organizations) forwarding Data Management advice, with an emphasis on professional training and quality assurance metrics.
- Read more at <https://www.dataversity.net/brief-history-data-management/>





## Data Management vs Data Governance vs Database Management

Data Governance is a set of practices and concepts which prioritize and organize data, as well as the enforcement of policies around data, while following various regulations and curtailing poor data practices.

Data Governance is essentially a part of the greater whole of Data Management.

Database Management, on the other hand, is focused on the tools and technology used to create and alter the foundation of data, rather than the overall system used to organize the data.

Database Management is also a subdivision of Data Management.





# What are the processes of data management?

- ✓ A common approach to data management is to utilize a master data file, called **Master Data Management (MDM)**.
- ✓ This file provides a common definition of an asset and all its data properties in an effort to eliminate ambiguous or competing data policies and give the organization comprehensive stewardship over its data.



# Data Management Knowledge Areas

1. **Data governance**, which is the planning of all aspects of data management. This commonly includes ensuring availability, usability, consistency, integrity, and security of data managed by an organization.
2. **Data architecture**, or the overall structure of an organization's data and how it fits into a broader enterprise architecture.
3. **Data modeling and design**, which covers data analytics and the design, building, testing, and maintenance of analytics systems.
4. **Data storage and operations**, which is concerned with the physical hardware used to store and manage data.
5. **Data security**, which encompasses all elements of protecting data and ensuring only authorized users have access.



# Data Management Knowledge Areas

**6. Data integration and interoperability**, which includes everything to do with the transformation of data into a structured form (i.e., in an organized database) and the work necessary to maintain it.

**7. Documents and content**, which includes all forms of [unstructured data](#) and the work necessary to make it accessible to, and integrated with, structured databases.

**8. Reference and master data**, or the process of managing data in such a way that redundancy and other mistakes are reduced by standardizing data values.

**9. Data warehousing and business intelligence**, which involves the management and application of data for analytics and business decision making.

**10. Metadata**, which involves all elements of creating, collecting, organizing, and managing metadata (data that references other data, like headers, etc.).

**11. Data quality**, which involves the practices of monitoring data and data sources to ensure quality information is being delivered, integrity is being maintained, and poor quality data is being filtered out.

<https://www.techrepublic.com/article/data-management-a-cheat-sheet/>





# Data Management Practices and Concepts

## 1. Data Governance

- Data asset
- Data governance
- Data steward
- Data Ethics

## 2. Data Architecture

- Data architecture
- Data flows

## 3. Data modeling and Design

## 4. Database & Storage Management

- Data maintenance
- Database administration
- Database management system
- Business continuity planning
- Data subsetting

## 5. Data Security

- Data access
- Data erasure
- Data privacy
- Data security

## 6. Reference and Master Data

- Data integration
- Master data management
- Reference data

## 7. Data Integration and Inter-operability

- Data movement (Extract, transform, load )
- Data Interoperability

## 8. Documents and Content

- Document management system
- Records management

## 9. Data Warehousing and Business Intelligence

- Business intelligence
- Data analysis and Data mining
- Data warehouse and Data mart

## 10. Metadata

- Metadata management
- Metadata
- Metadata discovery
- Metadata publishing
- Metadata registry

