CDMP Study Group

CHAPTER 5- DATA MODELING & DESIGN

APRIL 13, 2022

Lynn Noel, DAMA New England – VP, Membership
Email: lynnoel@lynnoel.com

AGENDA

Introduction

Activities

Tools

Best Practices

Governance

- Facilitator Introduction
- Chapter Overview
- Data Modeling and Design
  - 1. Introduction
  - 2. Activities
  - 3. Tools
  - 4. Best Practices
  - 5. Data Model Governance
- Q & A
- Next Session

Facilitator

LYNN E. NOEL

- **Principal, Digital Heritage Consulting**
- Data Governance Analyst, Deloitte
- Executive Certificate in Digital Business Strategy, MIT
- Mini-Masters in User Experience Design, Rutgers
- M.S. Geography, University of Wisconsin-Madison
- Director/mentor for IM & analytics, architecture & modeling, collaboration & content mgmt at Big Five IT, midsize, & startup firms

CONTACT INFO:

EMAIL: lynnoel@lynnoel.com

PHONE: +1-978-985-2707

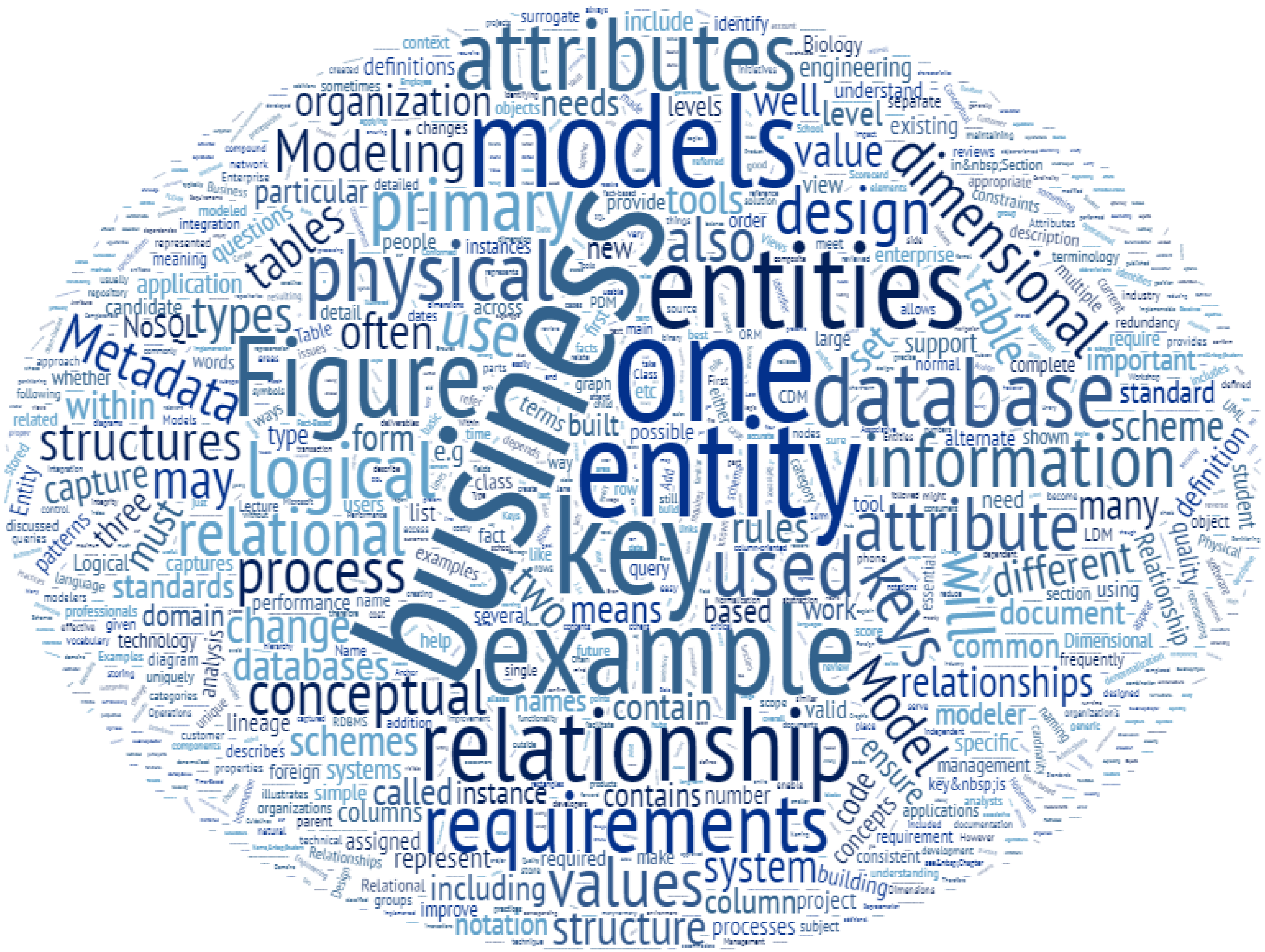
: /IN/lnoel

Possible Study Group Scenarios

1. You meant to read the chapter but section 1 was impenetrable
2. You started the chapter and got lost somewhere in Section 1.3
3. You made it through the chapter but you don't remember much
4. You are a data architect and know this cold, so you are wondering how we're going to get through this chapter in under an hour
5. You've been homeschooling your kids and want a break right now so you're hoping this won't be as boring as you fear it might be

