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

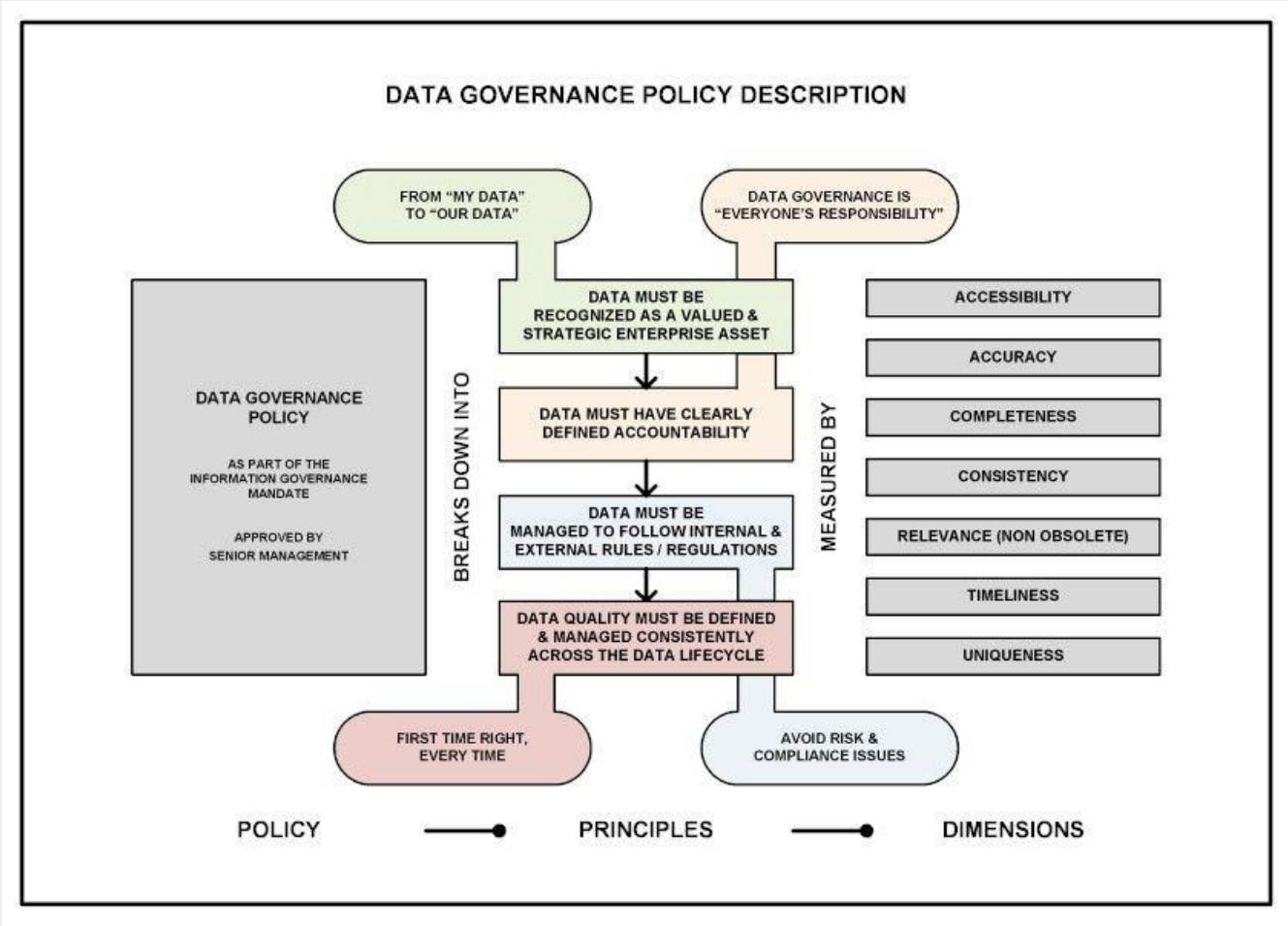


Image Credit: Robert S. Seiner, KIK Consulting

Governance Roles and Responsibilities

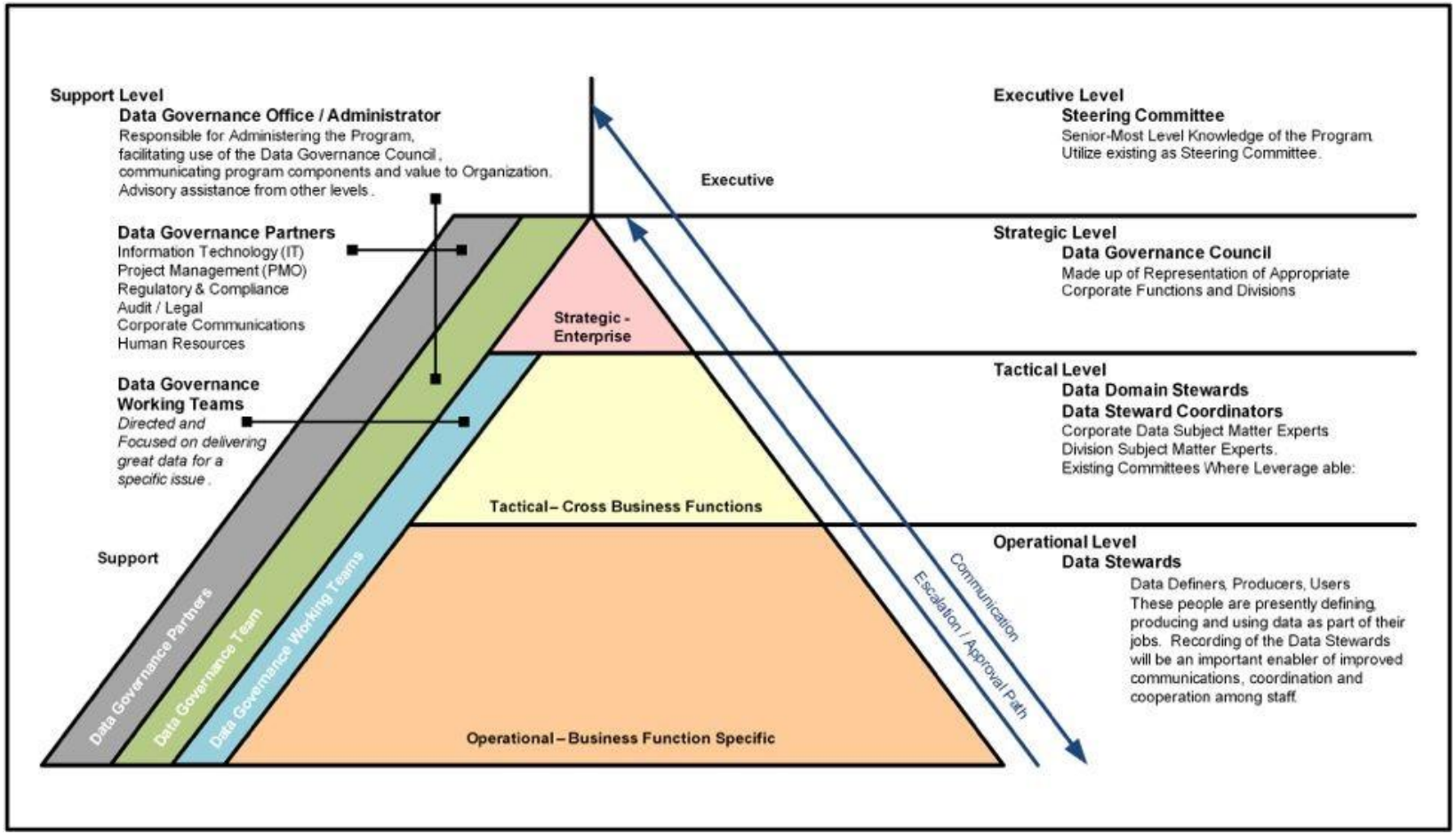


Image Credit: Robert S. Seiner, KIK Consulting

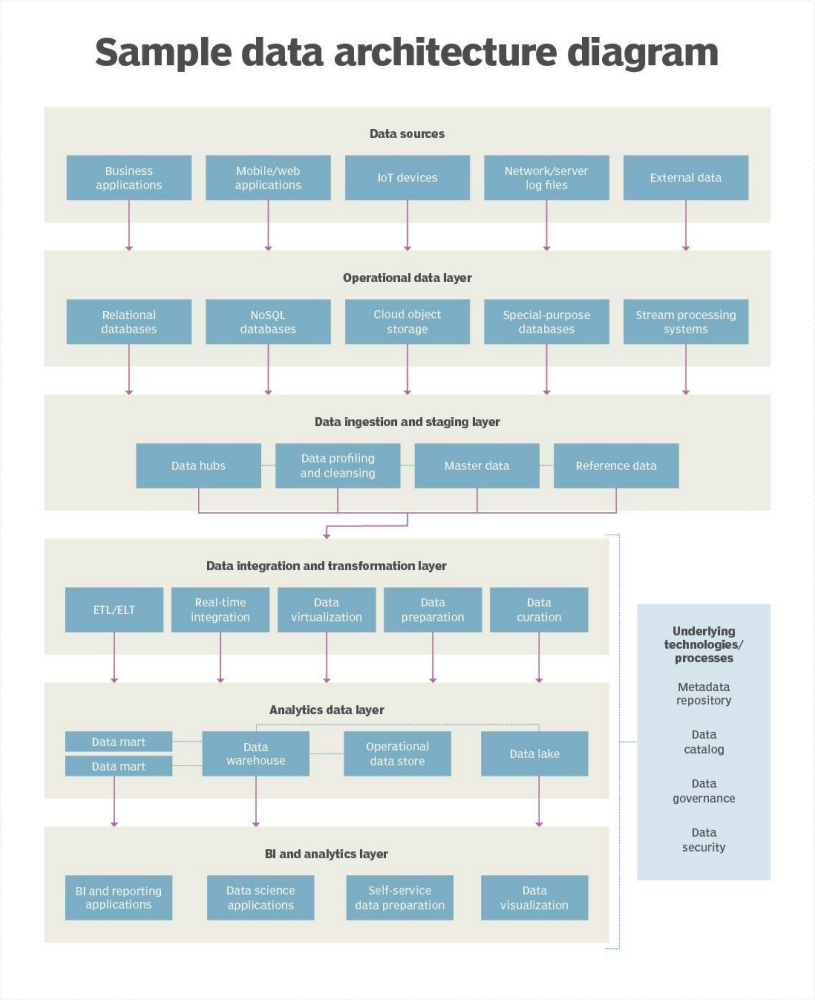
Module 3 - Data Architecture

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

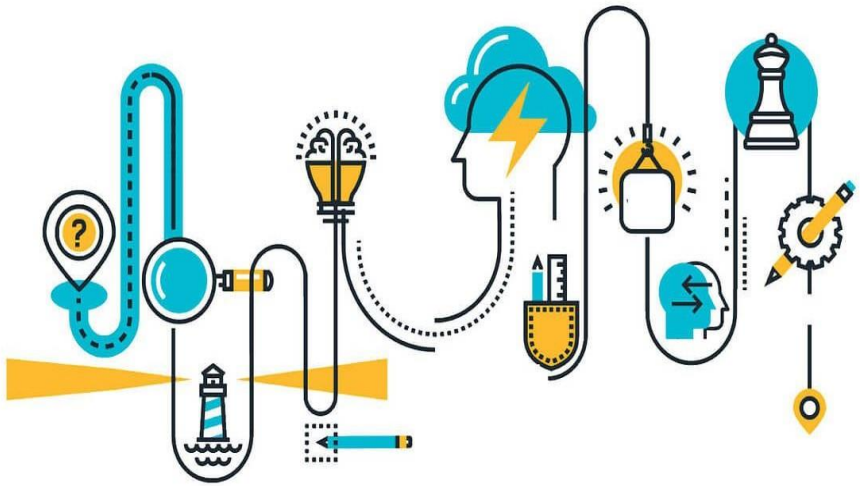
"**Data architecture** is a set of rules, policies, standards and models that govern and define the type of data