



**WELCOME!**

*DMBoK for Executives*

# *What will you learn in this course?*

**Module 1 - The Basics**

**Module 2 - Data Governance**

**Module 3 - Data Architecture**

**Module 4 - Data Modeling and Design**

**Module 5 - Data Storage and Operations**

**Module 6 - Data Security**

**Module 7 - Data Integration**

**Module 8 - Document and Content Management**

**Module 9 - Master & Reference Data Management**

**Module 10 - Metadata Management**

**Module 11 - Data Quality**

**Module 12 - Data Warehousing and Business Intelligence**

# *Why take this course?*

**Reason 1** - We cover every Data Management subject area

**Reason 2** - Real examples of use cases

**Reason 3** - Tips and tricks from experience

**Reason 4** - Answers most questions that you can expect in interviews

**Reason 5** - Gain confidence in attending any data meeting

**Reason 6** - Additional resources provided for further reading



# *Download the course resources*

**Main resource** - PDF Presentation file with all lessons

**Other resources** - additional files/resources provided in various lessons

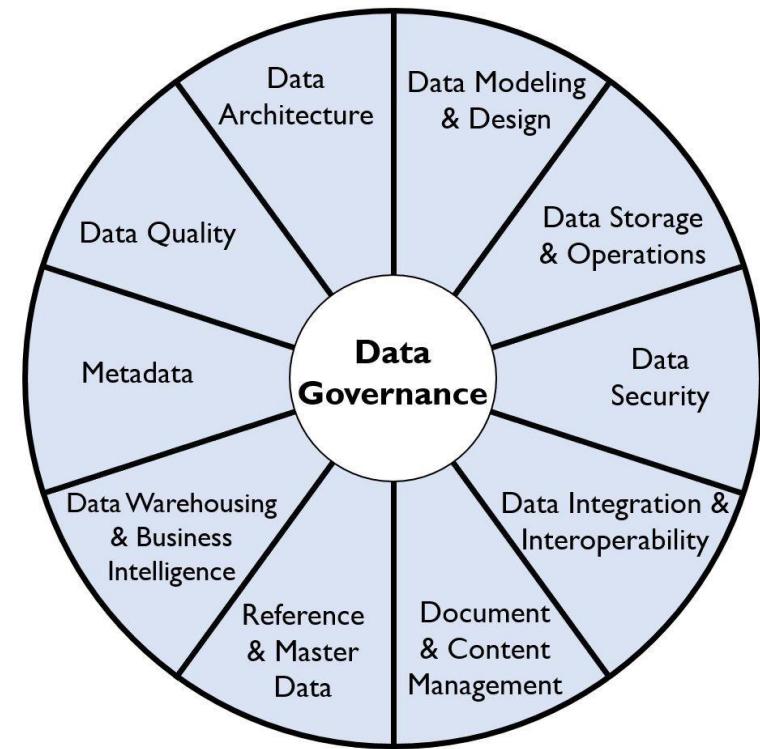
# *Module 1 - The Basics*

# What is Data Management?

**‘Data management is the practice of collecting, organizing, protecting and storing an organization’s data so it can be analyzed for business decisions.’**

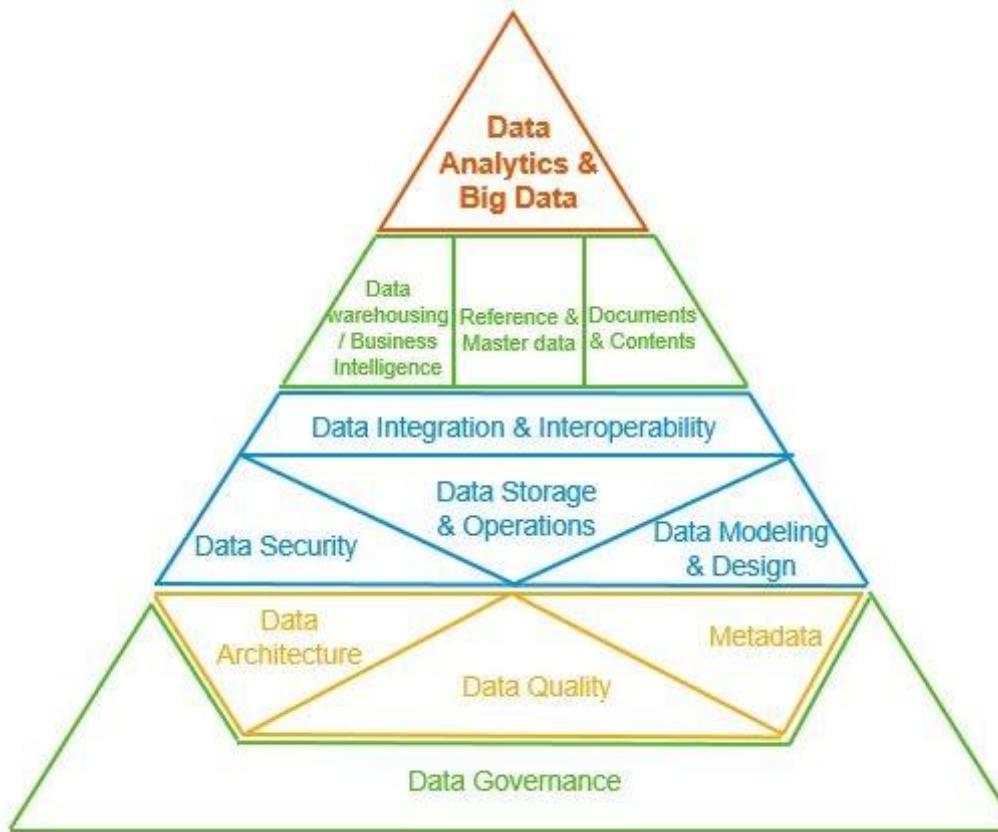
# Data Management Subject Areas

1. Data Governance
2. Data Modeling & Design
3. Data Storage & Operations
4. Data Security
5. Data Integration & Interoperability
6. Document & Content Management
7. Reference & Master Data
8. Data Warehousing & Business Intelligence
9. Metadata
10. Data Quality
11. Data Architecture



Copyright © 2017 DAMA International

# DMBOK pyramid by Peter Aiken



Source: Peter Aiken developed the DMBOK pyramid

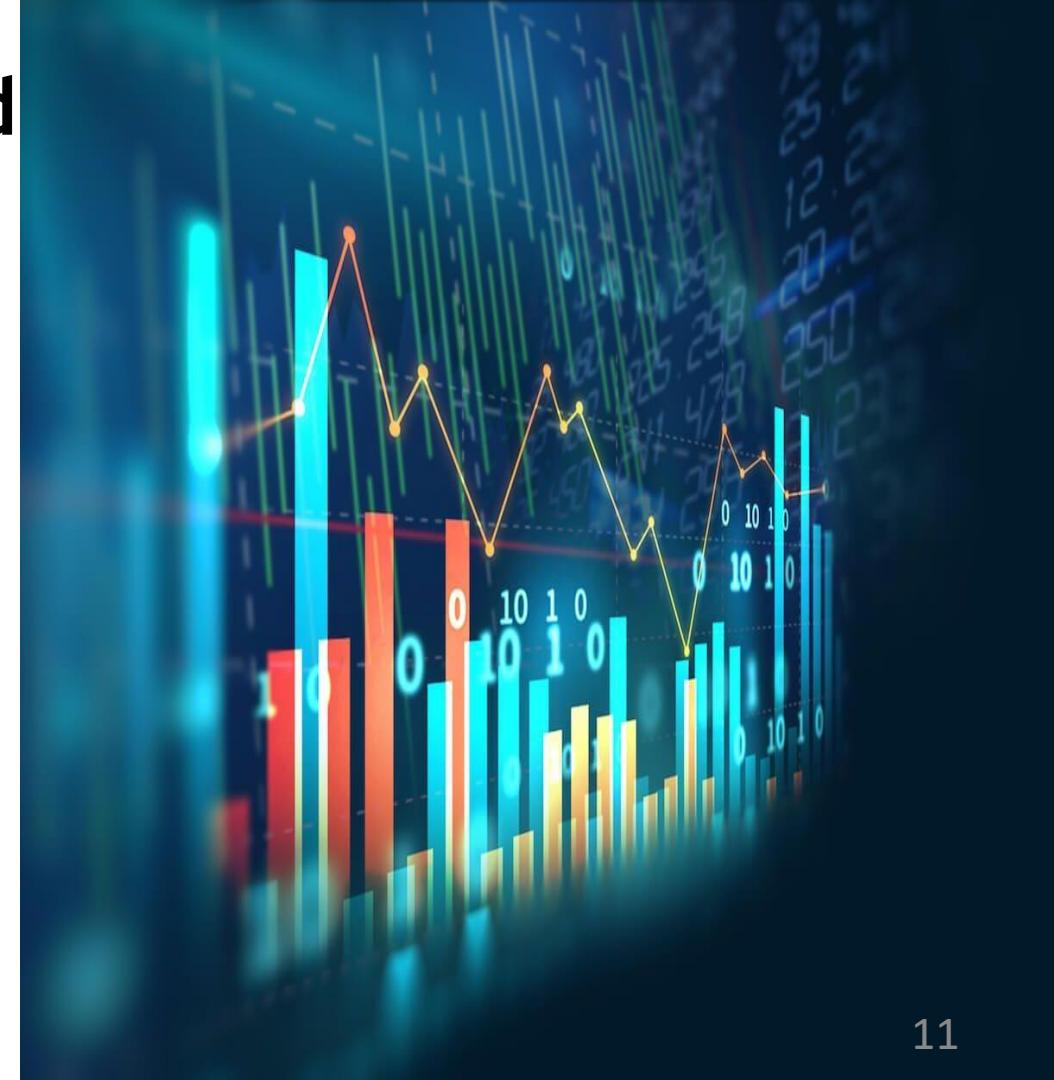
# Module 2 - Data Governance

# What is Data Governance?

- **Rules, Processes and Accountability** that allow the organization to better manage the availability, usability, security and integrity of the corporate data sources.
- **Tip - Think about it as bringing data under control and keeping it secure and consistent.**

# 7 reasons why you need Data Governance

1. Secure your data
2. Ensure compliance with regulations and data privacy laws
3. Improve the data quality
4. Avoid inconsistent data silos
5. Improve trust in the data
6. Better decision making
7. Improve efficiency



# Regulations and Data

Amazon: \$886 million

WhatsApp - \$267 million

Home Depot: \$200 million

Uber: \$148 million

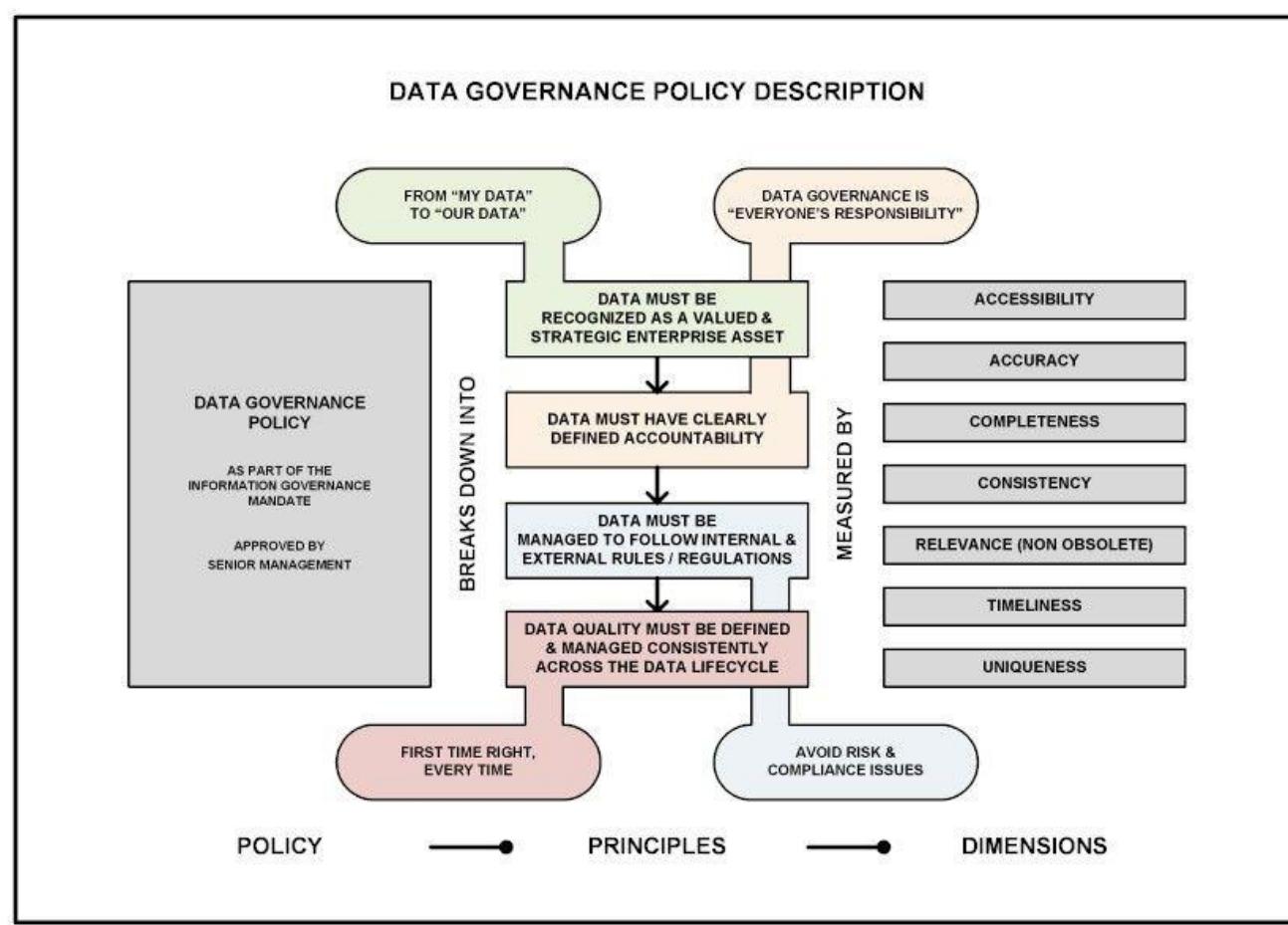
Yahoo: \$85 million

Capital One: \$80 million

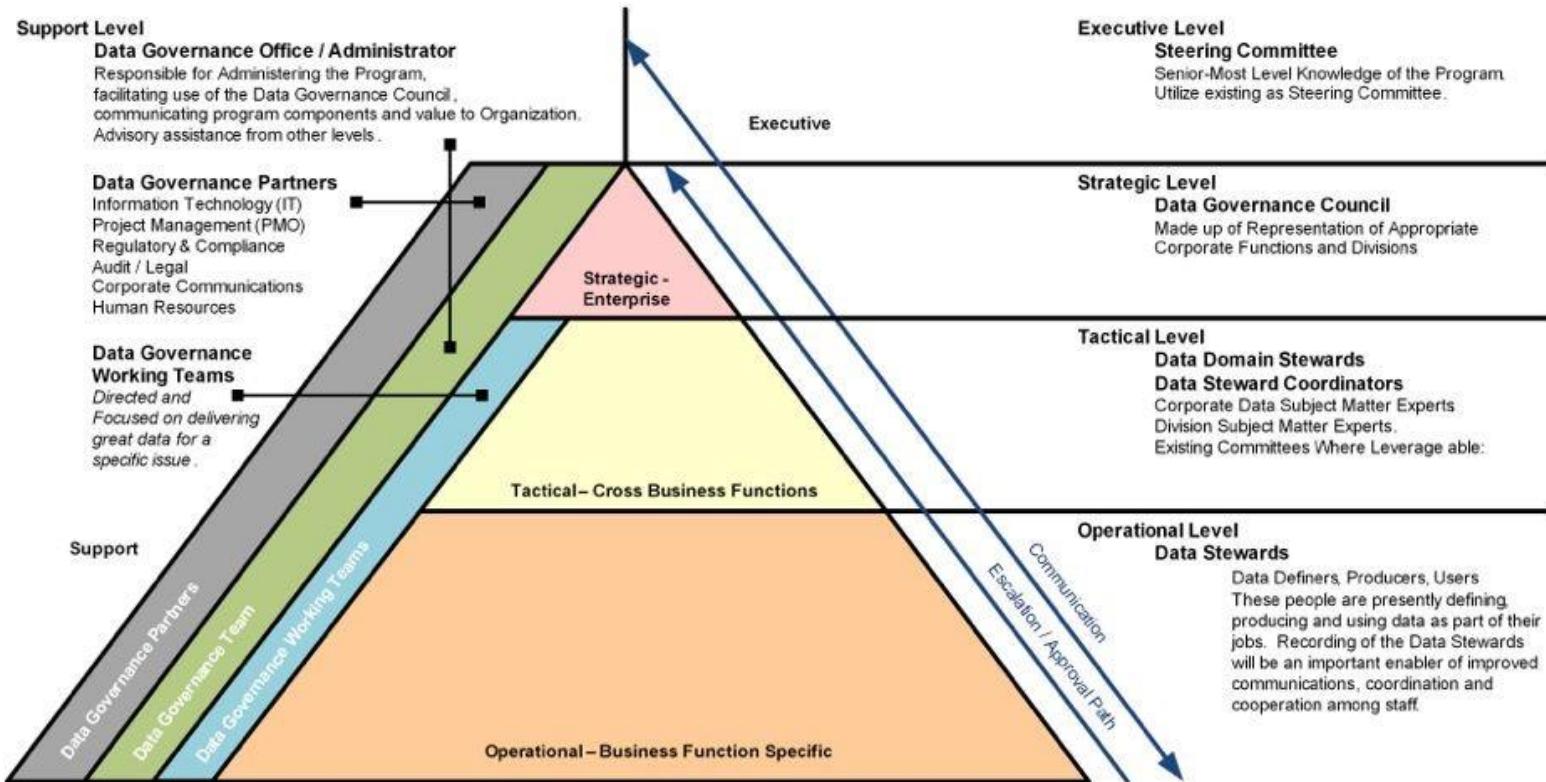


\*The U.S. relies on a “[combination of legislation, regulation and self-regulation](#)” rather than government intervention alone. There are approximately 20 industry- or sector-specific federal laws, and more than 100 privacy laws at the state level (in fact, there are 25 privacy-related laws in California alone).