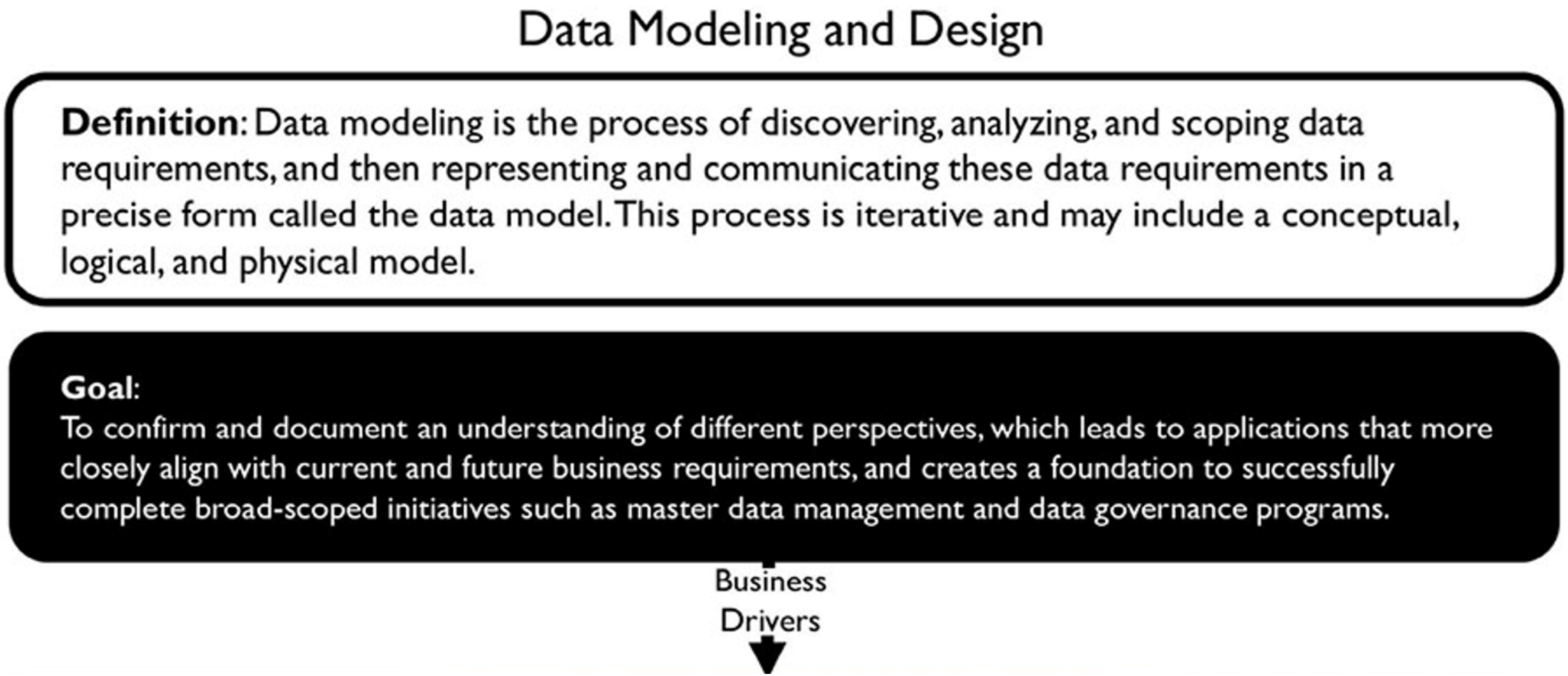
I DREW YOU A MAP – I HOPE IT HELPS

MODELS are Requirements for Business Entities, Attributes, and