collected and how it is used, stored, managed and integrated within an organization and its database systems. It provides a formal approach to creating and managing the flow of data and how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



Source of image:
https://www.techtarget.com/searchdatamanagement/definition/What-is-data-architecture-A-data-management-blueprint

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*



Sources: https://www.atscale.com/blog/the-six-modern-principles-of-modern-data-architecture/

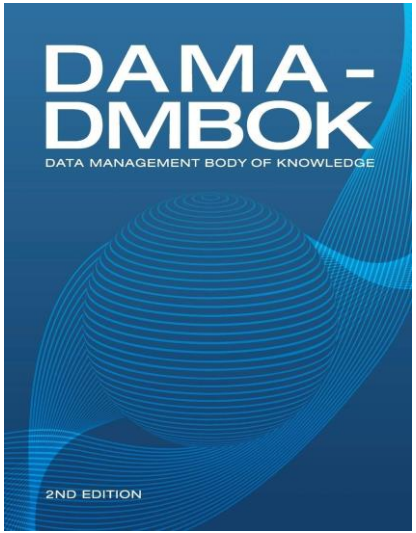
Data Architecture components

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

Sources: https://www.cio.com/article/190941/what-is-data-architecture-a-framework-for-managing-data.html

Data Architecture frameworks

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



ZACHMAN CERTIFIED™ ENTERPRISE ARCHITECT

TOGAF®

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*

Source: https://medium.com/@sutherl99/top-7-characteristics-of-a-modern-data-architecture-83b1e61f97f2

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.”