# Data Governance Core Principles



# Governance Roles and Responsibilities



# Module 3 - Data Architecture

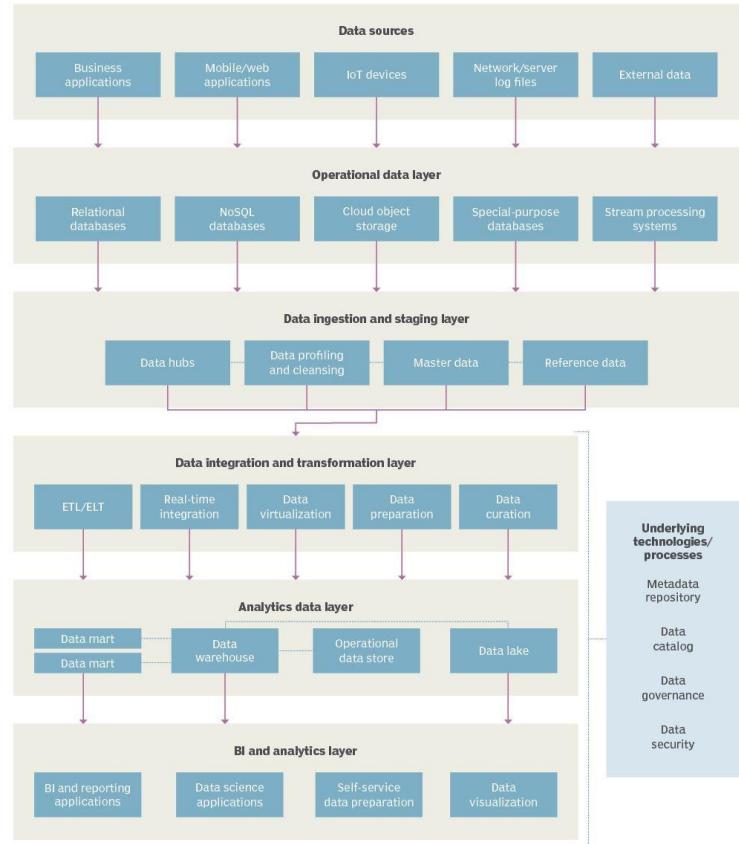
# What is Data Architecture

‘Data architecture is a set of rules, policies, standards and models that govern and define the type of data collected and how it is used, stored, managed and integrated within an organization and its database systems. It provides a formal approach to creating and managing the flow of data and how it is processed across an organization’s IT systems and applications’

# Data Architecture Example

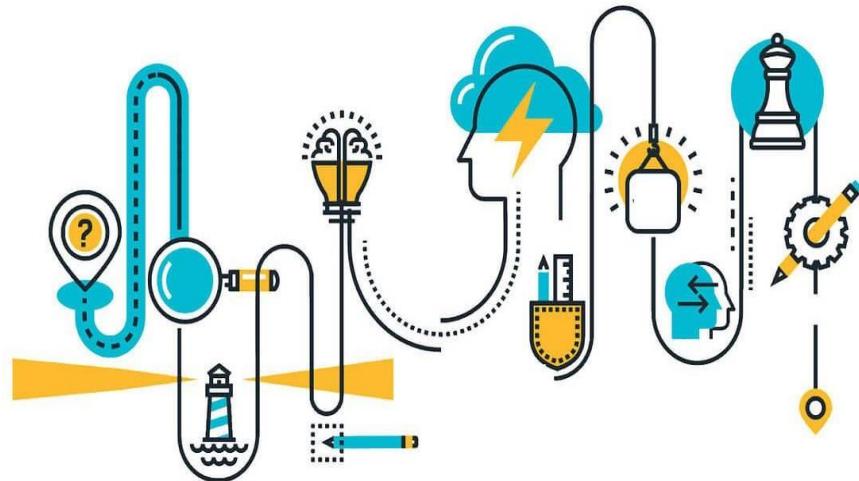
“Data architecture is a set of rules, policies, standards and models that govern and define the type of data collected and how it is used, stored, managed and integrated within an organization and its database systems. It provides a formal approach to creating and managing the flow of data and how it is processed across an organization’s IT systems and applications”

Sample data architecture diagram



# Data Architecture Principles

1. *Data is a shared asset*
2. *Users require adequate access to data*
3. *Security is essential*
4. *Common vocabularies ensure common understanding*
5. *Data should be curated*
6. *Data flows should be optimized for agility*



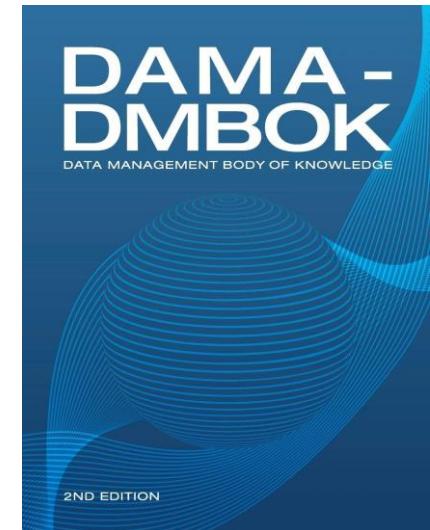
# Data Architecture components

1. *Data pipelines*
2. *Cloud storage*
3. *Cloud computing*
4. *AI and ML models*
5. *Data streaming*
6. *Container orchestration*
7. *Real-time analytics*



# Data Architecture frameworks

1. *DAMA-DMBOK-2*
2. *Zachman Framework for Enterprise Architecture*
3. *The Open Group Architecture Framework (TOGAF)*



**TOGAF®<sup>®</sup>**  
20

# Data Architecture best practices

1. *Cloud-native*
2. *Robust and scalable data pipelines*
3. *Seamless data integration*
4. *Real-time data enablement*
5. *Decoupled and extensible*
6. *Domain-driven*
7. *Balanced*



# Data Architecture Roles

Data Architect!!!

Data Modeler

Data Integration Developer

Data Engineer



# Module 4 - Data Modeling and Design

# What is Data Modeling

‘**Data modeling** is the process of creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structure.

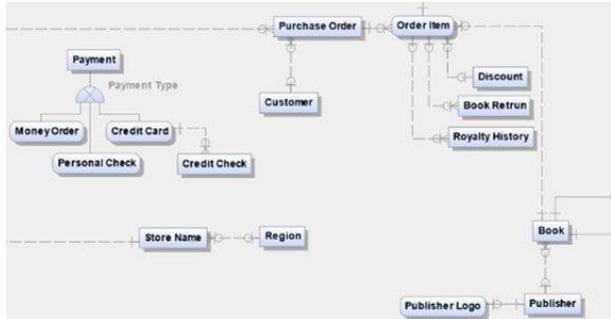
The goal is to illustrate the types of data used and stored within the system, the relationships among these data types, the ways the data can be grouped and organized and its formats and attributes.’

# Data Modeling vs Data Architecture

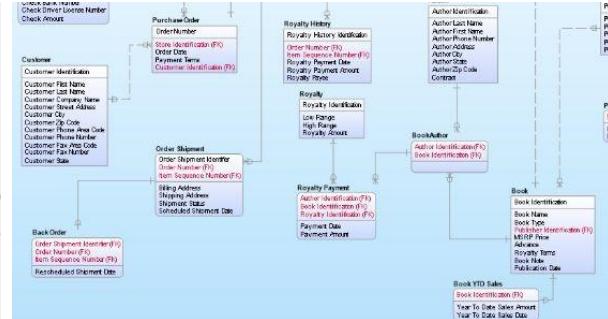
5 Key Differences	
Data Modeling	Data Architecture
Focuses on the representation of the data	Focuses on what tools and platforms to use for storing and analyzing data
Focus on accuracy of the data	Focus is on the infrastructure housing the data
Focus on reliability of the data	Focus on keeping the data safe
Representation of reality	Framework of systems and logistics
Represents a limited set of business concepts	Covers the data infrastructure of the entire organization

# The 3 Levels of Data Models

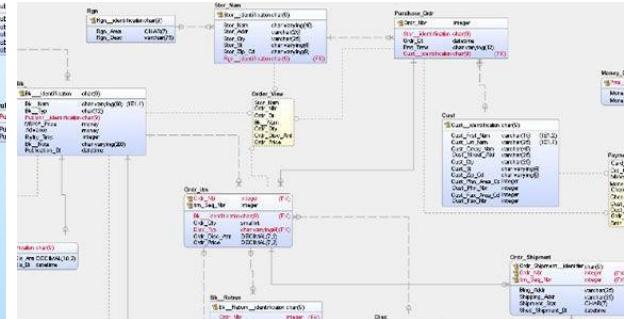
# Conceptual data model



## Logical data model



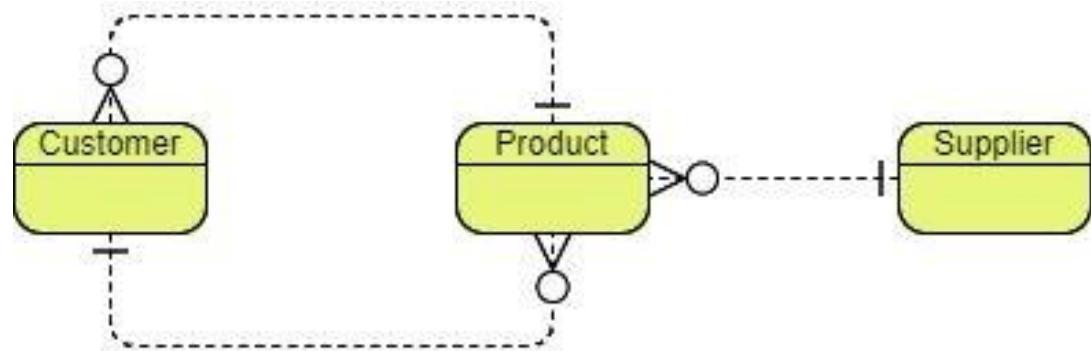
## Physical data model



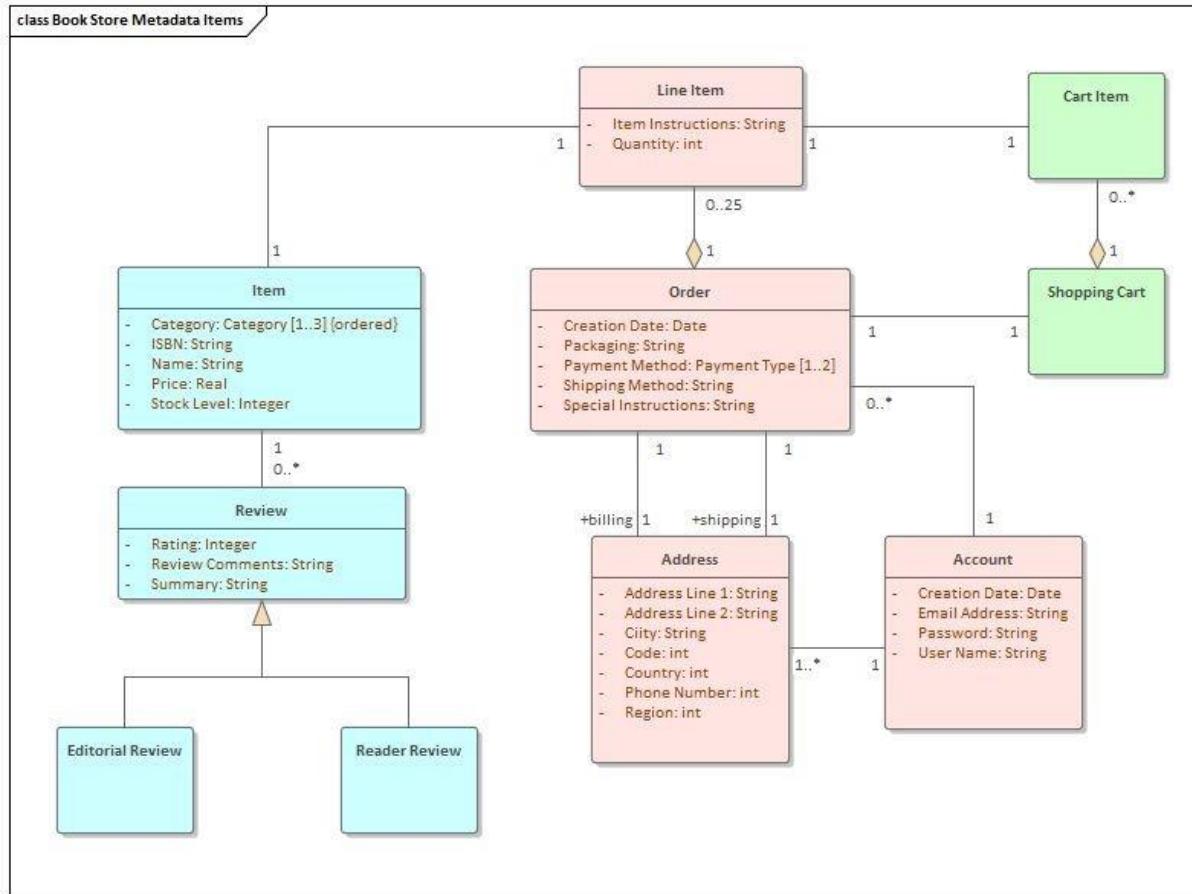
# Conceptual data model

Elements of a Conceptual data model:

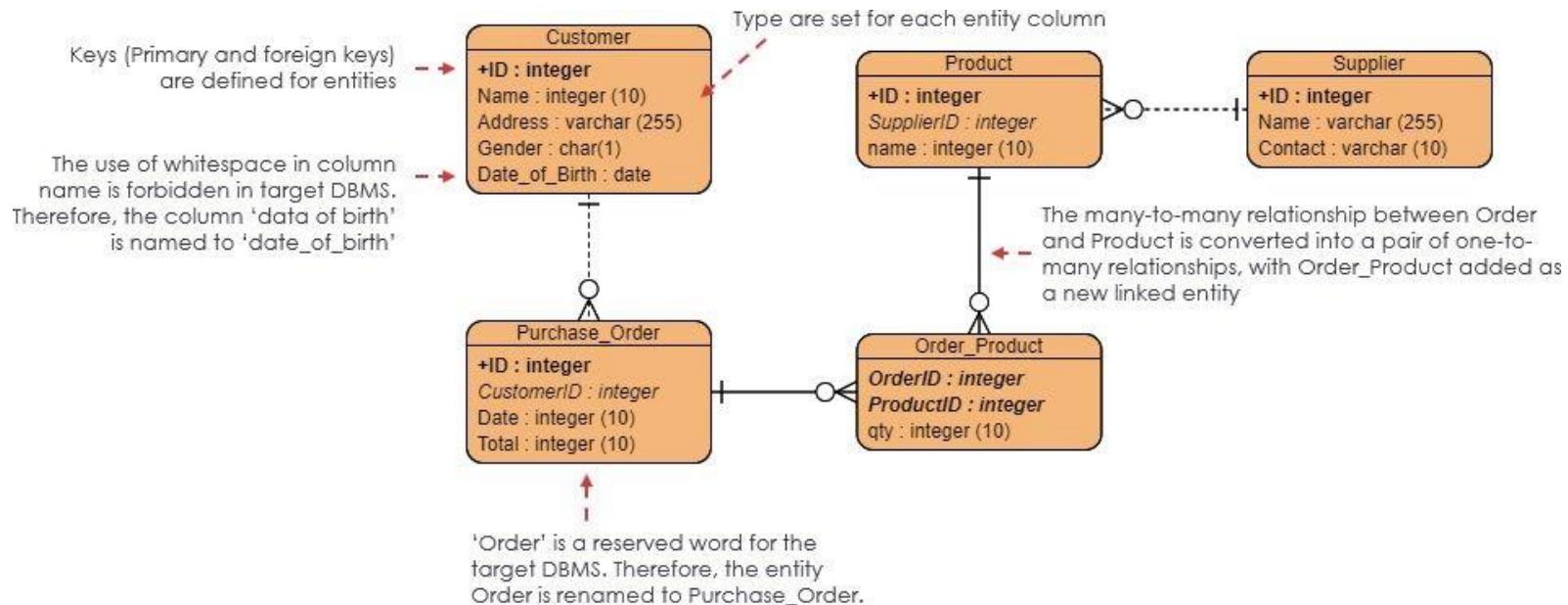
- Entity
- Attribute
- Relationship



# Logical data model

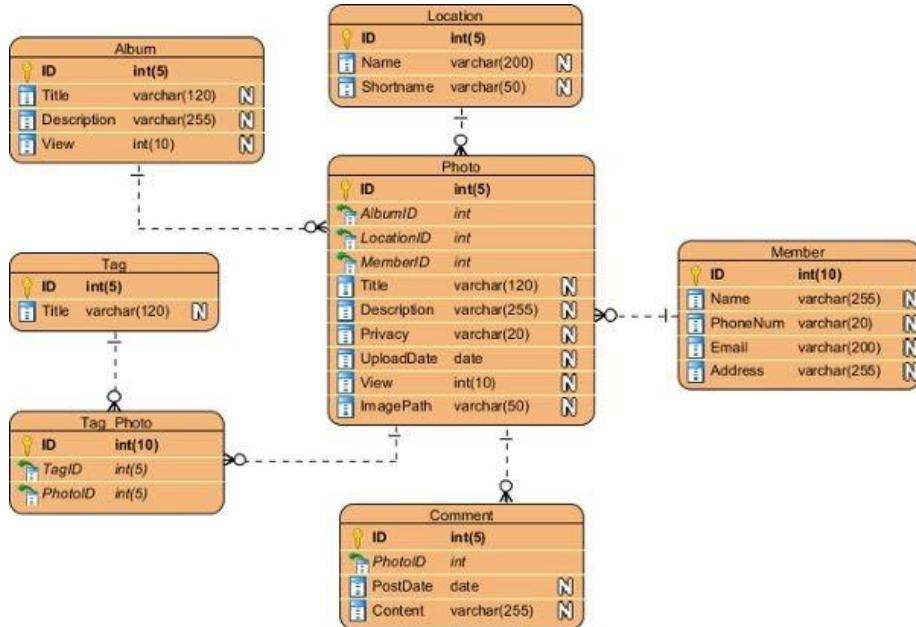


# Physical data model



# Data Modeling Process

1. Identify the entities
2. Identify the attributes of each entity
3. Identify relationships among entities
4. Map attributes to entities completely
5. Assign keys, decide on degree of normalization
6. Finalize and validate the data model



# Benefits of Data Modeling

- Reduce errors in software and database development
- Reduced cost
- Better documentation
- Improve database performance
- Improved communication between developers and BI teams
- Ease and speed the process of database design



# Data Modeling Tools

- erwin Data Modeler
- ER/Studio
- DbSchema
- ERBuilder
- HeidiSQL
- Navicat Data Modeler
- Toad Data Modeler
- SQL Database Modeler



# Module 5 - Data Storage and Operations

# What is Data Storage and Operations

“**Data storage and Operations** includes the design, implementation, and support of stored data, to maximize its value throughout its lifecycle, from creation/acquisition to disposal”

Two sub-activities:

1. Database support
2. Database technology support

# Benefits of good Data Storage Strategy

1. *Reduce capital expenses*
2. *Reduce operational expenses*
3. *Easier data management*
4. *Optimized resource utilization*
5. *Easier scalability*
6. *Better performance*
7. *Better user experience*



# Data Storage activities

Monitoring  
Hardware and  
Software assets



Resolving  
performance issues



Relocating storage  
resources



Process automation,  
memory management,  
storage provisioning



Track storage activities  
and metrics



Report on KPIs  
related to storage



# Data Storage Management key attributes

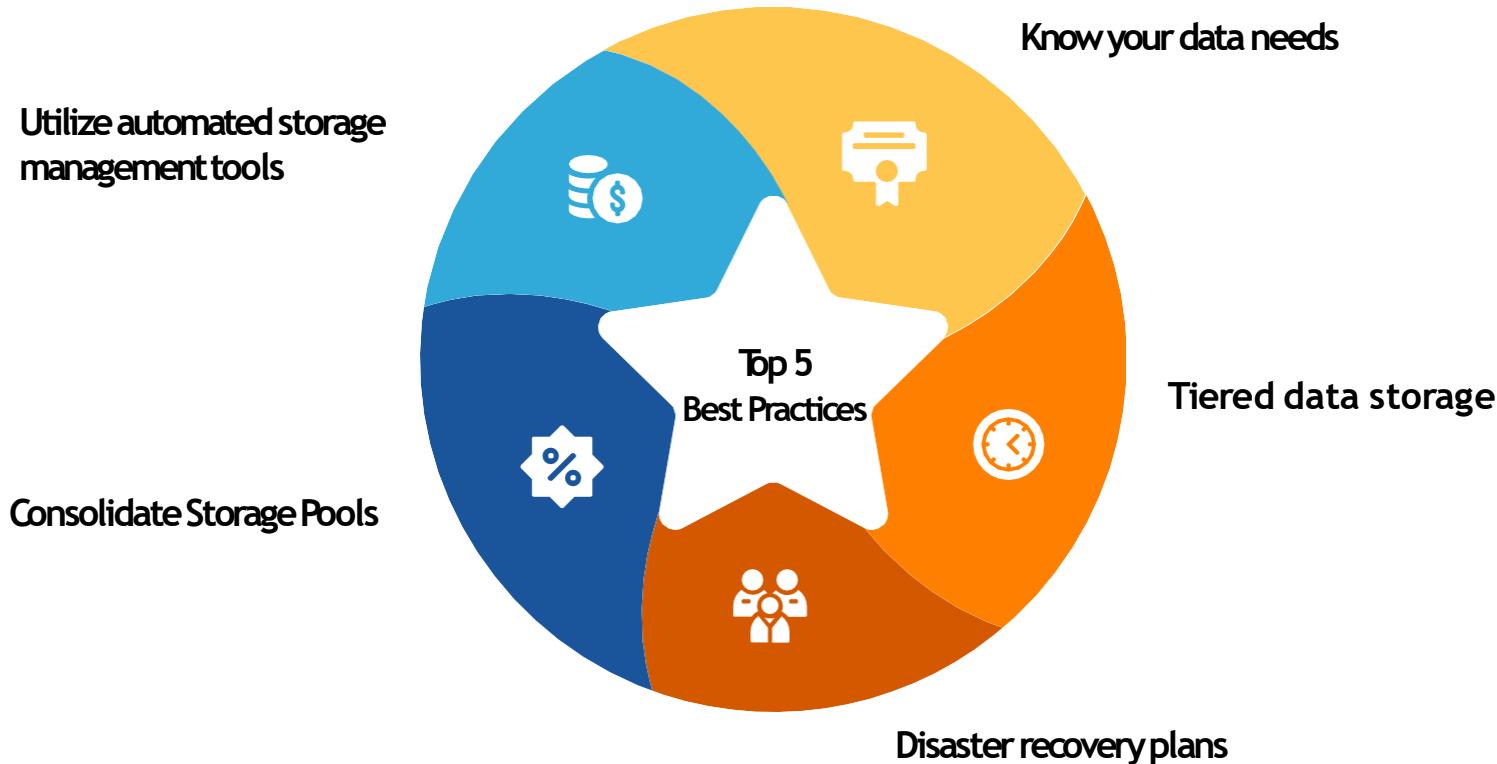
Performance

Reliability

Recoverability

Capacity

# Data Storage Best Practices



# Module 6 - Data Security

# What is Data Security

‘**Data Security** is the process of protecting digital information from unauthorized access, corruption, or theft throughout its entire lifecycle.’

# Why is Data Security important

- The average total cost of a ransomware breach is \$4.62 million, slightly higher than the average data breach of \$4.24 million ([IBM](#))
- The average per record (per capita) cost of a data breach increased by 10.3 percent from 2020 to 2021 ([IBM](#))
- The average cost of a breach with a lifecycle over 200 days is \$4.87 million ([IBM](#))
- 39 percent of costs are incurred more than a year after a data breach ([IBM](#))
- In 2021, the United States was the country with the highest average total cost of a data breach was at \$9.05 million ([IBM](#))
- 34 percent of data breaches in 2018 involved internal actors ([Verizon](#))
- It took an average of 287 days to identify a data breach ([IBM](#)).

# Types of Data Security

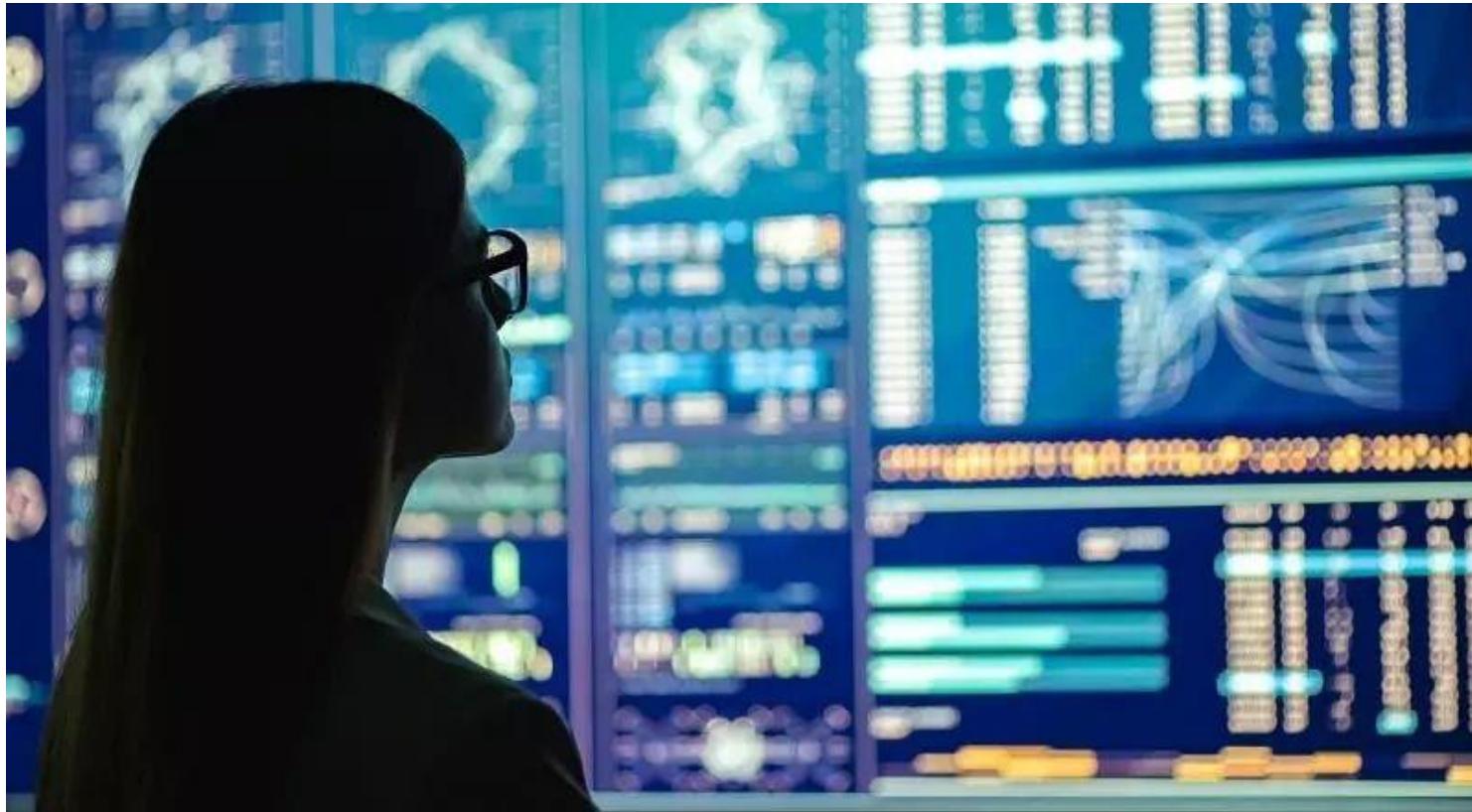


# Data Security Risks

- Accidental Data Exposure
- Phishing
- Malware
- Insider Threats
- Password Attack
- Denial-of-Service (DOS)
- Man-in-the-Middle (MITM)
- SQL Injections
- Zero-day Exploit



# Accidental Data Exposure



# Phishing



# Malware

## Types of Malware

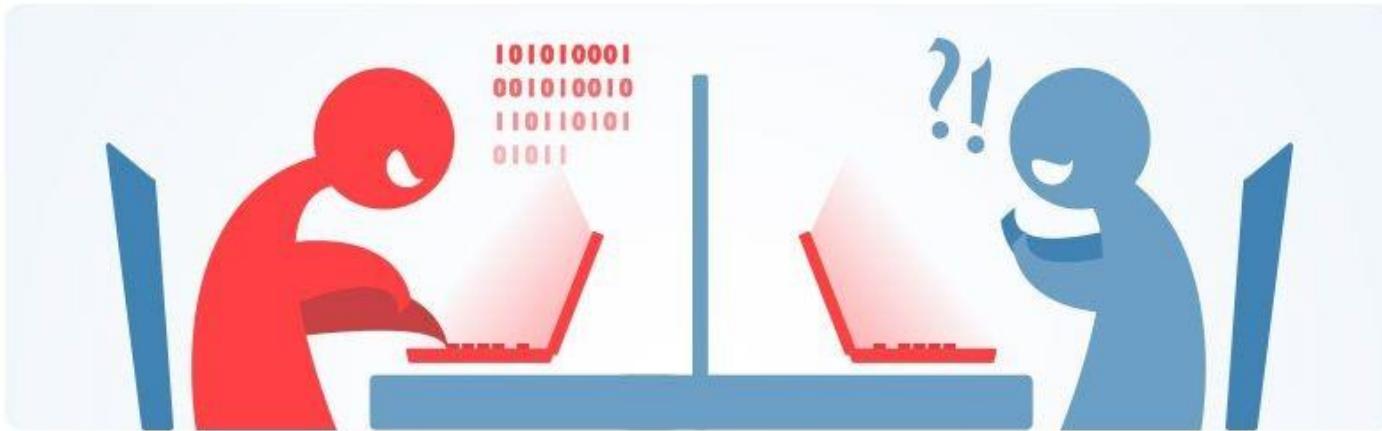


Source:

<https://us.norton.com/internetsecurity-malware-types-of-malware.html>

# Insider Threats

## Insider threat classification by CA Technologies



### Malicious insiders

Intentionally use their access to sensitive data to harm the company

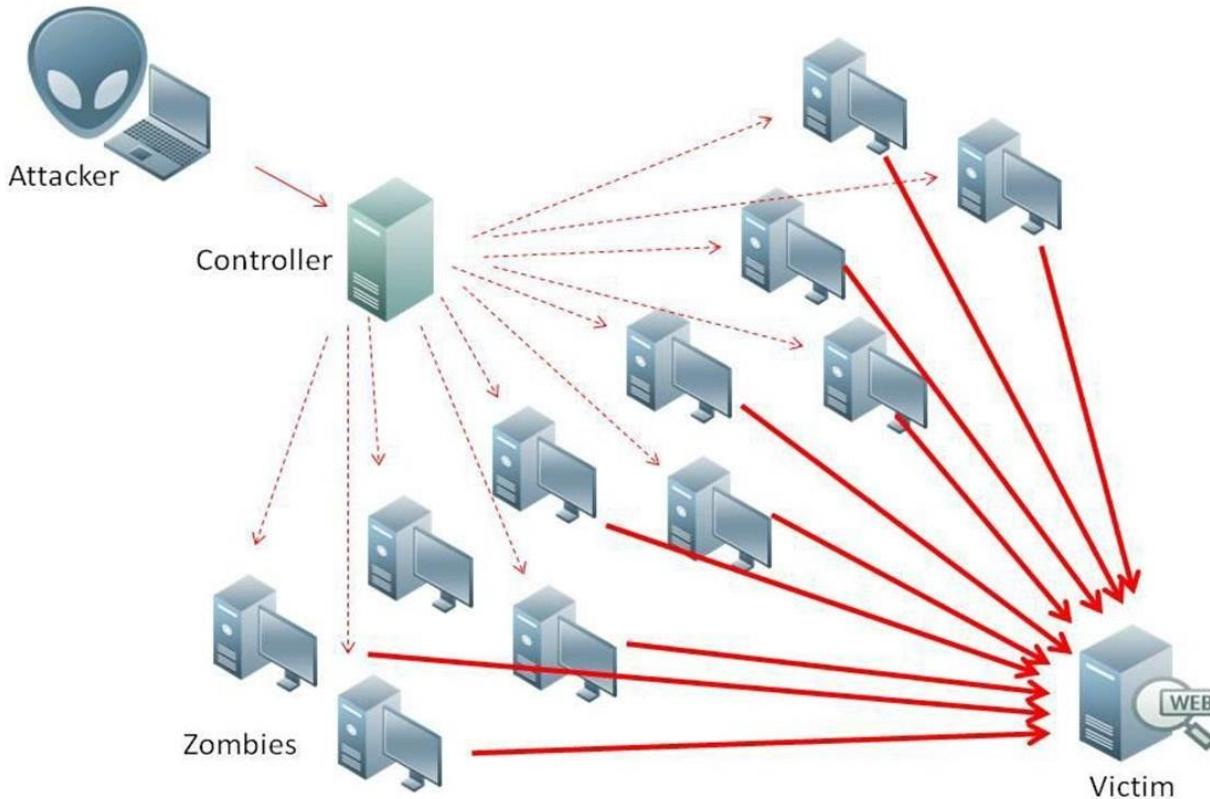
### Inadvertent insiders

Cause damage to the company unintentionally

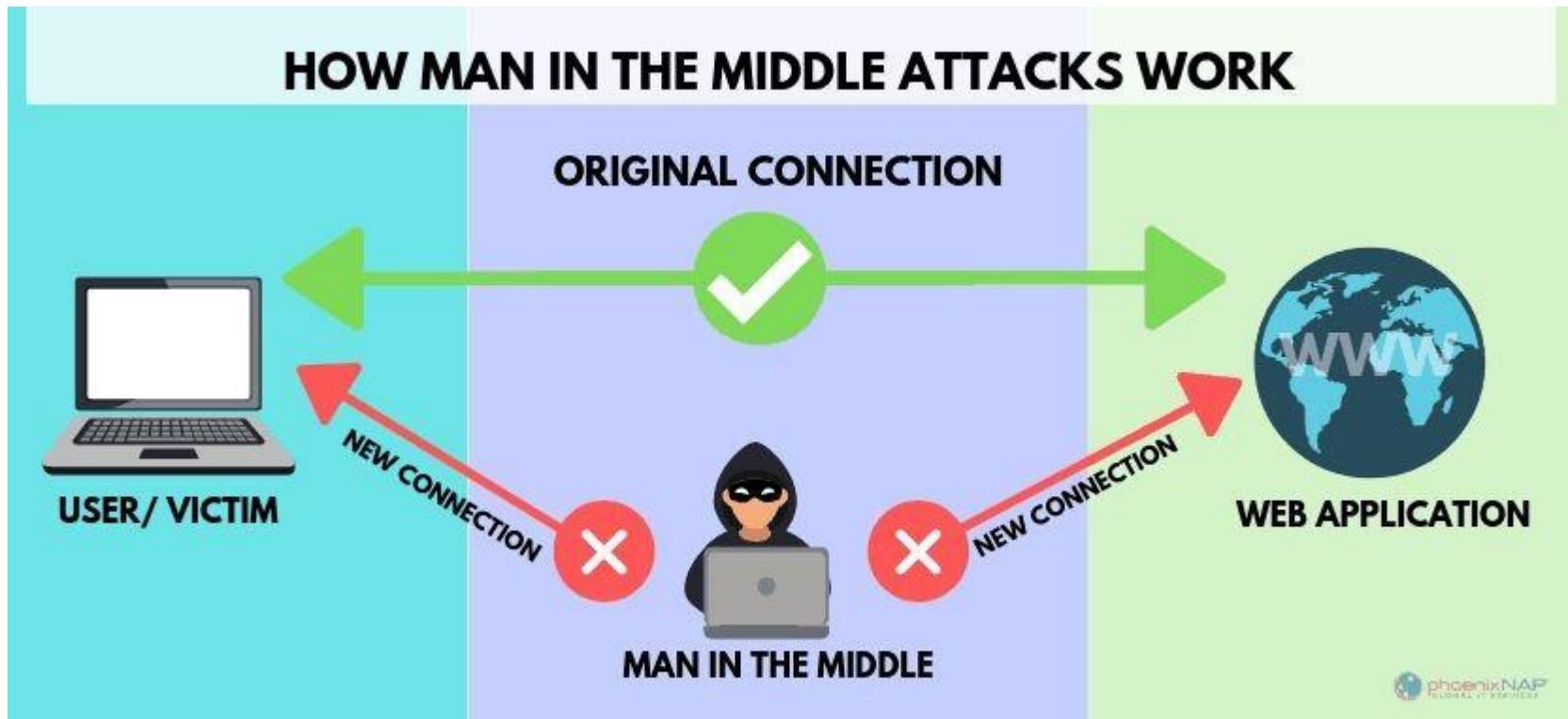
# Password Attacks



# Denial-of-Service (DOS)



# Man-In-The-Middle

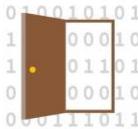


# SQL Injections



# Zero-Day Exploit

## 'Zero-Day' Defined

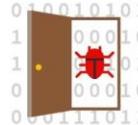


010010101  
1 00010  
1 01101  
0 00010  
000111011

A **zero-day vulnerability** is a security software flaw that's unknown to someone interested in mitigating the flaw.



A **zero-day attack** is when hackers leverage their zero-day exploit to commit a cyberattack.