Relationships

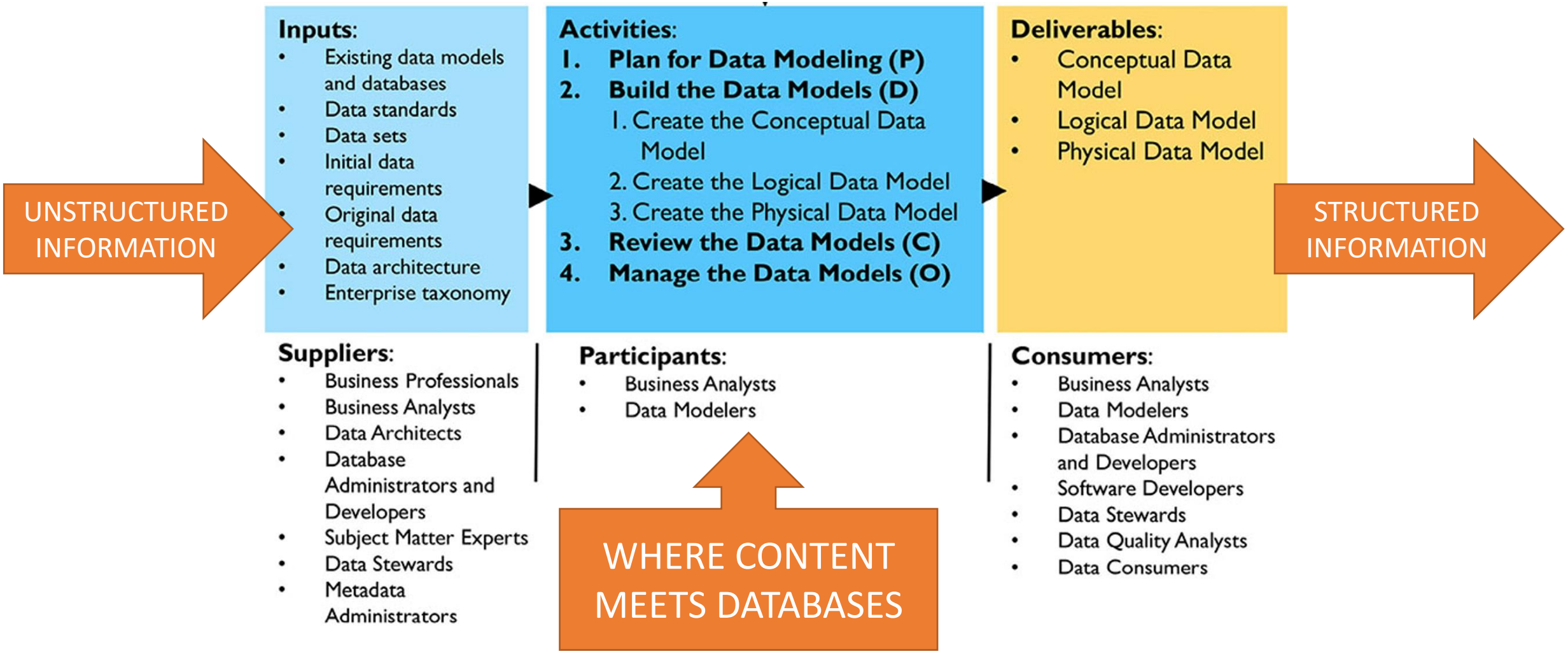


TIP: Prepare by first reading **Navigating the Labyrinth** Chapter 