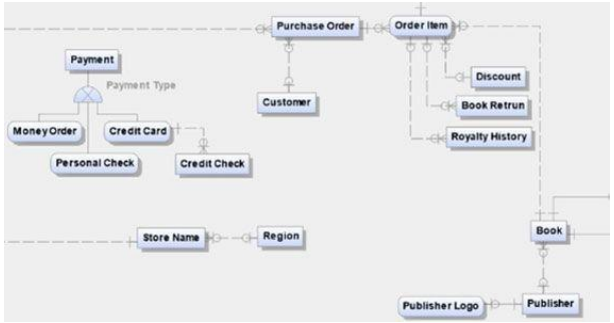
https://www.ibm.com/cloud/learn/data-modeling

Data Modeling vs Data Architecture

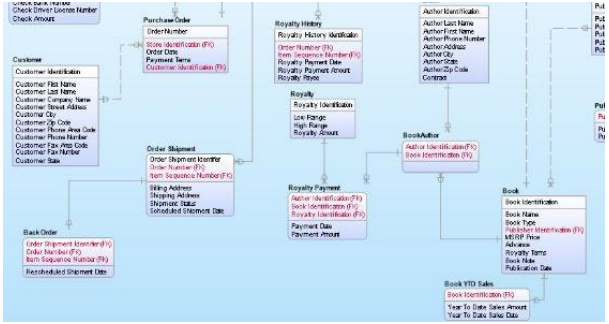
5 Key Differences	
Data Modeling	Data Architecture
Focuses on the representation of the data	Focuses on what tools and platforms to use for storing and analyzing data
Focus on accuracy of the data	Focus is on the infrastructure housing the data
Focus on reliability of the data	Focus on keeping the data 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