010010101  
1 00010  
1 01101  
0 00010  
000111011

A **zero-day exploit** is when hackers take advantage of a zero-day vulnerability for malicious reasons.

# 9 Best Practices to secure your data

1. Employees education
2. Create Insider Threat Policies
3. Phishing Simulations
4. Backup data
5. Update Systems and Software
6. Utilize HTTPS
7. Maintain Compliance
8. Use multi-factor authentication
9. Employ latest secure coding practices



# Module 7 - Data Integration

# What is Data Integration

‘**Data Integration** is the process of consolidating data from different sources into one, unified view for efficient data management’

# Example of Data Integration

Company A uses the following:

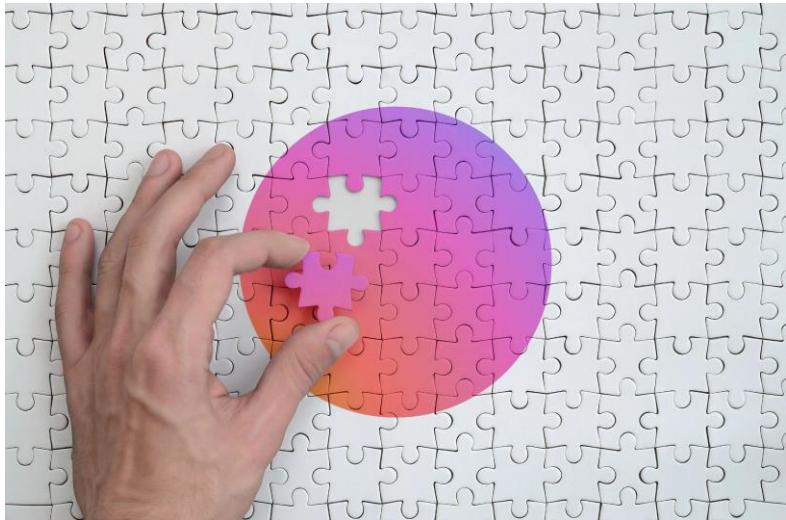
- Salesforce for customer information and sales pipeline data
- External vendor for additional customer firmographics
- Internal database that tracks customer satisfaction ratings from surveys
- Internal financial system that tracks the sales revenue per customer
- Marketing department database on customer campaigns



**Goal: Integration all the above sources into single source!**

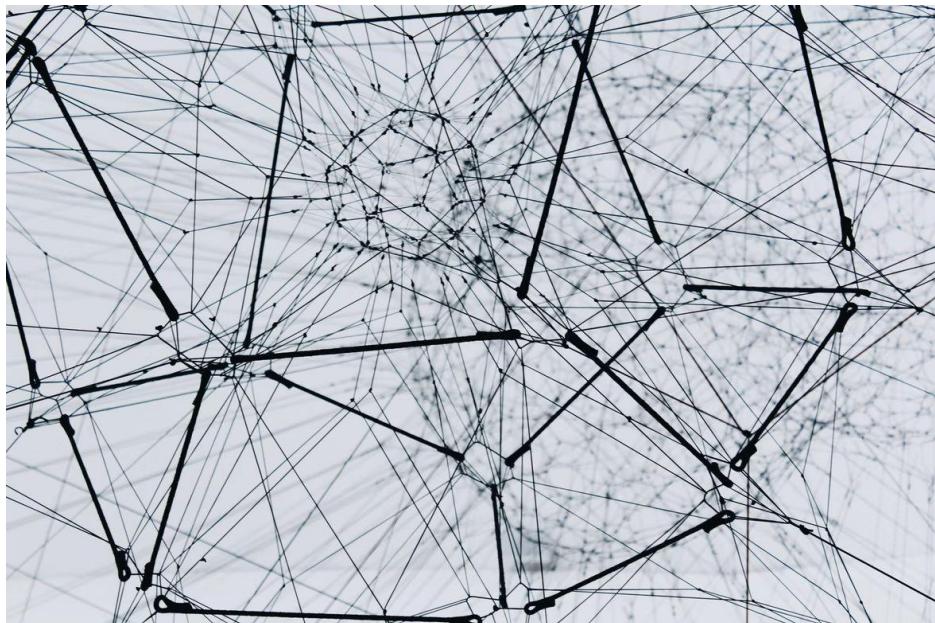
# Importance of Data Integration

- Complete view of business intelligence, insights and analytics
- Increased efficiency and ROI - no need for manual data gathering
- Better employee, customer and partner experience
- Improves Collaboration
- Eliminates Data Silos
- Reduces errors
- Faster innovation, sales, time to market
- Improves Data Quality and Integrity



# Techniques for Data Integration

- Manual Data Integration
- Middleware Data Integration
- Application Based Integration
- Uniform Access Integration
- Common Storage Integration (Data Warehousing)
- Data Virtualization



# Manual Data Integration

Pros:

- Low cost
- Total control

Cons:

- Difficult to scale
- Human error



# Middleware Data Integration

Pros:

- Better data streaming
- Easier access between systems

Cons:

- Maintenance
- Limited functionality



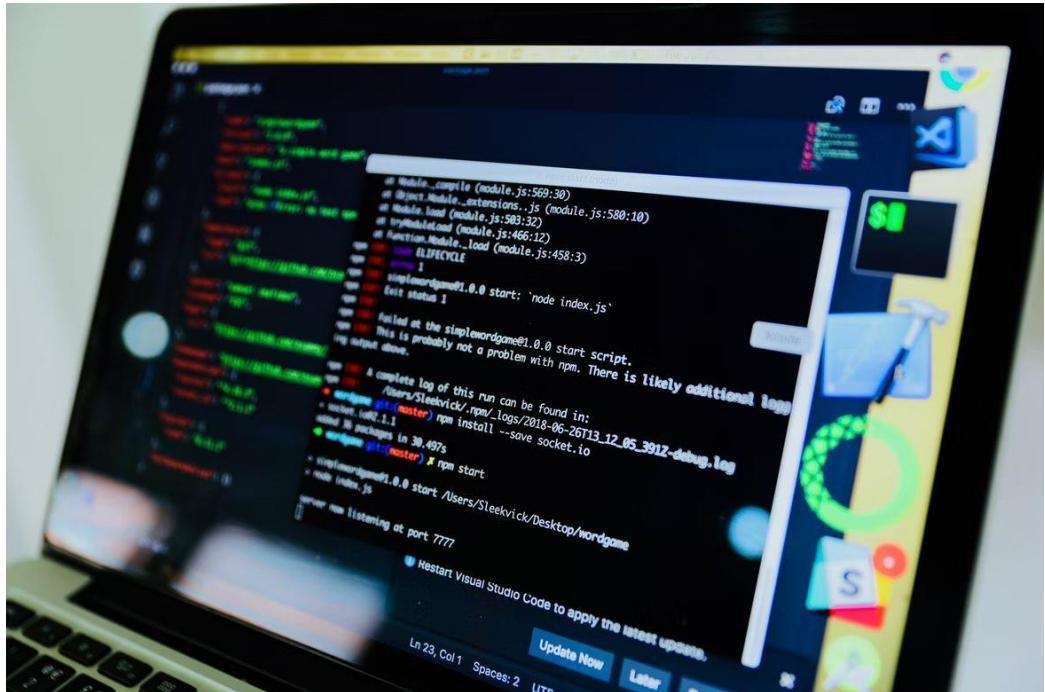
# Application Based Integration

Pros:

- Simplified processes
- Easier information exchange
- Fewer resources used

Cons:

- Maintenance
- Inconsistent results
- Complicated setup
- Difficult data management



# Uniform access Integration

Pros:

- Lower storage requirements
- Easier data access
- Simplified view of data

Cons:

- Data integrity challenges
- Strained systems



# Common Storage Integration (Data Warehousing)

Pros:

- Reduced burden
- Increased data version management control
- Cleaner data
- Enhanced data analytics

Cons:

- Increased storage costs
- Higher maintenance costs



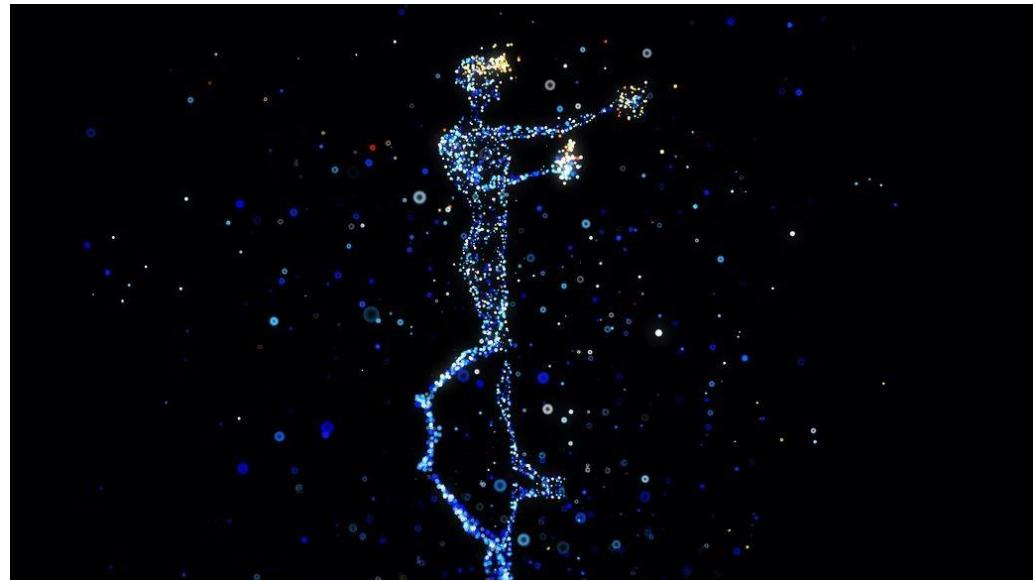
# Data Virtualization

Pros:

- No need to move data
- Scalable
- No need to maintain data in multiple locations

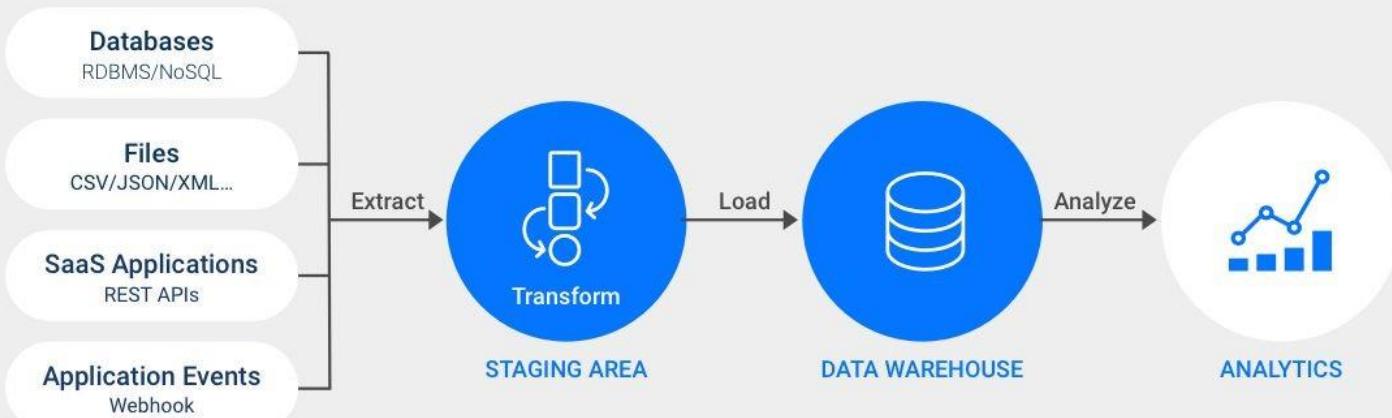
Cons:

- Cost



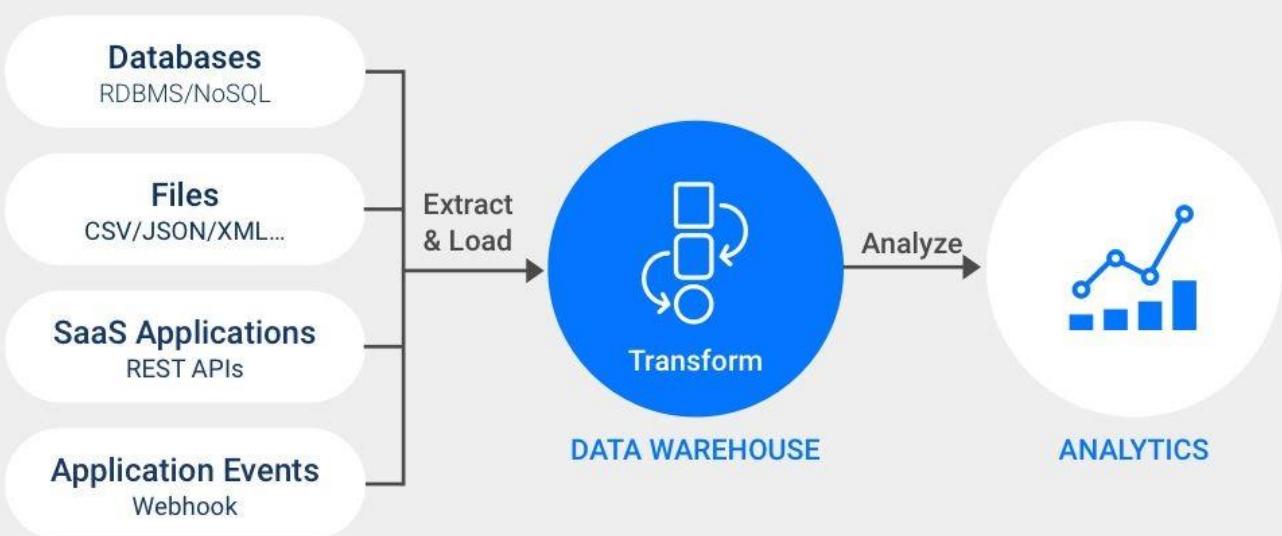
# ETL

## ETL PROCESS





# ELT PROCESS



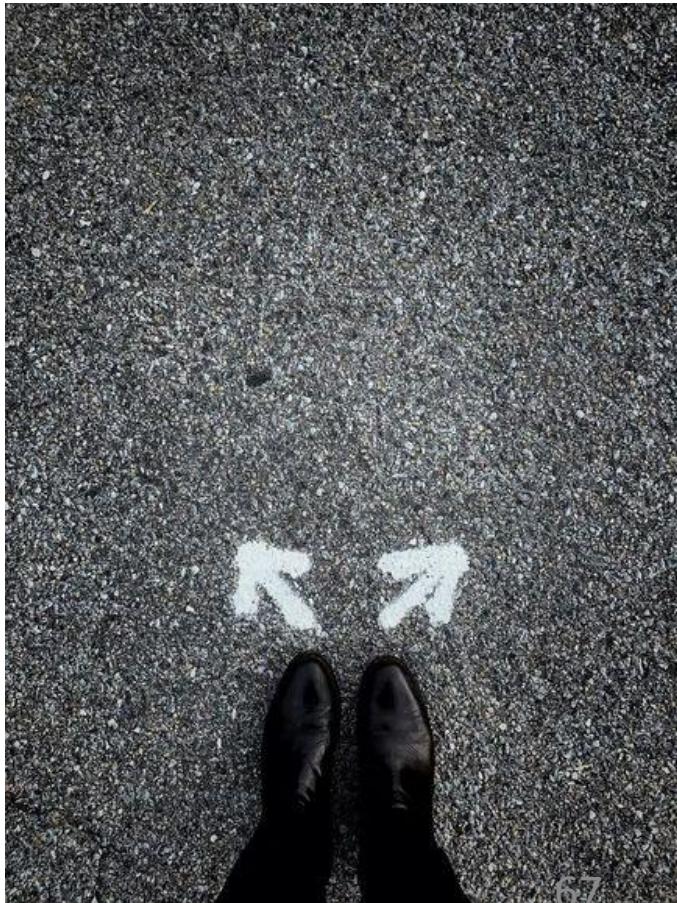
# ETL vs ELT

## Pros of ETL:

- **Compliance** - ETL is better for compliance with GDPR, HIPAA, CCPA and other standards
- **Implementation** - many ETL tools and experts
- **Maturity** - more documentation, tools and best practices

## Pros of ELT:

- **Maintenance** - all data is always flowing with automated process
- **Speed** - data available in Data Warehouse faster since no transformation layer
- **Cost** - only cloud-based platforms needed at lower cost. On premise ETL processes require expensive hardware



# Data Integration Tools

## On-premise tools

- Oracle Data Service Integrator
- Informatica PowerCenter
- IBM InfoSphere Information Server

## Cloud-based tools

- SnapLogic
- Talend Cloud Integration

## Open-source tools

- Talend Open Studio
- Tibco Jaspersoft



# Data Integration Best Practices

1. Identify your business needs first
2. Include internal business expert in the data integration team
3. Consider Long-Term Goals
4. Take into consideration the total cost of all methods
5. Avoid very complex data integration solutions
6. Choose a flexible solution



# Module 8 - Document & Content Management

# What is Document & Content Management

**‘Document & Content management** is the process of establishing planning, implementation and control activities for lifecycle management of data and information found in any form or medium - outside of relational databases”

# Why do we need Document & Content Management?

- Comply with legal obligations
- Comply with customer expectations regarding records management
- Effective and efficient storage, retrieval and use of documents and content
- Integration between structured and unstructured content



# What is a DMS?

- DMS stands for Document Management System
- What are the benefits?
- Types of DMS



# What is a CMS?

- CMS stands for Content Management System
- What are the benefits?
- Types of CMS



# ECMS

- ECMS stands for Enterprise Content Management System
- What are the benefits?
- CMS vs ECM



Figure 1: Magic Quadrant for Content Services Platforms



Source: Gartner (October 2021)

# Document Management vs Enterprise Content Management

Comparison	Document Management System (DMS)	Enterprise Content Management System (ECMS)
Type of Data	Structured data in traditional formats (Word, PDF, PowerPoint, Excel, etc)	Structured + unstructured data such as images, audio, video files, HTML, etc
Main purpose	Workflow management and regulatory compliance	Storage, retrieval and publishing of content
Key difference	DMS is a software	ECM is a set of tools and processes. ECM is a broader version of DMS
Company size	DMS only solution can work well for small companies	ECM solution needed in bigger organizations

# Module 9 - Master & Reference Data Management

# What is Master Data

**DAMA Guide to Data Management Body of Knowledge:** “Master Data represents data about the business entities that provide context for business transactions”

**Gartner:** “Master Data is the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies and chart of accounts”

# **Master+Data.xlsx**

# What is Reference Data

**DAMA Guide to Data Management Body of Knowledge:** “Reference data is data used to classify or categorize other data”

Examples of Reference Data:

- Postal codes
- Language codes
- Customer segments
- Country codes
- Cost centers

# Master+Reference Data

# Master Data vs Reference Data

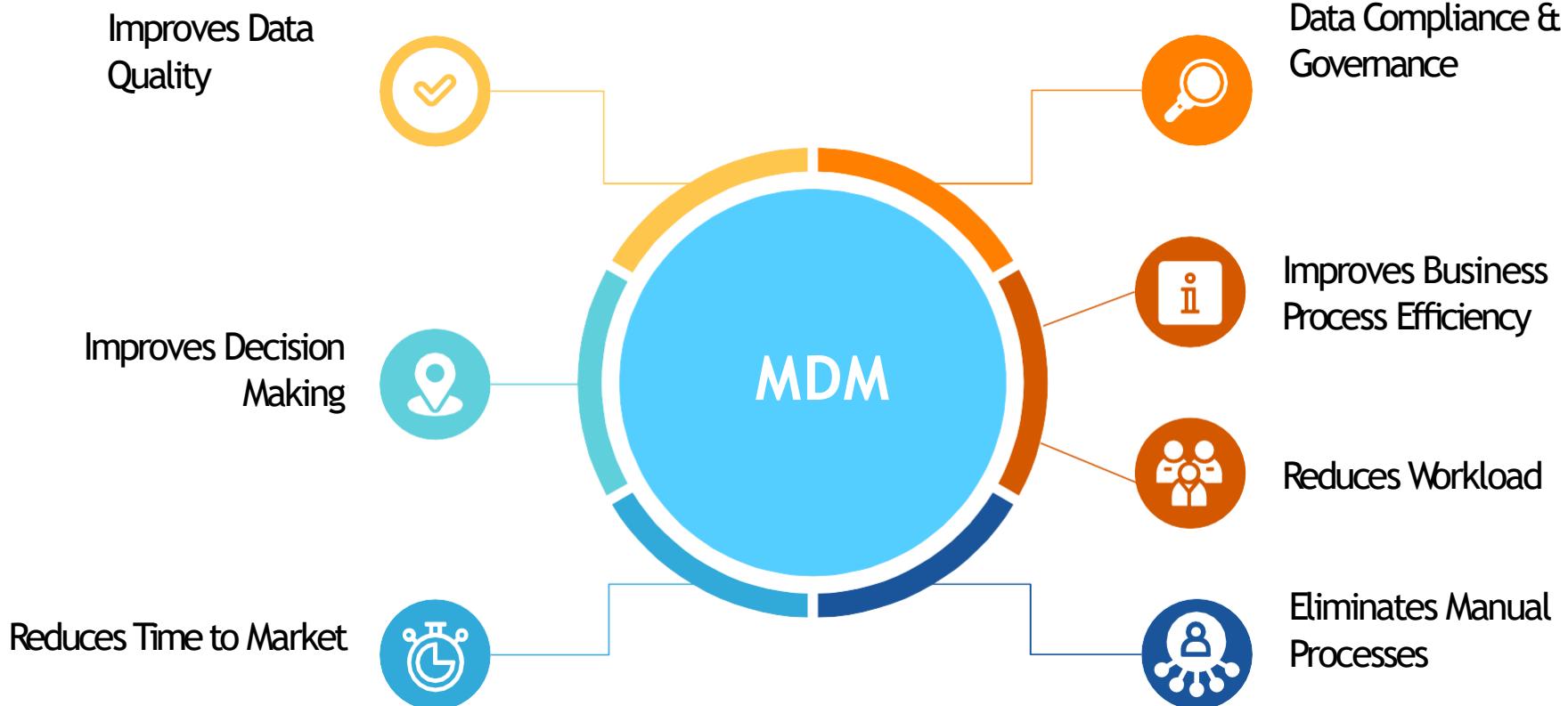
Comparison	Master Data	Reference Data
Main purpose	Represents the business objects which contain the most valuable, agreed upon information shared across the organization	Data that defines the set of permissible values to be used by other data fields
More on usage	Master data is the data shared by multiple systems, applications, processes in the organization	Reference data is a type of master data that is used by other data fields
Examples	<ul style="list-style-type: none"><li>Customer information - names, phone numbers and addresses</li><li>Product information - product name and location</li><li>Partner data - partner name and address</li></ul>	<ul style="list-style-type: none"><li>Fixed conversion rates - weight, temperature, length, etc</li><li>Currency codes</li><li>Language codes</li><li>Customer Segments</li><li>Cost centers</li><li>Postal codes</li><li>Units of measurement</li></ul>

# What is Master Data Management (MDM)

**Master Data Management** is the process of creating and maintaining a single master record - or single source of truth - for each person, place, and thing in a business.

Through MDM, organizations gain a trusted, current view of key data that can be shared across the business and used for better reporting, decision-making, and process efficiency.

# Why MDM is important



# Steps to implement MDM



- Learn about MDM:
  - Important Capabilities
  - Implementation Styles
  - Domain Categories

Analyze your current data environment

Define your business goals

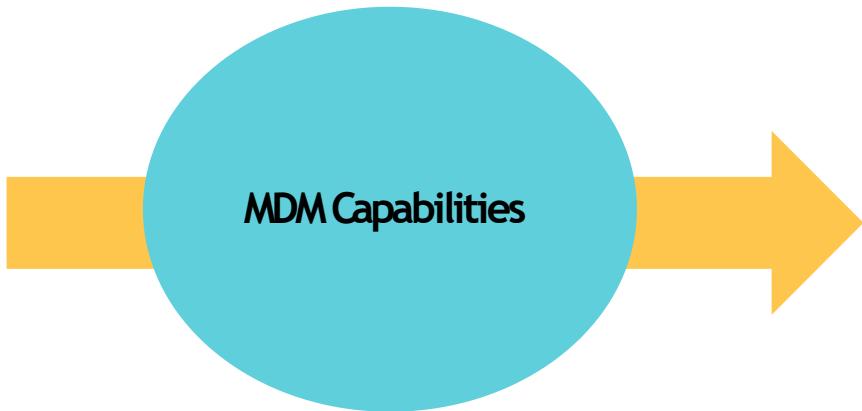
Secure the right sponsorship

Get the right team and tool

Implement a quick win

Repeat Step 6 and monitor

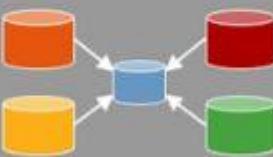
# MDM Solution Capabilities



- Workflow/BPM
- Loading/Sync/Business Services
- Data Modeling
- Information Quality/Semantics
- Perform/Scale/Availability/Security
- Hierarchy Management
- Data Stewardship
- Data Governance
- Multiple Implementation Styles
- Multiple Usage Scenarios
- Multiple Domain and Multidomain
- Product Suite Internal Integration

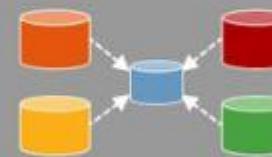
## MDM Implementation Styles

### Consolidation



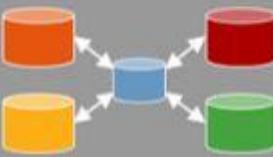
- Ideal for reporting or analytics that reside in a BI/data warehouse
- Nonintrusive to the business
- BI is the business platform
- Any industry
- Benefits dependent on success of BI strategy
- No attempt to clean up source data

### Registry



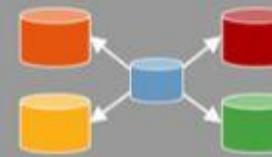
- Low control, autonomous environments
- Nonintrusive to edge applications
- Emphasis is on remote data and application-to-application integration (lots of real-time network access)
- Distributed governance
- Faster to implement than coexistence and centralized

### Coexistence



- Large-scale distributed model
- Largest change to information infrastructure
- Greatest need to mirror data
- Global and local governance
- Greatest risk over control, security
- Focused on shared services

### Centralized

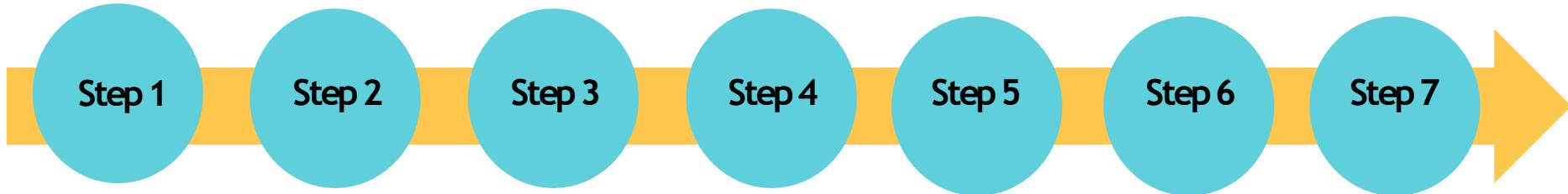


- High-control, top-down environments
- Largest change to application infrastructure
- Hugely invasive to the business
- Centralized governance
- Greatest control over access, security
- Focus on common services

# MDM Domains



# Steps to implement MDM



## Learn about MDM:

Important Capabilities  
Implementation Styles  
Domain Categories

Analyze your current  
data environment

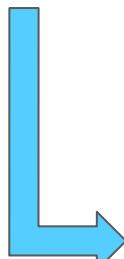
Define your business  
goals

Secure the right  
sponsorship

Get the right team and  
tool

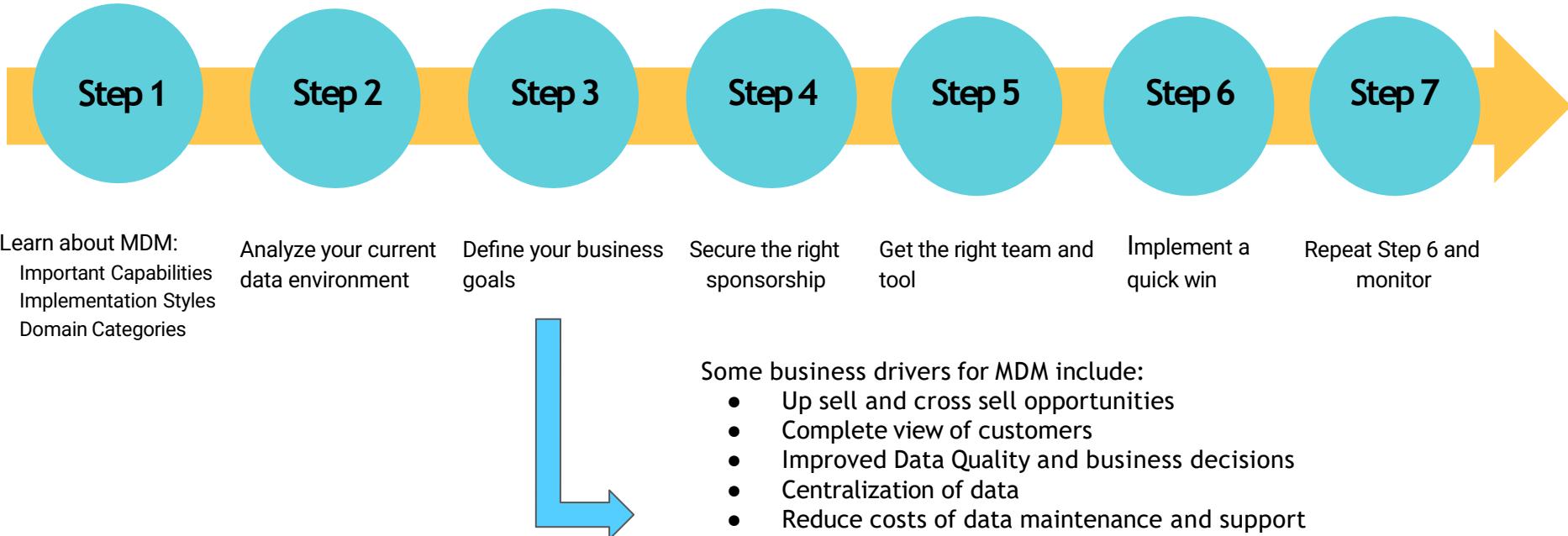
Implement a  
quick win

Repeat Step 6 and  
monitor

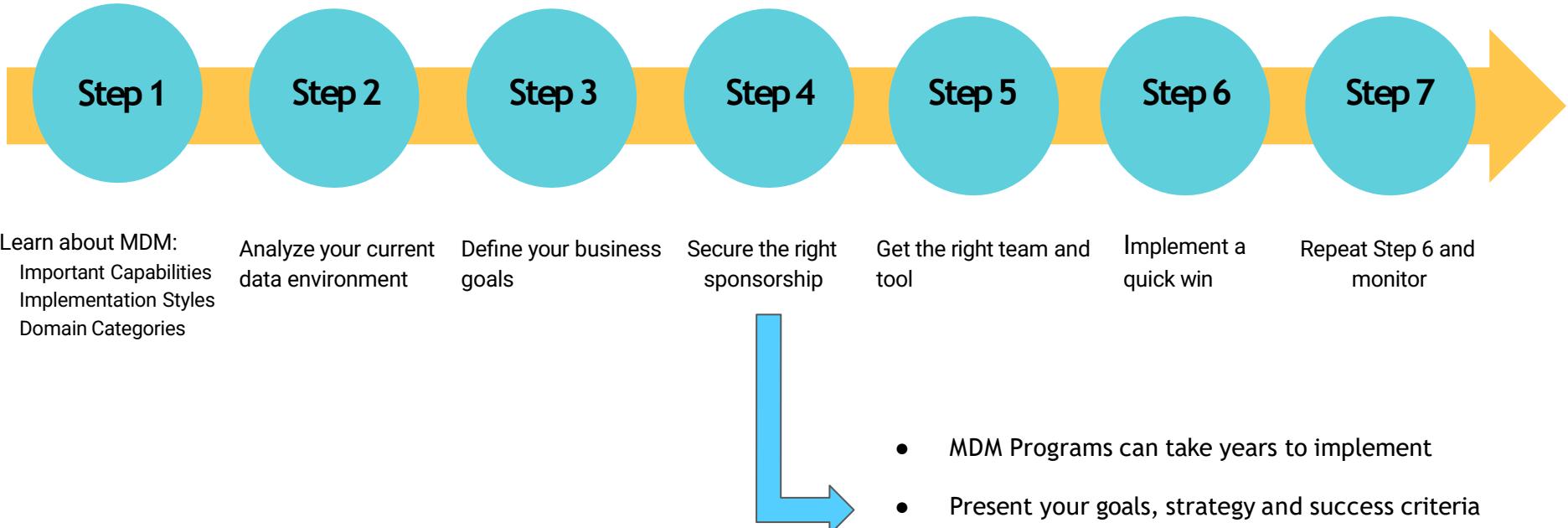


1. What are the different company departments?
2. What kind of data do they use?
3. Where is the data coming from?
4. Is it well integrated with the company tools/apps and central repositories?
5. Any data silos?
6. What are the main data problems for the business?
7. Are there data quality controls in place?
8. How is the data being governed?
9. Who are the main data stewards in the department?
10. Any documentation that will help the MDM program?

# Steps to implement MDM

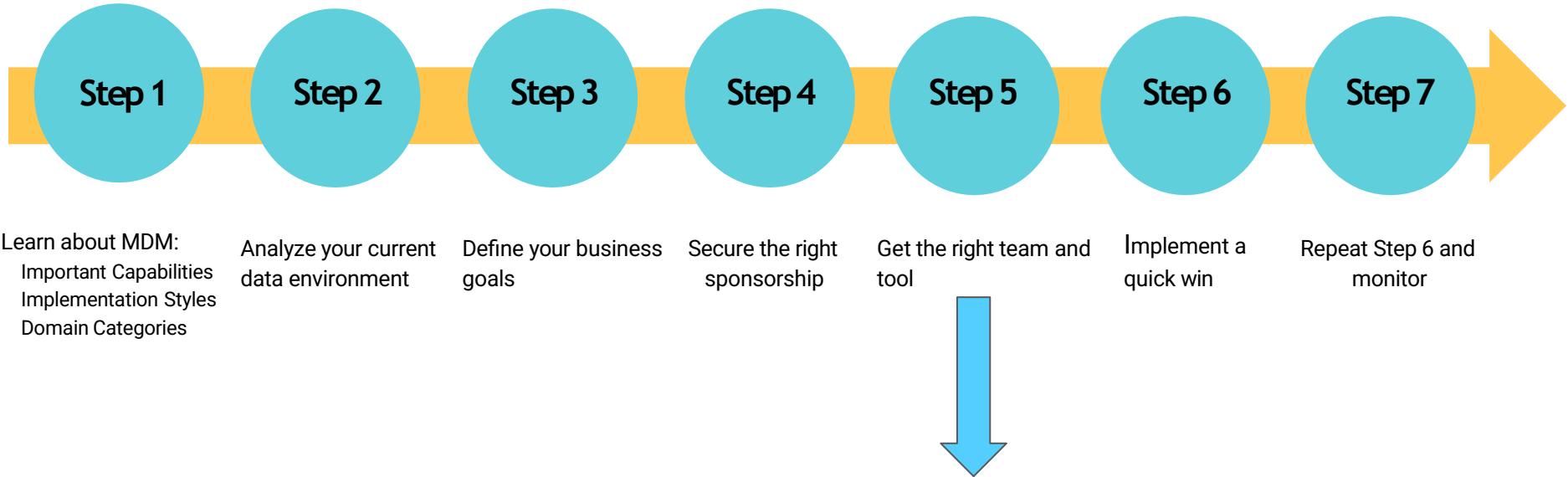


# Steps to implement MDM



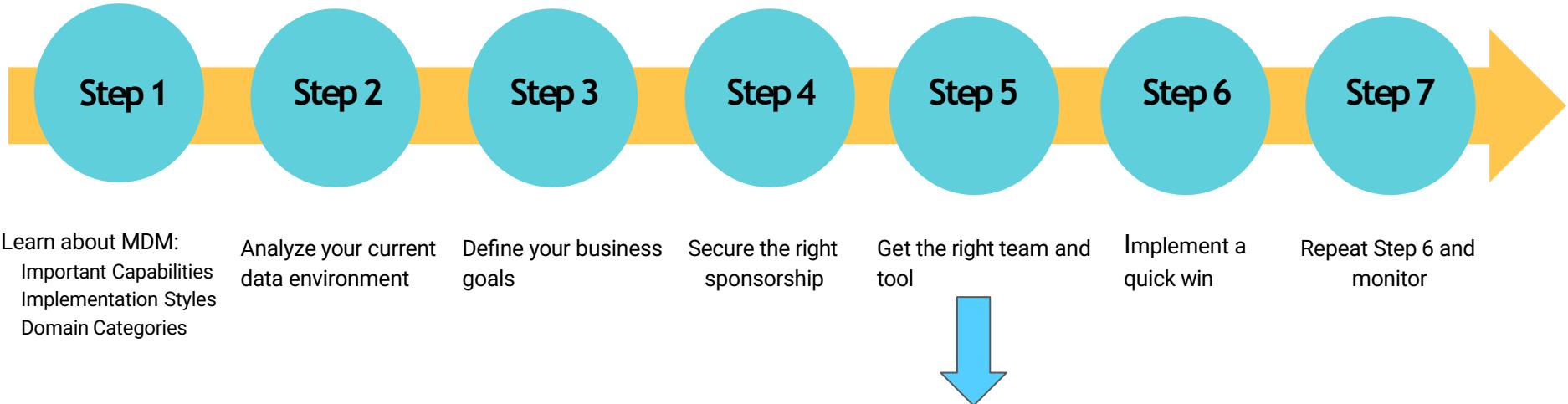
- MDM Programs can take years to implement
- Present your goals, strategy and success criteria
- Go to the top for long term budget commitment