6, *Planning and Design in Data Lifecycle Management*

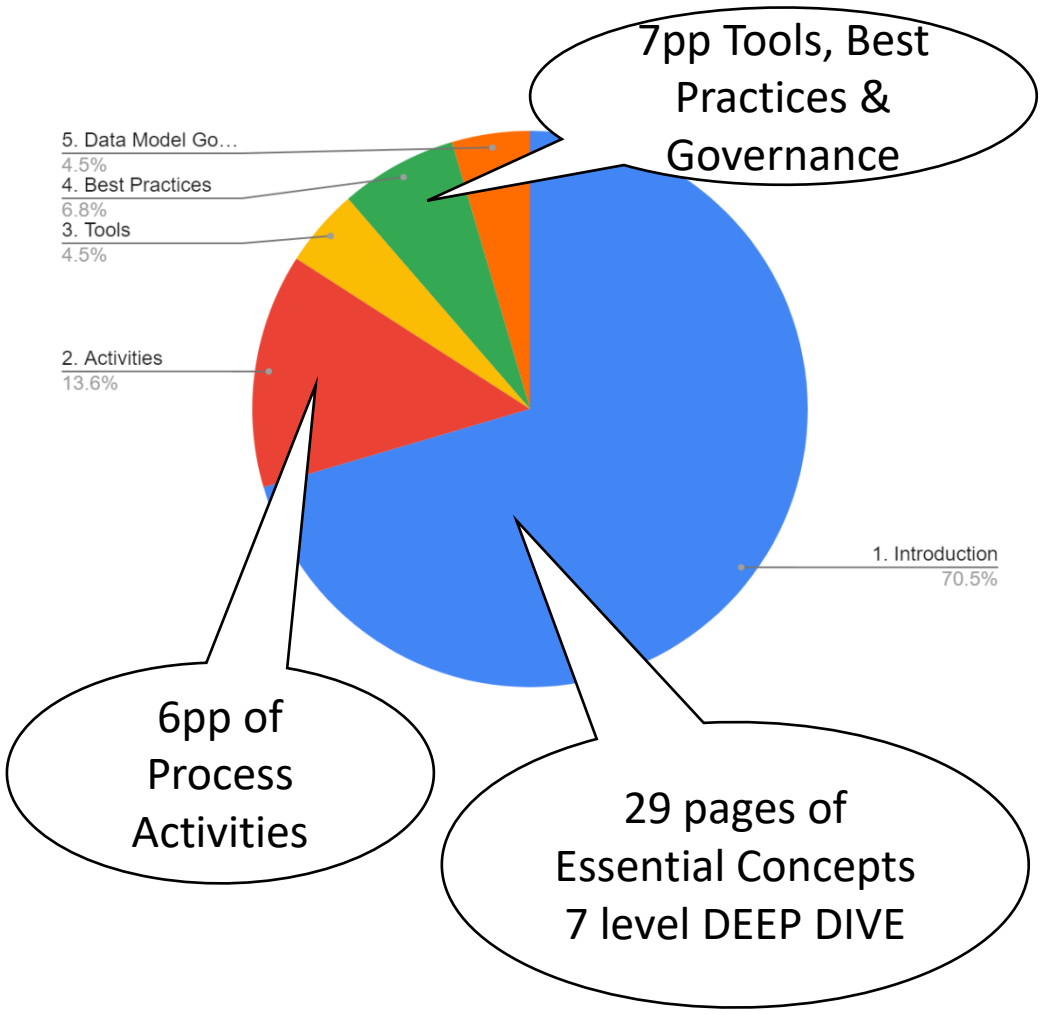
What Is Data Modeling and Design?