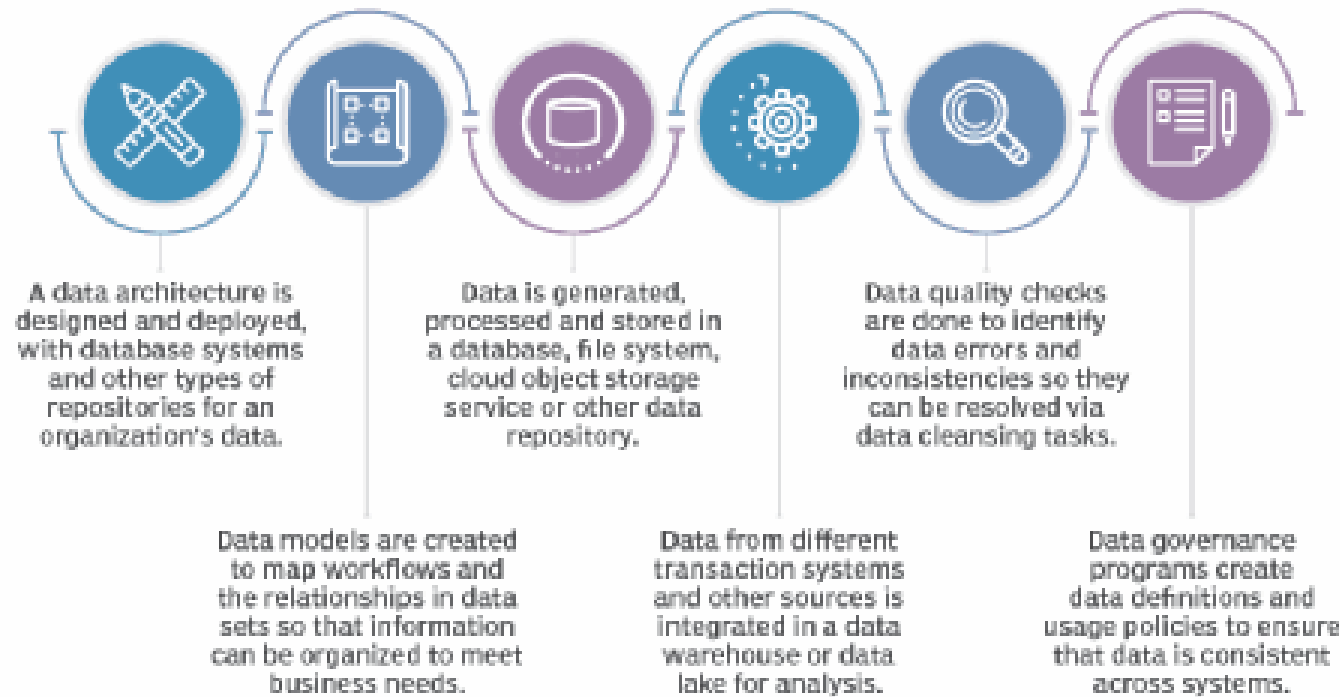
## 11. Data Quality

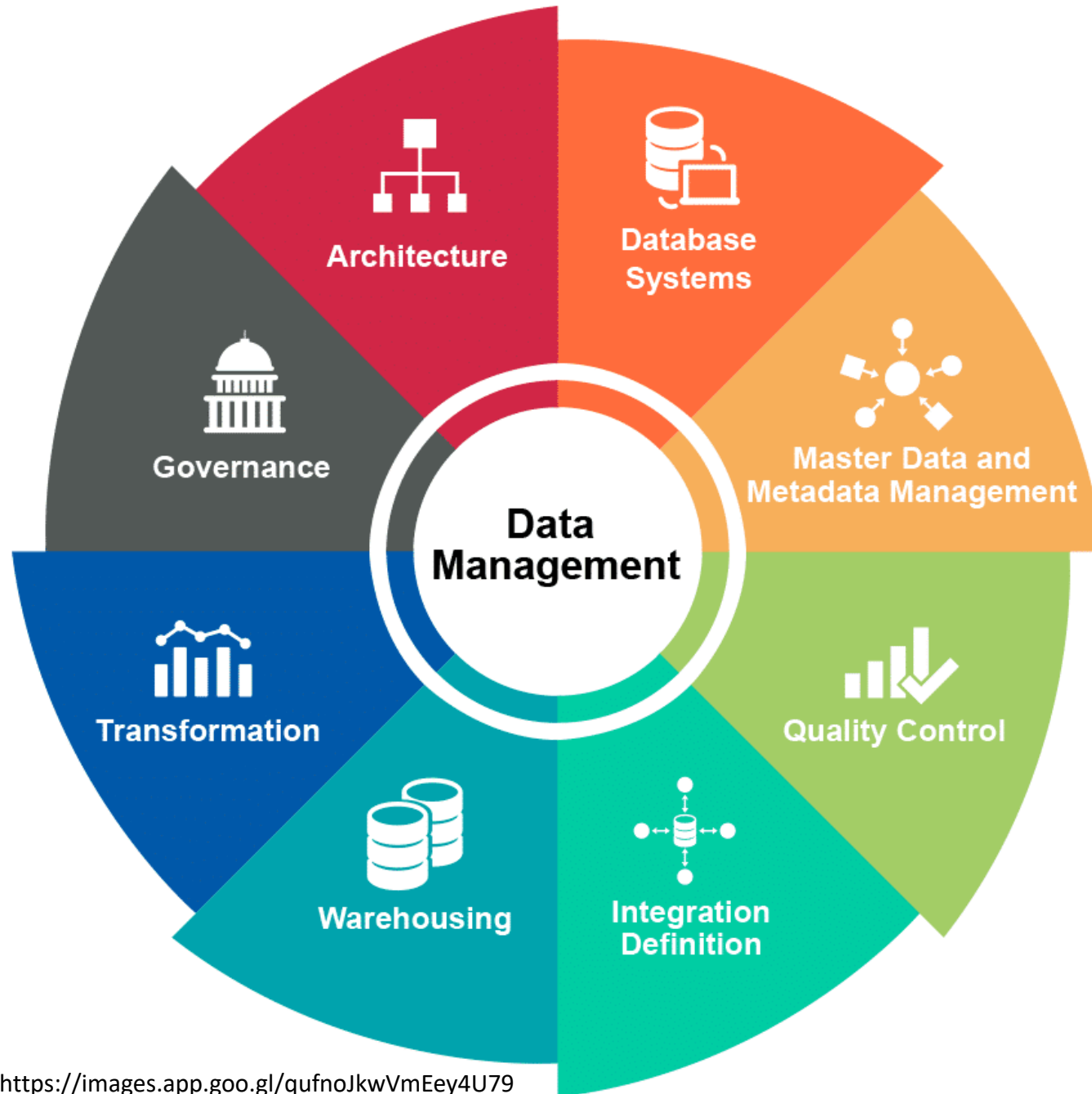
- Data discovery
- Data cleansing
- Data integrity
- Data enrichment
- Data quality
- Data quality assurance
- Secondary data

[https://en.wikipedia.org/wiki/Data\\_management](https://en.wikipedia.org/wiki/Data_management)



## Key parts of the data management process







# Data Management Life Cycle



# Data Management Life Cycle



<https://www.fws.gov/data/life-cycle>

## Plan

Planning for a project involves making decisions about data management, potential products, as well as data stewardship roles and responsibilities. It is important to document all stages of the data management life cycle and quality control prior to beginning a new project.

## Acquire

Data can be acquired by collecting new data, processing old or legacy data, collaborating with partners, and contracting others to collect data.

## Maintain

Data maintenance includes processing data for analysis, creating metadata, and making sure data are in a format that can be accessed by others in the future.

## Access

The ability to prepare, release, and share quality data to the public, other agencies, and internally is an important part of the life cycle process.

## Evaluate

Evaluate represents steps associated with processing and analyzing data. Important goals for both the processing and analysis of data are maximizing accuracy and productivity while minimizing costs.

## Archive

Data archiving supports the long-term storage of scientific data and the methods used to read or interpret them.

## Quality Assurance / Quality Control

Data quality management is the prevention of data defects that reduce our ability to apply data towards our science-based conservation efforts.

# The Benefits of Data Management Include

A well-developed Data Management program within an organization has the ability to positively effect change around the administration and use of data assets across all levels, departments, and lines of business.