# Steps to implement MDM



- Get the Right Team:
- **Business Team** - \*Sponsor (already covered), business analysts/SME, end users/data stewards
  - **Program Team** - MDM Specialist, Data Architects, Program Manager
  - **Tech Team** - DBA, Developers, Integration Experts, System Admins

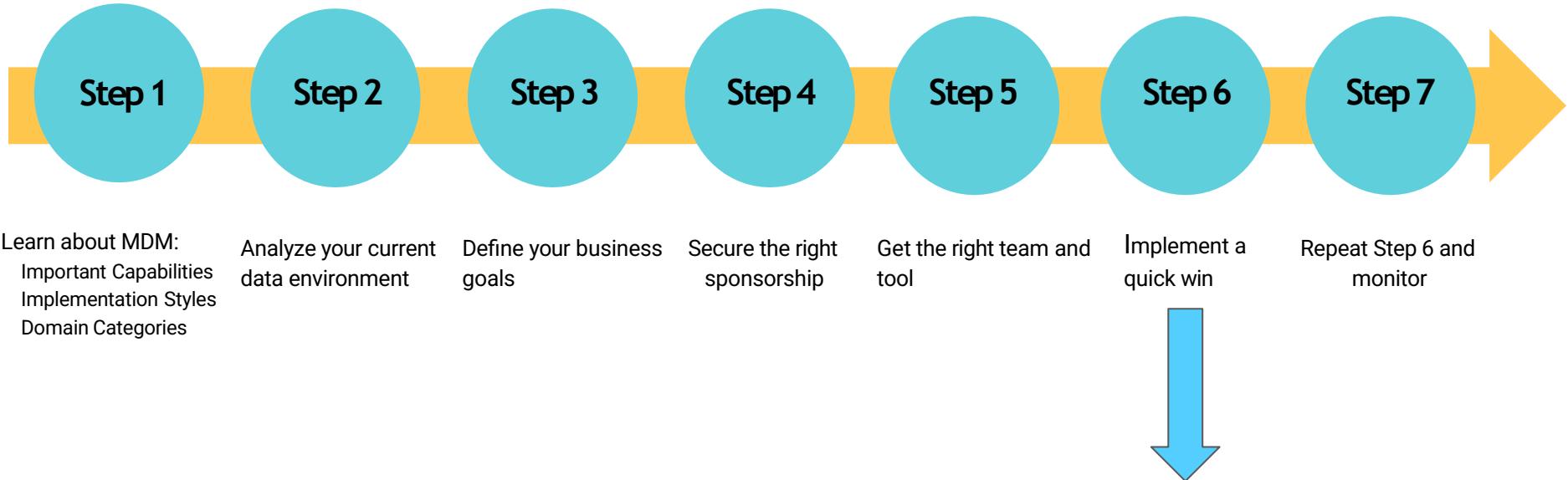
# Steps to implement MDM



**Get the Right MDM Solution** (Master data management software that consists of policies, governance, standard tools and processes that facilitate defining and managing organization's data from a single point):

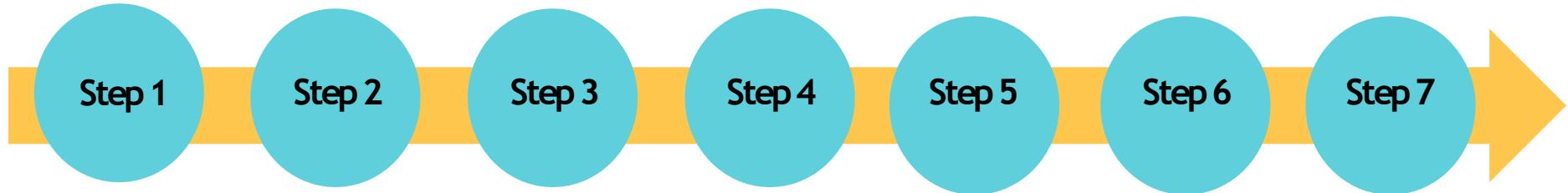
1. Understand your needs - features, functionality and business processes
2. Make a list of top 3-4 MDM solutions to further explore
3. Explore with a free trial

# Steps to implement MDM



- Start with a small pilot project to show the power of MDM (look for the stakeholders that were vocal about their data problems in Step 2)
- Leads to your first happy customers
- Secures your budget for the long term

# Steps to implement MDM



Learn about MDM:

- Important Capabilities
- Implementation Styles
- Domain Categories

Analyze your current data environment

Define your business goals

Secure the right sponsorship

Get the right team and tool

Implement a quick win

Repeat Step 6 and monitor



- Maintain the long term vision and release “wins” at regular intervals
- Keep on tracking the KPIs
- Maintain leadership’s interest in the MDM Program

# *Module 10 - Metadata Management*

# What is Metadata

**“Metadata is data that provides information about other data”**

# Example of Metadata



dog picture Properties

General Security Details Previous Versions

Property	Value
Copyright	
Image	
Image ID	
Dimensions	3024 x 4032
Width	3024 pixels
Height	4032 pixels
Horizontal resolution	72 dpi
Vertical resolution	72 dpi
Bit depth	24
Compression	
Resolution unit	
Colour representation	
Compressed bits/pixel	
Camera	
Camera maker	
Camera model	
F-stop	
Exposure time	
ISO speed	

[Remove Properties and Personal Information](#)

OK Cancel Apply

# What is Metadata Management?

It is the portfolio of best-practice processes and technologies that allow businesses to manage this data about their data and derive insights for more effective data management. It allows users of all kinds – business, technical, and operational – to search for, understand, and securely access the data they need to do their jobs.

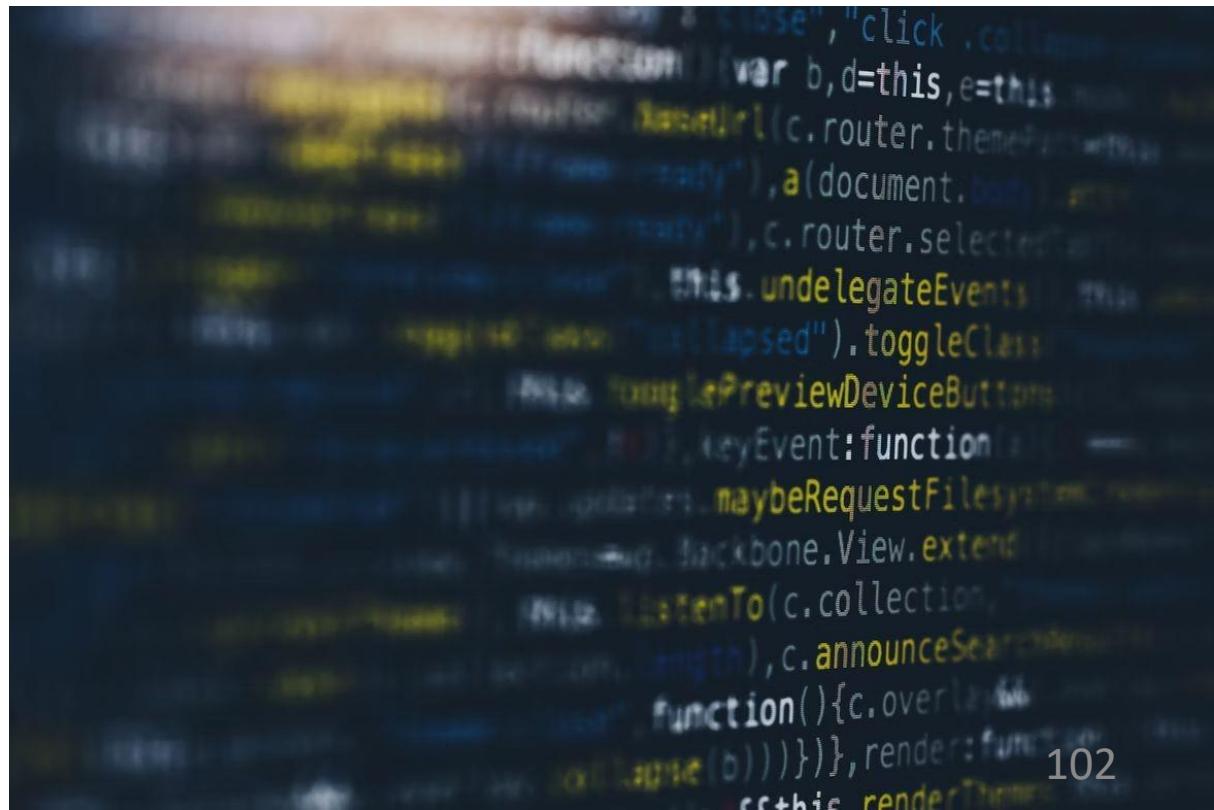
# Why do you need Metadata Management?

- Improved Consistency -establish a common business language
- Capture institutional knowledge
- Better data quality
- Faster access to insights
- Faster project delivery timelines
- Reduced costs
- Improved regulatory compliance



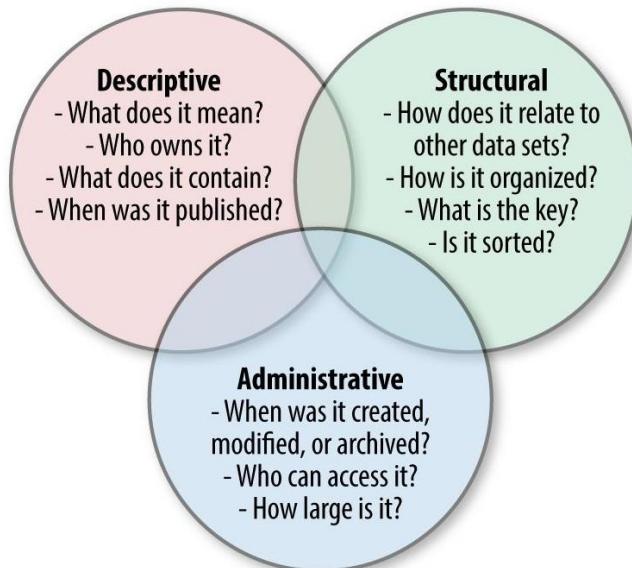
# Types of Metadata

- Descriptive metadata
- Structural metadata
- Administrative metadata



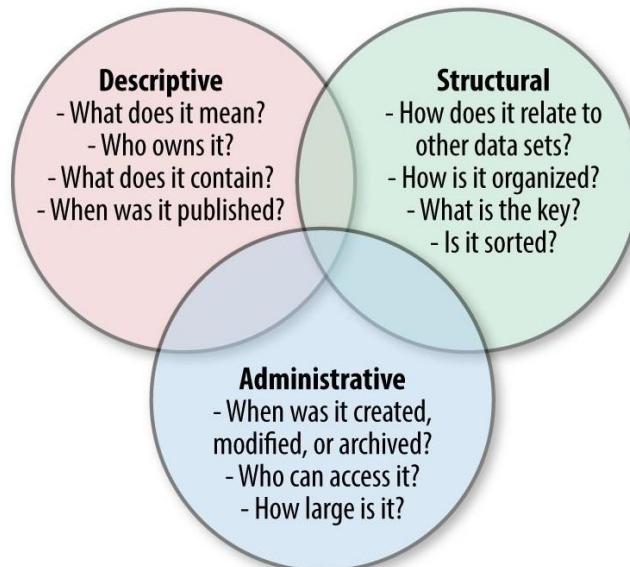
# Descriptive Metadata

Descriptive metadata describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.



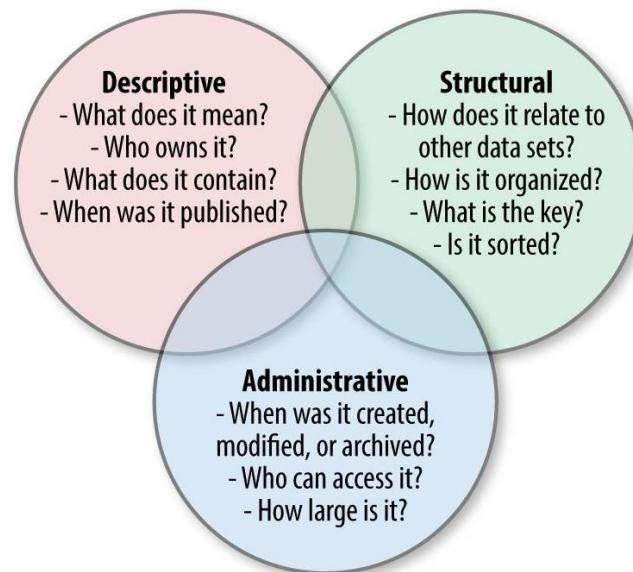
# Structural Metadata

Structural metadata is used to specify the relationships between components of a digital object (internal structure) and between different digital objects (external structure)



# Administrative Metadata

Administrative metadata provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.



# Magic Quadrant for Metadata Management Solutions

Figure 1. Magic Quadrant for Metadata Management Solutions



# Implement Metadata Management

Step 1

Step 2

Step 3

Step 4

Step 5



Select the Metadata  
Program team

Define the  
Metadata Strategy

Adopt Metadata  
Standards

Get the right  
Metadata  
Management tool

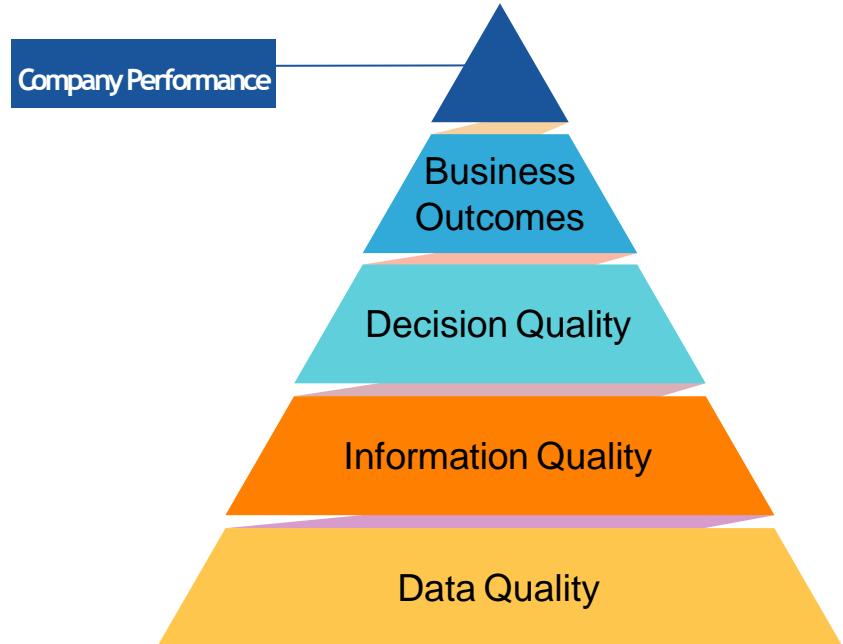
Deploy and expand

# Module 11 - Data Quality Management

# What is Data Quality?

Simple Definition of Data Quality:

*"Data quality is defined by how well a given dataset meets a user's needs. Data quality is an important criteria for ensuring that data-driven decisions are made as accurately as possible"*



# Data Quality Management

Definition of Data Quality Management:

*"Set of practices that aim at improving and maintaining a high quality of information within the organization"*

## Pillars of Data Quality Management

People

Data Profiling

Defining Data  
Quality

Data reporting

Data Repair

# Cost of Poor Data Quality

- According to IBM's estimate, the US lost \$3.1 trillion yearly due to bad data.
- Gartner.com suggests that organizations lose between \$10 to \$14 Million USD annually due to poor data.
- Integrate reported that around 40% of all leads have inaccurate data.
- Cio.com identified that around 80% of companies believe they lost revenue due to data challenges.
- MIT Sloan reported that employees spend half of their time coping with managing data quality tasks.
- Pragmaticworks states 20 to 30 percent of operating expenses are due to bad data.
- Econsultancy.com reported that due to poor data, companies having mail delivery issues lost about 30% of their revenue, in addition to the 21% of businesses experienced reputation damages.
- Gartner also reported that data scientists spend around 80% of their time cleaning and organizing data.