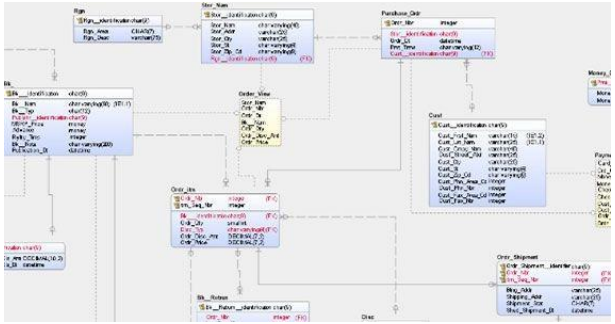
Conceptual data model



Logical data model



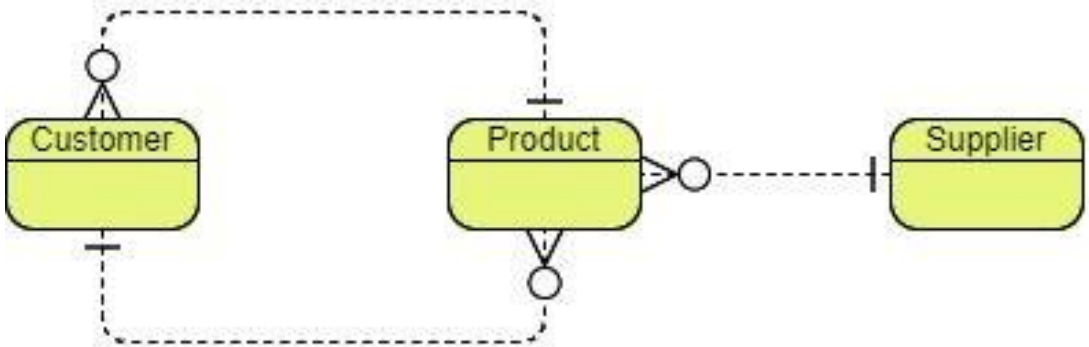
Physical data model



Conceptual data model

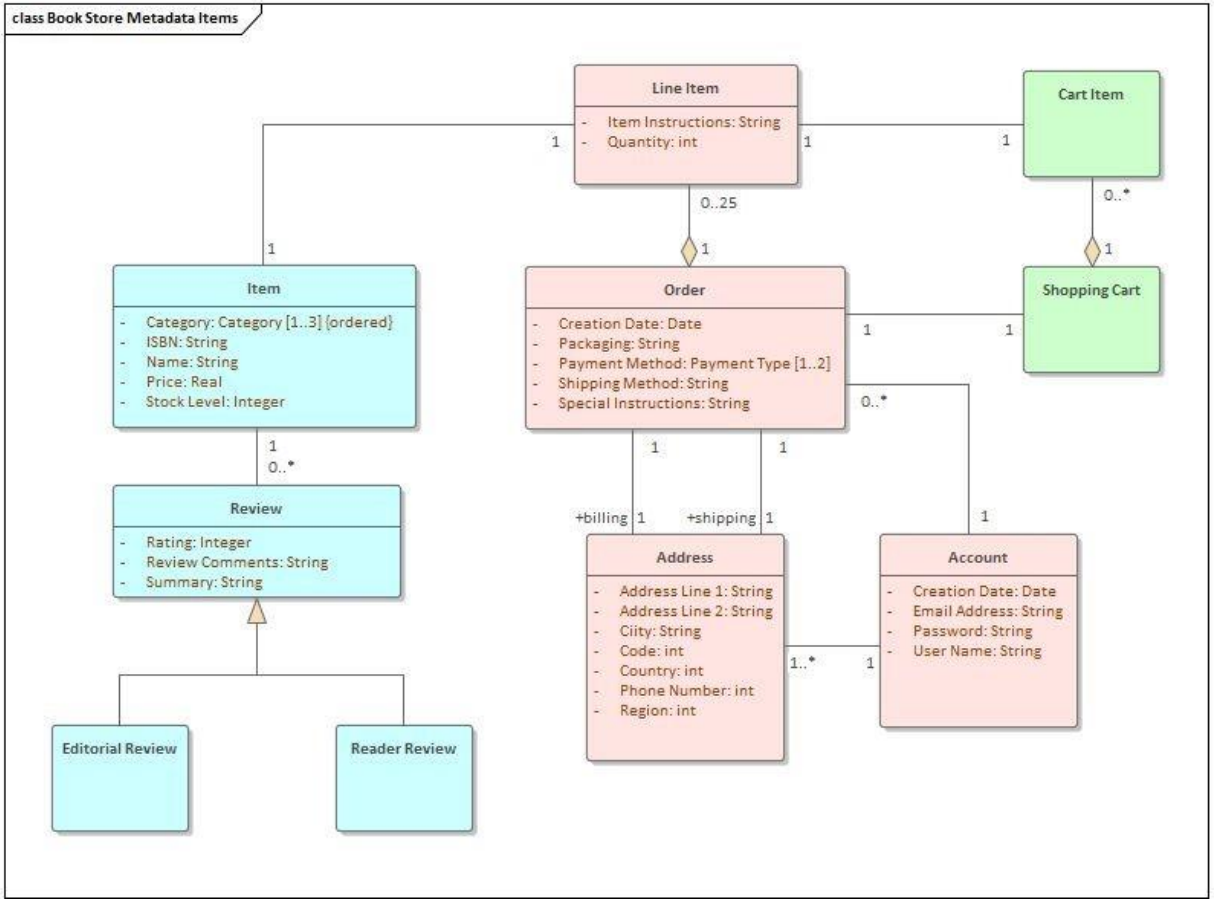
Elements of a Conceptual data model:

- Entity
- Attribute
- Relationship

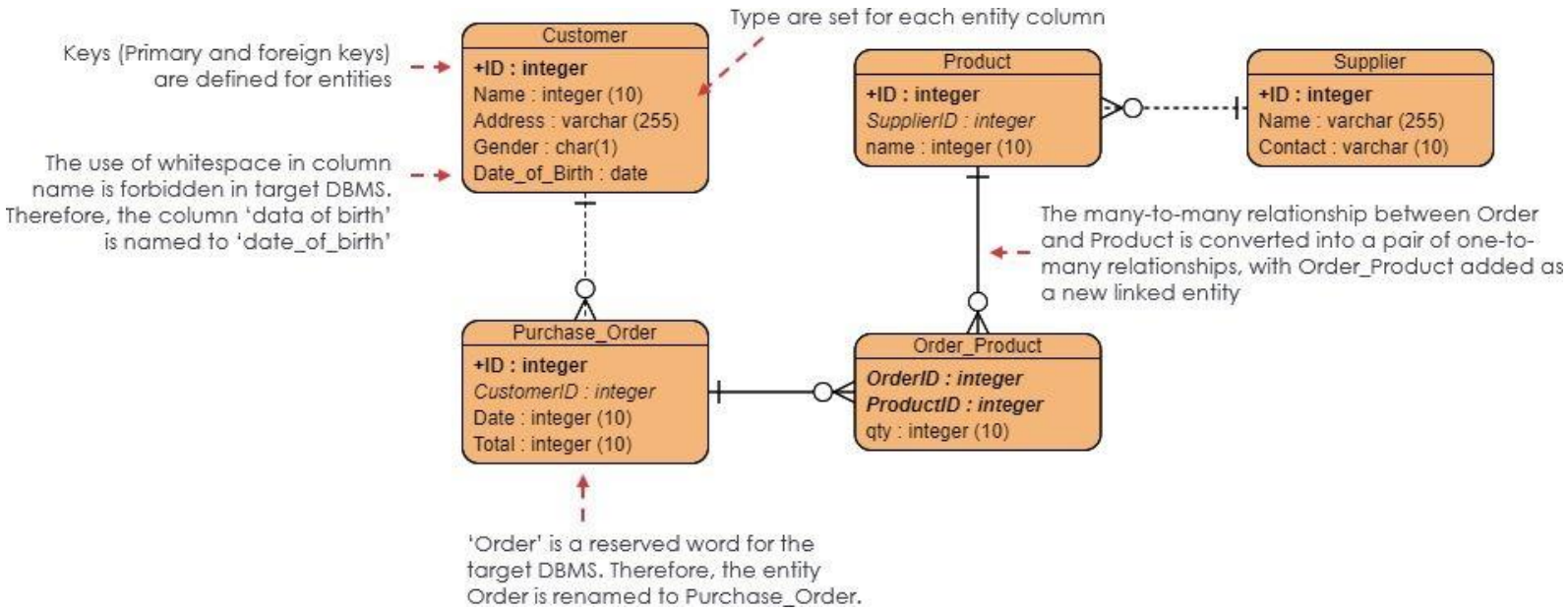


Source of Image: https://online.visual-paradigm.com/knowledge/visual-modeling/conceptual-vs-logical-vs-physical-data-model

Logical data model



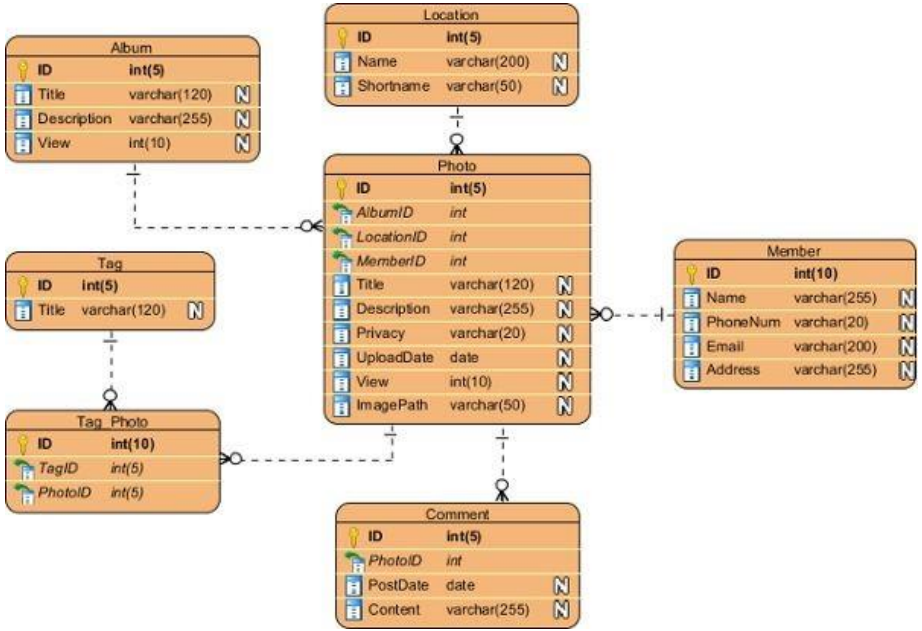
Physical data model



Source of Image: https://online.visual-paradigm.com/knowledge/visual-modeling/conceptual-vs-logical-vs-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



Source: https://www.ibm.com/cloud/learn/data-modeling

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

Source: https://spectralops.io