- Improved operations management.
- More effective marketing and sales.
- Better regulation and compliance controls.
- Enhanced security and privacy.
- Reduction of risk across the board.
- Faster application and system development.
- Improved decision making and reporting.
- Sustained business growth.
- Business and technical alignment.
- Automated and/or streamlined operations.
- Greater collaboration and revenue growth.
- More consistency across all enterprise processes.

<https://www.dataversity.net/what-is-data-management/>





# The Benefits of Data Management Include

- The benefits of data management include enhanced compliance, greater security, improved sales and marketing strategies, better product classification, and improved **data**

**governance to reduce  
organizational risk.**



# A Data Management Professional is:

- “Any person who works in any facet of data management...
  - from the highly technical
    - (e.g., database administrators, network administrators, programmers)
  - to strategic business
    - (e.g., Data Stewards, Data Strategists, Chief Data Officers).”

<https://www.dataversity.net/what-is-data-management/>



# What skills do data management professionals need?

- **General computer science:** A qualified data management professional should be trained in the basics of computer science--they're going to be spending a lot of time using basic skills to organize data.
- **Database programming:** Some of the most important database languages in the data management world include SQL, [Python](#), R, Hadoop, XML, and PERL. Be sure to learn at least one of these languages and get familiar with its corresponding database platforms.
- **BI/BA:** Business intelligence (BI) and business analytics (BA) are at the core of why companies collect and organize data. Data management pros should be able to understand the hows and whys of analytics.
- **Cloud computing:** Data hosting can take up a lot of storage space, which is why many businesses turn to the cloud to host, manage, and analyze their data. Skilled data management professionals should be familiar with [AWS](#), [Microsoft Azure](#), [Google Cloud](#), [IBM Cloud](#), and other major platforms.
- **Machine learning:** Data analytics, in particular its later stages like [predictive analytics](#) and [prescriptive analytics](#), make extensive use of [machine learning](#) technology to decrease the computing time needed to deliver results.
- **Data management certifications:** Data management is a science in and of itself, and there are several certificates that data management professionals can pursue. DAMA International offers the [Certified Data Management Professional \(CDMP\) certification](#). [Oracle](#), [IBM](#), and others also offer certifications.
- **Soft skills:** Making use of data requires a lot of collaboration with non-IT departments to plan and execute big data strategies. Good writing, speaking, and innovative thinking are an essential set of skills for successful data management professionals.

<https://www.techrepublic.com/article/data-management-a-cheat-sheet/>





# Data Management in Today's World



## Data Quality: What You Need to Know

Outdated or unreliable data leads to mistakes and missteps. Yet many organizations distrust the quality of their data. Learn about the key features of data quality, why it's so crucial and how to fix data quality dilemmas.



## The SAS Data Governance Framework

Today's barrage of data demands critical governance decisions. An overarching approach to collecting, managing and storing data across the enterprise helps you keep pace with changing technologies, trends and regulations.



## Build a Data and Analytics Strategy

Wondering how to build a world-class analytics organization? Make sure information is reliable. Empower data-driven decisions. Drive the strategy. And know how to wring every last bit of value out of your data.



## Self-Service Data Preparation

Imagine the results if business users could prep data for analytics without relying on IT- no coding or special skills required. SAS Data Preparation lets business users access, cleanse, profile and transform data on their own.

# Who's Using Data Management?

## Retail



Understanding customers and responding appropriately to expectations requires having an accurate, up-to-date view of all the data – whether it's streaming, cloud based, or stored in a data lake or warehouse. From marketing to merchandising to sales, trusted data management is essential to taking charge of retail data.

## Manufacturing



In the manufacturing industry, nothing speaks success like quality. With solid data management and data quality technologies, manufacturers can efficiently manage product inventory, and integrate structured and unstructured data from all sources to get an enterprise view of performance, drive better outcomes and make well-informed business decisions.



# Who's Using Data Management?

## Banking



More than ever, issues around data privacy, compliance and digitization require banks to have a trusted data foundation. Only with a complete, integrated view of all their data – and sound techniques for quality, governance and personal data protection – can banks gain customers' trust and pursue forward-looking digital transformation efforts.