# Data Quality Dimensions



# Data Quality Improvement Process

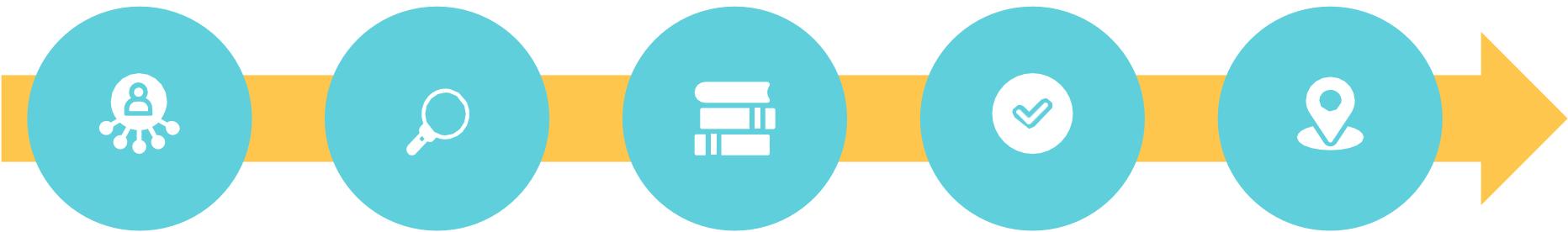
Step 1

Step 2

Step 3

Step 4

Step 5



Define the Data  
Quality improvement  
goals

Data Profiling

Conduct Data  
Quality  
Assessment

Resolve Data  
Quality Issues

Monitor and  
Control

# Data Quality Example.xlsx

# Data Quality Improvement Process

Step 1

Step 2

Step 3

Step 4

Step 5



Define the Data  
Quality improvement  
goals

Data Profiling

Conduct Data  
Quality  
Assessment

Resolve Data  
Quality Issues

Monitor and  
Control

# Data Root Cause Analysis - 5 Whys

1. Why do we have incorrect full name entries?

Due to manual data entry errors

2. Why do we have manual data entry errors?

The fields in the systems have no mandatory format

3. Why do the fields in the system do not have a mandatory format?

It was requested but not prioritized for delivery by the IT team

4. Why was it not prioritized for delivery by IT?

Limited resources for data team requests

5. Why are there limited resources for data team requests

Management strategy to dedicate 90% of IT resources to new software development

# Data Quality Improvement Process

Step 1

Step 2

Step 3

Step 4

Step 5



Define the Data  
Quality improvement  
goals

Data Profiling

Conduct Data  
Quality  
Assessment

Resolve Data  
Quality Issues

Monitor and  
Control

# Module 12 - Data Warehousing & Business Intelligence

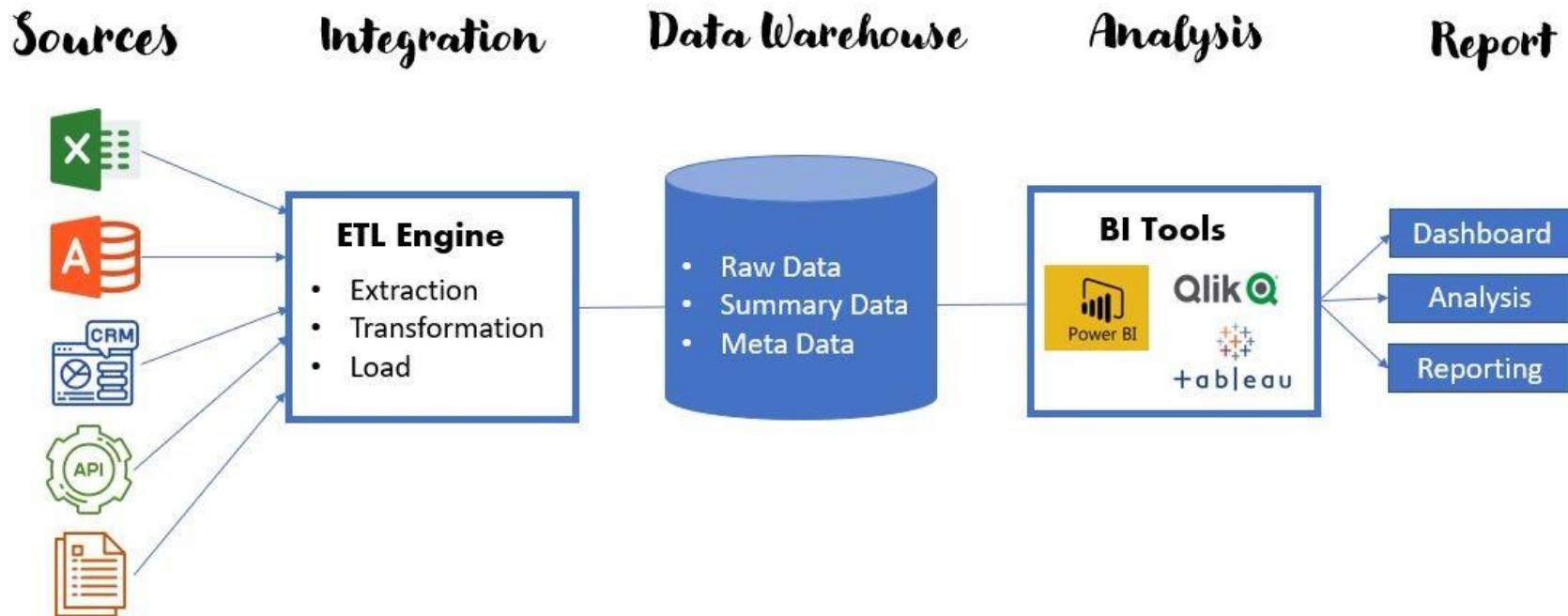
# What is Data Warehousing and Business Intelligence

“Data warehousing is the secure electronic storage of information by a business or other organization. The goal of data warehousing is to create a collection of historical data that can be retrieved and analyzed to provide useful insight into the organization's operations.

Data warehousing is a vital component of business intelligence. That wider term encompasses the information infrastructure that modern businesses use to track their past successes and failures and inform their decisions for the future.

”

# How the process works

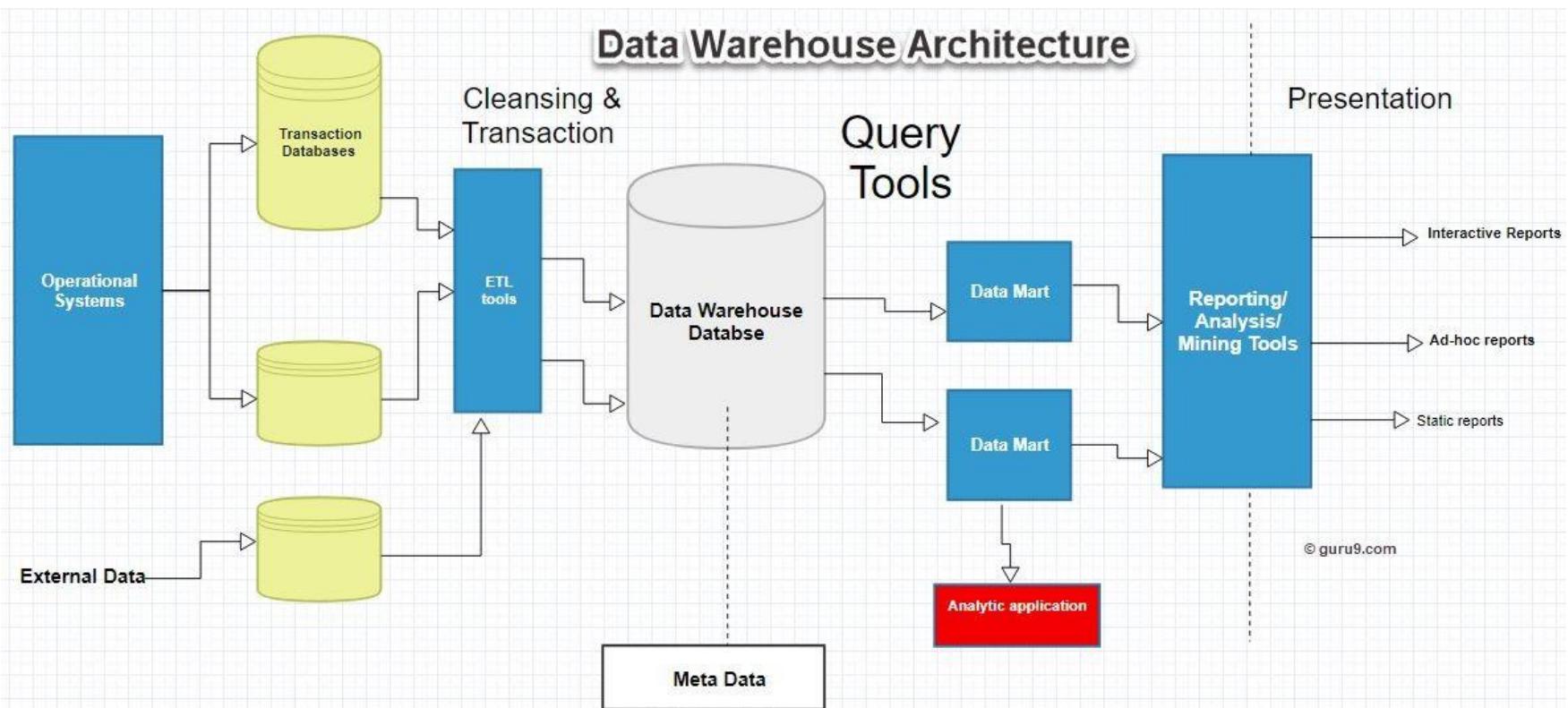


# What is a Data Warehouse (DW/DWH)

Data Warehouse (DW or DWH), and sometimes also referred to as Enterprise Data Warehouse (EDW) is a central repository of integrated data from one or more different sources.

Data Warehouses are used to store current and historical data in one single place.

# Data Warehouse Components



# What is a Data Warehouse (DW/DWH)

Data Warehouse (DW or DWH), and sometimes also referred to as Enterprise Data Warehouse (EDW) is a central repository of integrated data from one or more different sources.

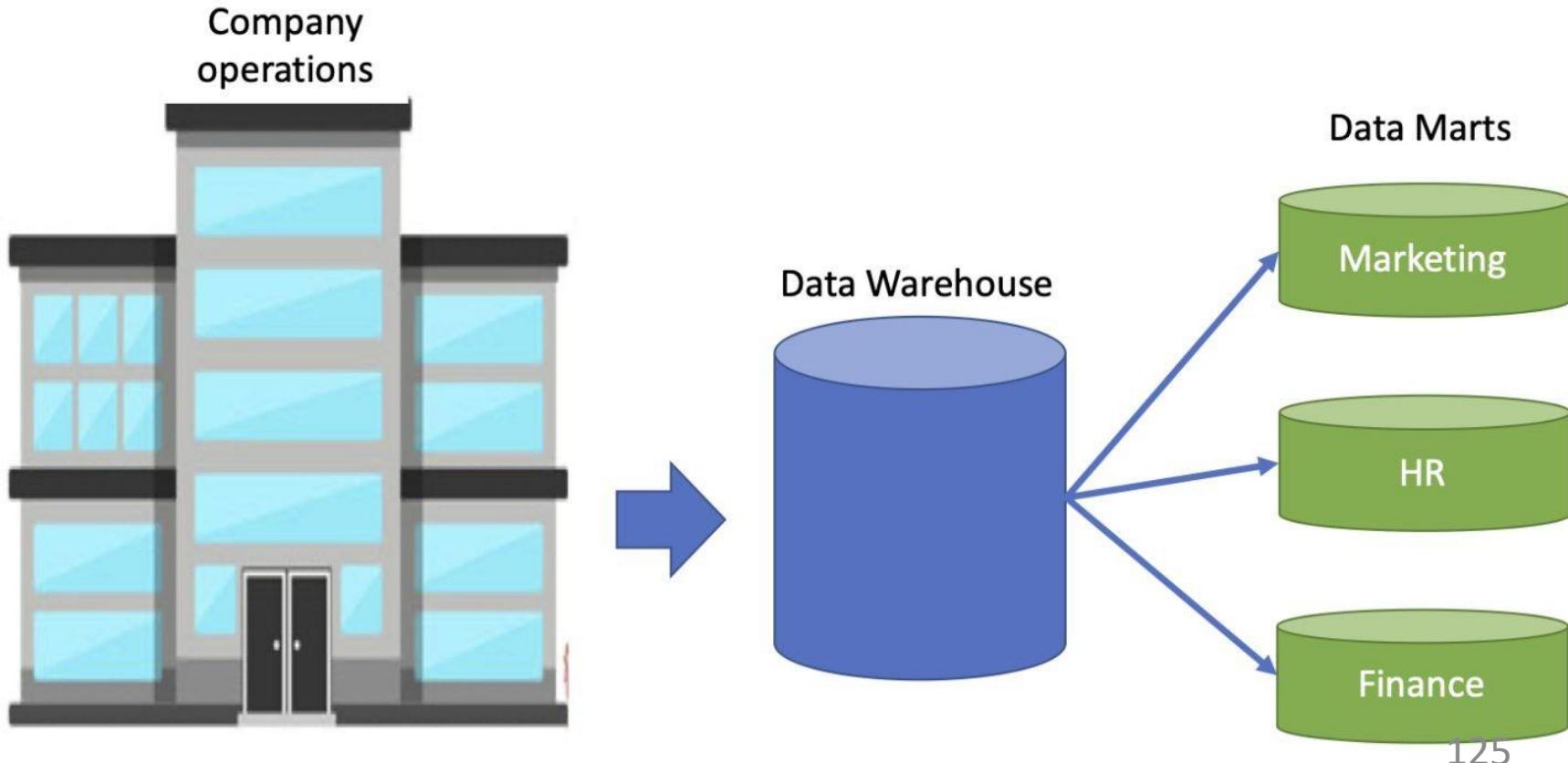
Data Warehouses are used to store current and historical data in one single place.

# Data Warehouse vs Database

Database vs. Data Warehouse	
While these two data storage elements may seem similar, they offer very different capabilities. Here is a brief breakdown of the differences:	
Database	Data Warehouse
Designed to record data	Designed to analyze data
Stores detailed data	Stores summarized data
Uses Online Transactional Processing OLTP	Uses Online Analytical Processing OLAP
Performs fundamental business operations and transactions	Allows users to analyze business data
Data is available in real time	Data must be refreshed when needed
Application-oriented data collection	Subject-oriented data collection
Limited to a single application	Draws data from a range of other applications

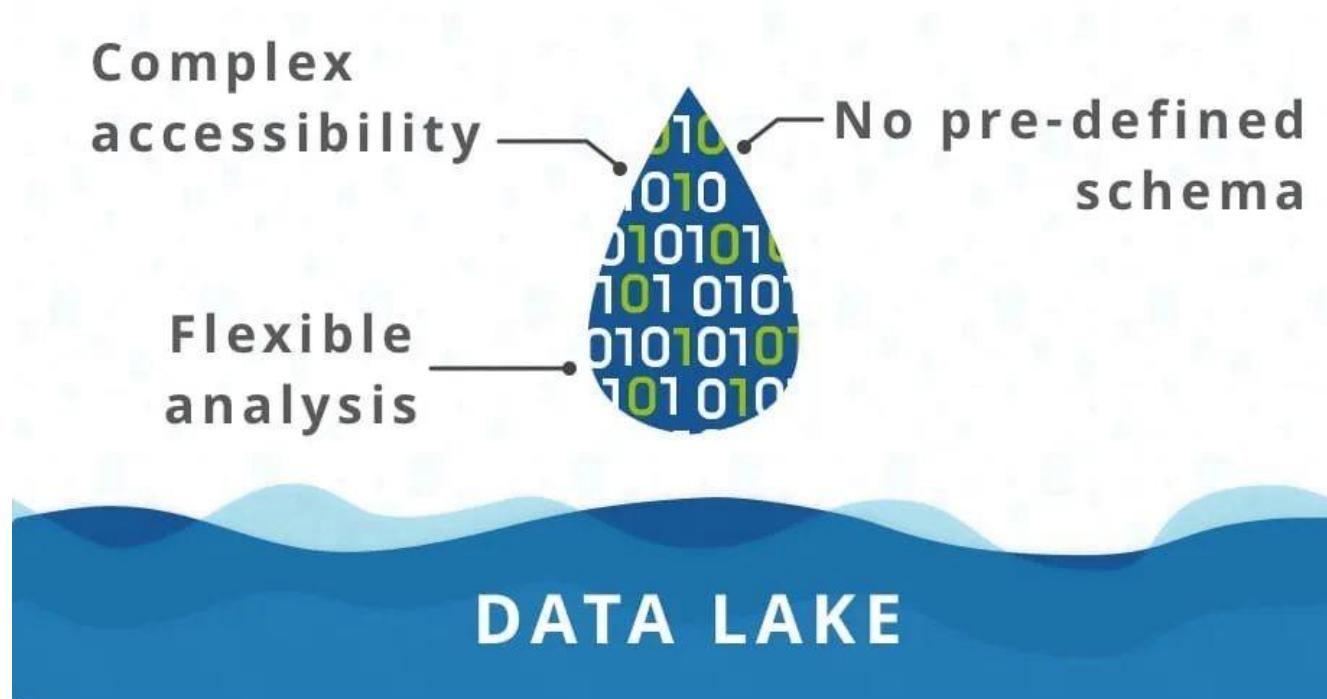
source:  SelectHub

# Data Warehouse vs Data Marts



# Data Warehouse vs Data Lake

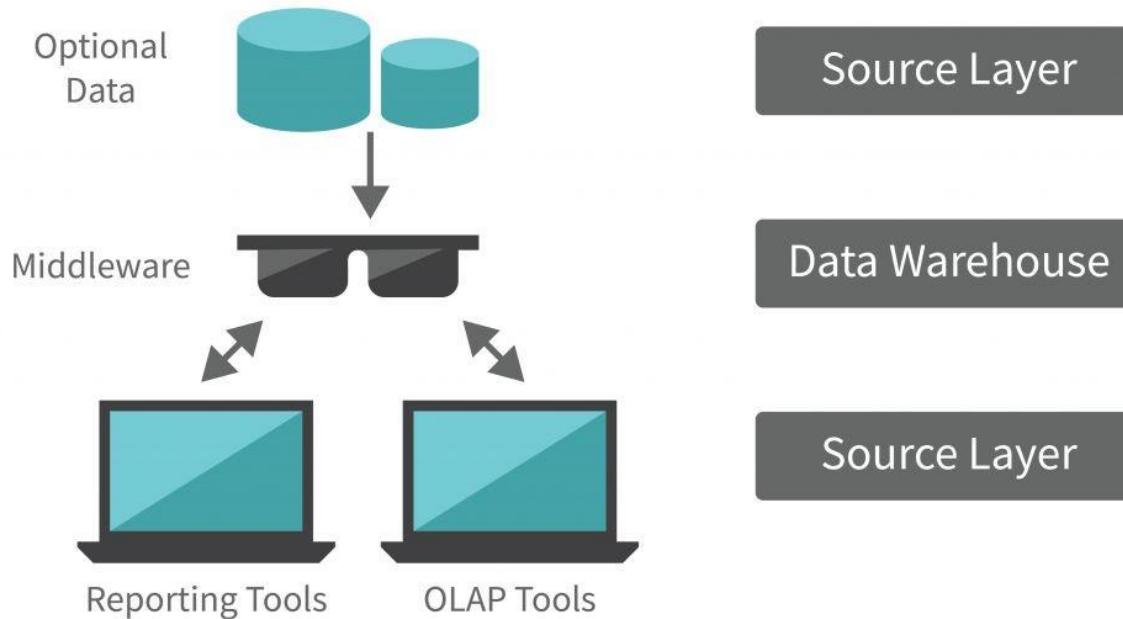
A data lake is a central storage repository that holds big data from many sources in a raw, granular format.



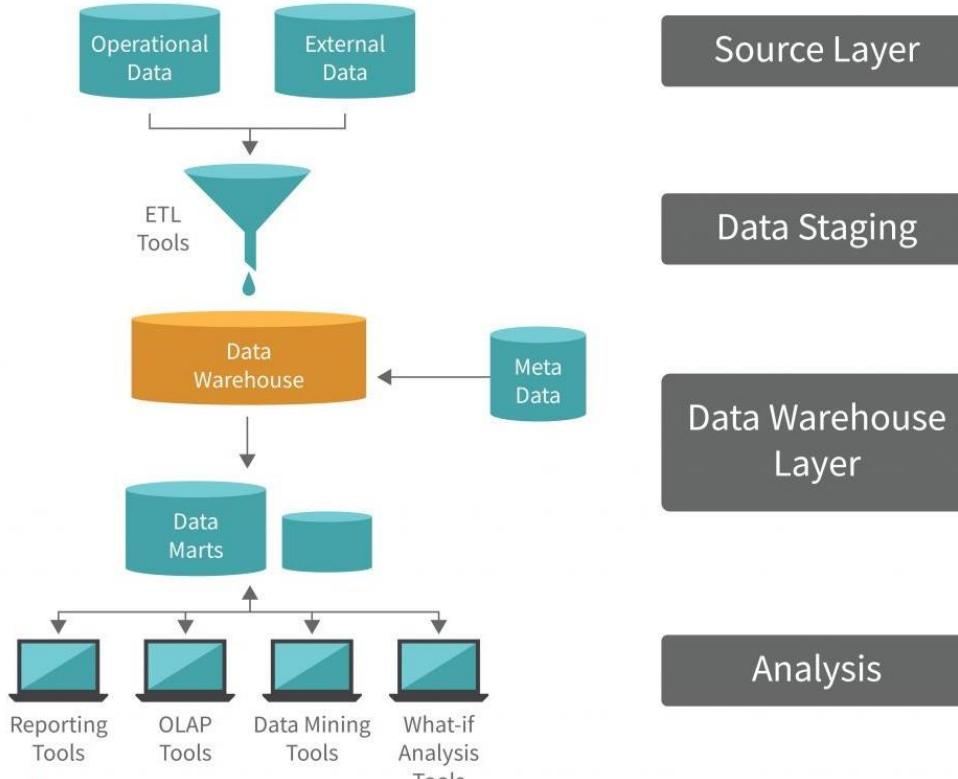
# Data Warehouse Architecture Types

- Single-Tier Architectures
- Two-Tier Architectures
- Three-Tier Architectures

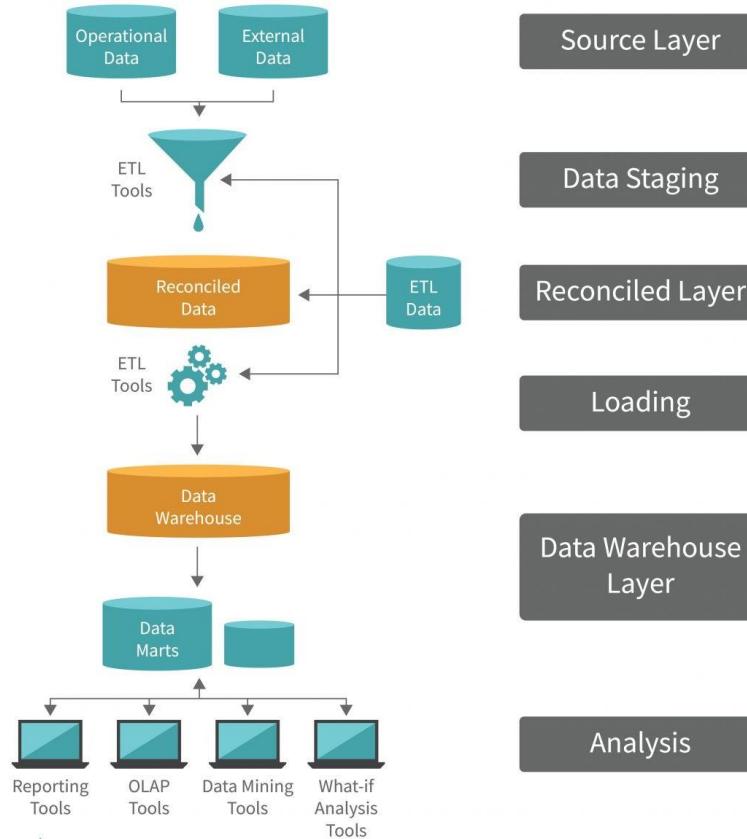
# Single-Tier Data Warehouse Architecture



# Two-Tier Data Warehouse Architecture



## Three-Tier Architecture for a Data Warehouse System



# What is Business Intelligence

“Business intelligence (BI) leverages software and services to transform data into actionable insights that inform an organization’s strategic and tactical business decisions.”

# Business Intelligence vs Business Analytics

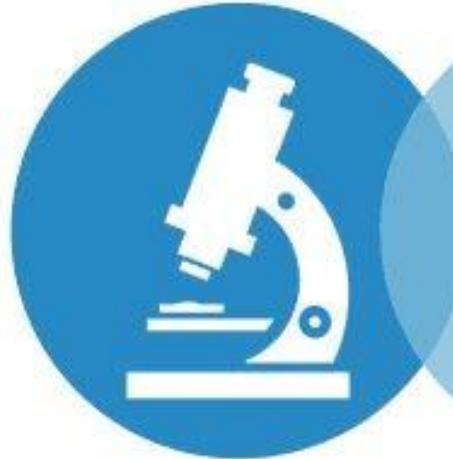
'The primary distinction between business intelligence and business analytics is the focus on when events occur. **Business intelligence is focused on current and past events that are captured in the data. Business analytics is focused on what's most likely to happen in the future..”**

# Applications of Business Intelligence

1. Sales Intelligence
2. Visualization
3. Reporting
4. Performance Management



# Categories of BI analysis



## Descriptive

Explains what happened.



## Diagnostic

Explains why it happened.



## Predictive

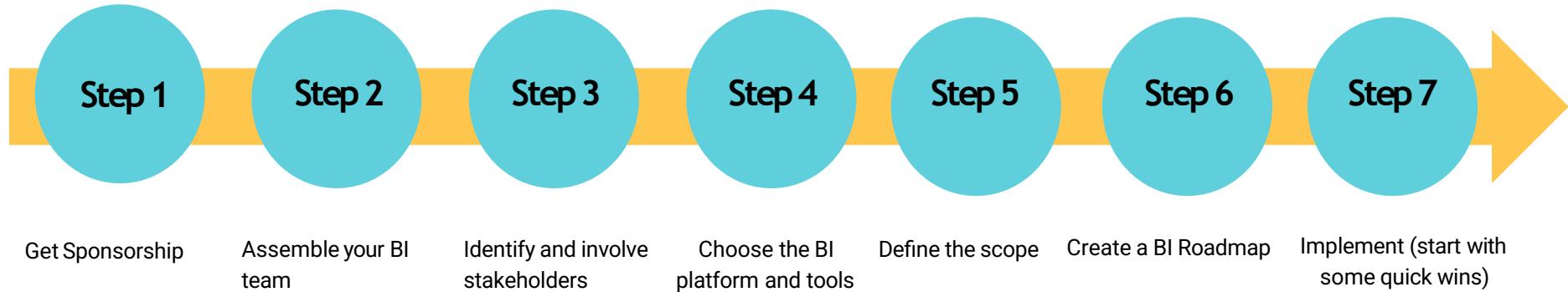
Forecasts what might happen.



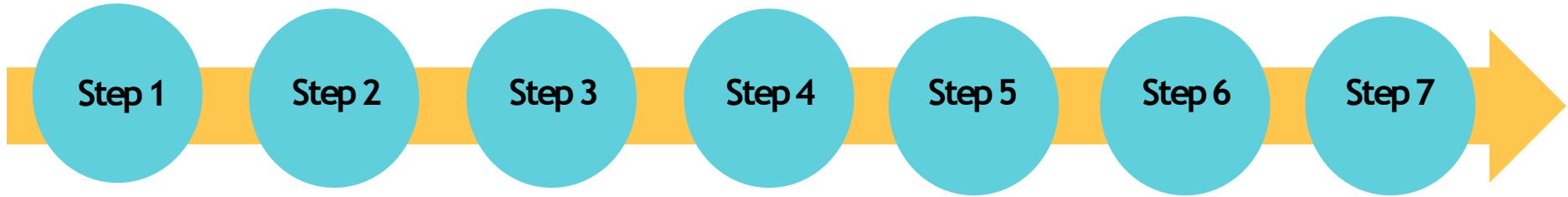
## Prescriptive

Recommends an action based on the forecast.

# Create a Business Intelligence Strategy



# Create a Business Intelligence Strategy



Get Sponsorship

Assemble your BI  
team

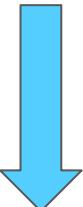
Identify and involve  
stakeholders

Choose the BI  
platform and tools

Define the scope

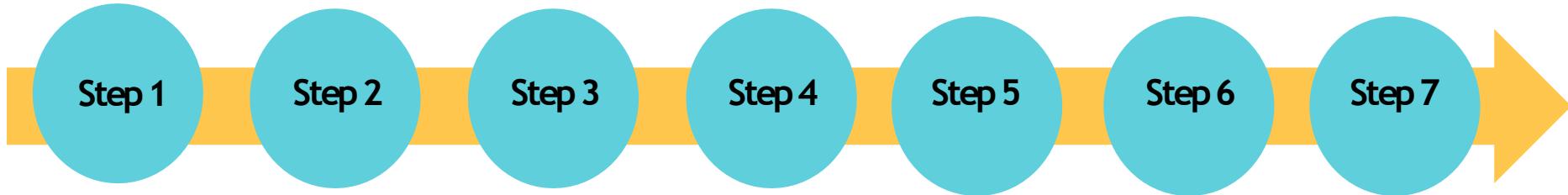
Create a BI Roadmap

Implement (start with  
some quick wins)



- Executive level sponsorship is key
- Update sponsor regularly
- Quick wins will help

# Create a Business Intelligence Strategy



Get Sponsorship

Assemble your BI  
team

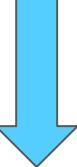
Identify and involve  
stakeholders

Choose the BI  
platform and tools

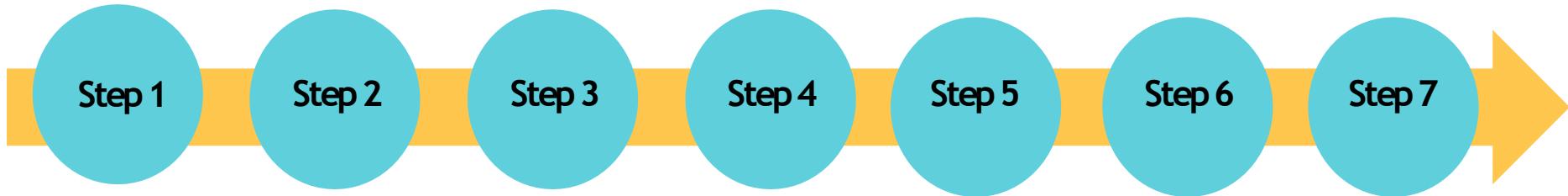
Define the scope

Create a BI Roadmap

Implement (start with  
some quick wins)

- 
- Program Manager
  - IT Owner
  - Enterprise Architect
  - Data Stewards
  - etc

# Create a Business Intelligence Strategy



Get Sponsorship

Assemble your BI team

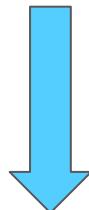
Identify and involve stakeholders

Choose the BI platform and tools

Define the scope

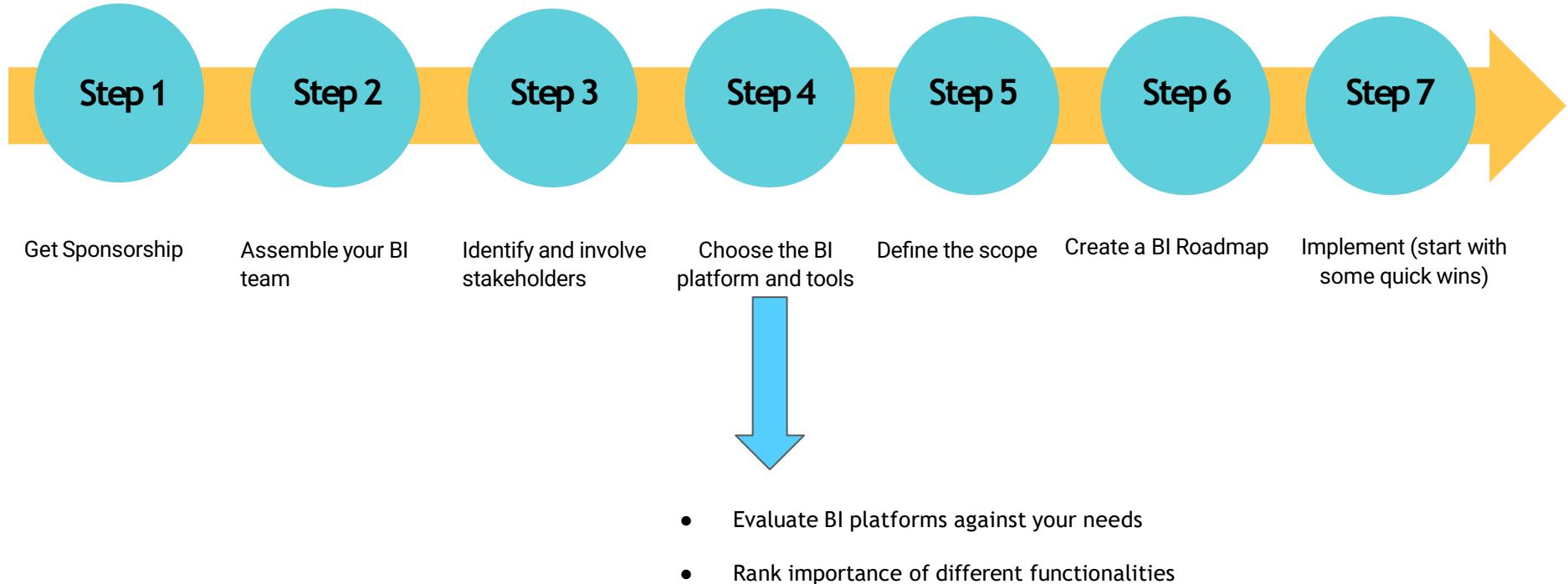
Create a BI Roadmap

Implement (start with some quick wins)

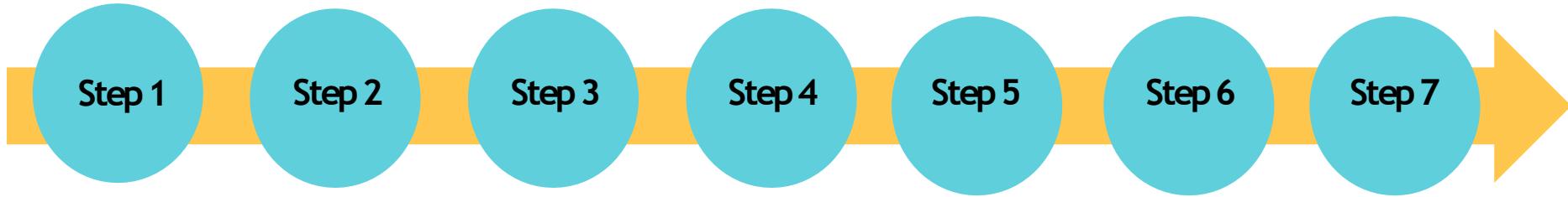


- Representative from every affected business group
- Involve early

# Create a Business Intelligence Strategy



# Create a Business Intelligence Strategy



Get Sponsorship

Assemble your BI team

Identify and involve stakeholders

Choose the BI platform and tools

Define the scope

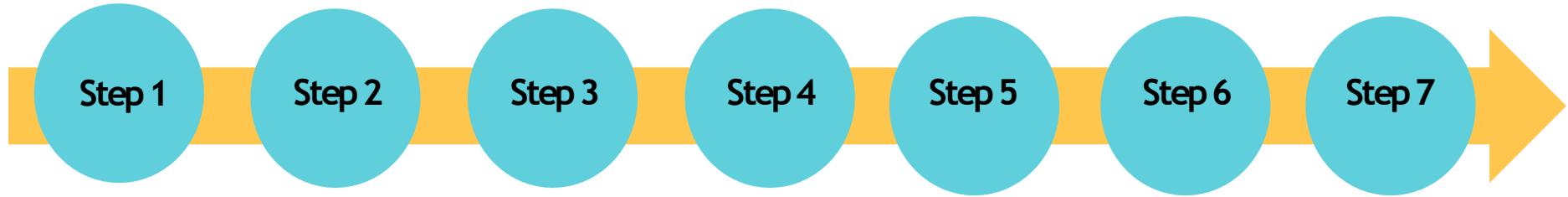
Create a BI Roadmap

Implement (start with some quick wins)



- What do you want to change?
- What do you envision the processes to look like?
- Do you start from a limited scope first?
- What will be the scope of analysis done?
- What is priority for the organization?

# Create a Business Intelligence Strategy



Get Sponsorship

Assemble your BI team

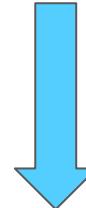
Identify and involve stakeholders

Choose the BI platform and tools

Define the scope

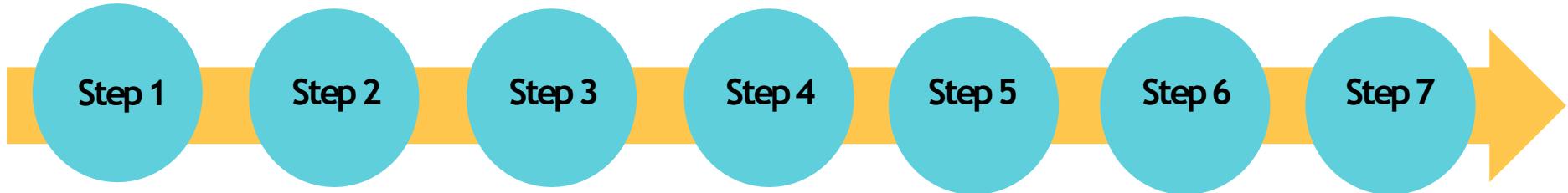
Create a BI Roadmap

Implement (start with some quick wins)



- Plan the scope in logical steps of release
- Keep in mind the dependencies!
- Adapt if needed
- Do not overcommit if not sure

# Create a Business Intelligence Strategy



Get Sponsorship

Assemble your BI team

Identify and involve stakeholders

Choose the BI platform and tools

Define the scope

Create a BI Roadmap

Implement (start with some quick wins)



- Have strong partnership between stakeholders, IT and BI teams
- Do not forget about security and security profiles
- Break up implementation into multiple phases
- Do not forget about training and change management

# Big Data and BI

- What is Big Data?
- Combining BI and Big Data



# Self-Service Business Intelligence

- What is Self-Service BI
- Why is it important
- What are the challenges
- Best Practices



# Magic Quadrant for Analytics and BI Platforms

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (March 2022)

# Congratulations!