## Health Care



Enterprise data management is a must-have in the health care industry. The industry counts on being able to integrate data from all formats and sources – including data from outside of the organization – all while spotting duplicate data, fixing data quality issues, and adhering to strict regulatory and compliance requirements for protecting personal data and privacy.





# Who's Using Data Management?

## Government



Local and national governments are responsible for a vast range of services and programs. Reliable data management technologies support all those efforts – from fighting fraud and improper payments to ensuring citizen safety to overseeing population health outcomes, economic development and smart city initiatives.

## Small and midsize business



As small and midsize businesses work toward digital transformation, they need to implement data-driven business models and modernize legacy IT so they can be competitive with their larger counterparts. One way to get there is with reliable data management technology that can be catered to the needs of smaller businesses.



# Data Management Challenges

- **They don't know what data they have.** Data from an increasing number and variety of sources such as sensors, smart devices, social media, and video cameras is being collected and stored. But none of that data is useful if the organization doesn't know what data it has, where it is, and how to use it.
- **They must maintain performance levels as the data tier expands.** Organizations are capturing, storing, and using more data all the time. To maintain peak response times across this expanding tier, organizations need to continuously monitor the type of questions the database is answering and change the indexes as the queries change—without affecting performance.
- **They must meet constantly changing compliance requirements.** Compliance regulations are complex and multijurisdictional, and they change constantly. Organizations need to be able to easily review their data and identify anything that falls under new or modified requirements. In particular, personally identifiable information (PII) must be detected, tracked, and monitored for compliance with increasingly strict global privacy regulations.
- **They aren't sure how to repurpose data to put it to new uses.** Collecting and identifying the data itself doesn't provide any value—the organization needs to process it. If it takes a lot of time and effort to convert the data into what they need for analysis, that analysis won't happen. As a result, the potential value of that data is lost.
- **They must keep up with changes in data storage.** In the new world of data management, organizations store data in multiple systems, including data warehouses and unstructured data lakes that store any data in any format in a single repository. An organization's data scientists need a way to quickly and easily transform data from its original format into the shape, format, or model they need it to be in for a wide array of analyses.



# Data Management Best Practices

- **Create a discovery layer to identify your data.** A discovery layer on top of your organization's data tier allows analysts and data scientists to search and browse for datasets to make your data useable.
- **Develop a data science environment to efficiently repurpose your data.** A data science environment automates as much of the data transformation work as possible, streamlining the creation and evaluation of data models. A set of tools that eliminates the need for the manual transformation of data can expedite the hypothesizing and testing of new models.
- **Use autonomous technology to maintain performance levels across your expanding data tier.** Autonomous data capabilities use AI and machine learning to continuously monitor database queries and optimize indexes as the queries change. This allows the database to maintain rapid response times and frees DBAs and data scientists from time-consuming manual tasks.
- **Use discovery to stay on top of compliance requirements.** New tools use data discovery to review data and identify the chains of connection that need to be detected, tracked, and monitored for multijurisdictional compliance. As compliance demands increase globally, this capability is going to be increasingly important to risk and security officers.
- **Use a common query layer to manage multiple and diverse forms of data storage.** New technologies are enabling data management repositories to work together, making the differences between them disappear. A common query layer that spans the many kinds of data storage enables data scientists, analysts, and applications to access data without needing to know where it is stored and without needing to manually transform it into a usable format.





# Data Management Evolves

- With data's new role as business capital, organizations are discovering what digital startups and disruptors already know: Data is a valuable asset for identifying trends, making decisions, and taking action before competitors. The new position of data in the value chain is leading organizations to actively seek better ways to derive value from this new capital.
- Within companies, the data management responsibilities of the DBA are also evolving, reducing the number of mundane tasks so that DBAs can concentrate on more strategic issues and provide critical data management support in cloud environments involving key initiatives such as data modeling and data security.