Who Does What in Data Modeling and Design?



What's In 44 Pages and Who Needs to Know What?



DATA MODELING ROLE	Supply	Participate	Consume
Business Professionals/SMEs	Y		
Business Analysts	Y	Y	Y
Data Architects	Y		
DBAs/Developers	Y		Y
Data Stewards	Y		Y
Metadata Admins	Y		
Data Modelers		Y	Y
Software Developers			Y
Data Quality Analysts			Y
Data Consumers			Y
10 ROLES	**6**	**2**	**7**

It takes more people to use models than to make them

Chapter Map

OVERVIEW: PART I

- **1. Introduction**
  - 1.1 Business Drivers
  - 1.2 Goals & Principles
  - 1.3 Essential Concepts

    29 PAGES AND 7 LEVELS WHERE EVERYONE GETS LOST
- **2. Activities**
  - 2.1 Plan for Data Modeling
  - 2.2 Build the Data Model
  - 2.3 Review the Data Models
  - 2.4 Maintain the Data Models
- **3. Tools**
  - 3.1 Data Modeling Tools
  - 3.2 Lineage Tools
  - 3.3 Data Profiling Tools
  - 3.4 Metadata Repositories
  - 3.5 Data Model Patterns
  - 3.6 Industry Data Models
- **4. Best Practices**
  - 4.1 Best Practices in Naming Conventions
  - 4.2 Best Practices in Database Design
- **5. Data Model Governance**
  - 5.1 Data Model and Design Quality Management
  - 5.2 Data Modeling Metrics

DEEP DIVE PART II

Data models are for **business communication**. Don't get lost in the weeds of learning how to make models before you learn to use and evaluate them. If you don't understand your data model, your modeler has more work to do.

Business Drivers, Goals, and Principles

- Introduction
- Activities
- Tools
- Best Practices
- Governance

1.1 Business Drivers

- **Common vocabulary**
- **Documentation**
- **Communication**
- **Starting point for app customization, integration, or replacement**

1.2 Goals and Principles

- **Formalization**
- **Scope definition**
- **Knowledge retention and documentation**

"The data model becomes a **reusable map** to help business professionals, project managers, analysts, modelers, and developers understand **data structure within the environment**. In much the same way as the mapmaker learned and documented a geographic landscape for others to use for **navigation**, the modeler enables others to understand an **information landscape** (Hoberman, 2009)."

Essential Concepts Overview

- Introduction
- Activities
- Tools
- Best Practices
- Governance

WE WILL COME BACK TO THIS IN THE DEEP DIVE

- 1.3.1 Data Modeling and Data Models
  - **Diagram**
  - **Pattern**
  - **Map**
  - **Blueprint**
  - **Communication**
- 1.3.2 Types of Data Modeled
  - **Categories**
    - **Attributes**
  - **Resources**
    - **Reference**
  - **Events**
    - **Transactional**
  - **Transaction Details**
    - **Big Data**
- 1.3.3 Data Model Components
  - 1.3.3.1 Entity
  - 1.3.3.2 Relationship
  - 1.3.3.3 Attributes
  - 1.3.3.4 Domain
- 1.3.4 Data Modeling Schemes
  - 1.3.4.1 Relational
  - 1.3.4.2 Dimensional
  - 1.3.4.3 Object-Oriented (UML)
  - 1.3.4.4 Fact-Based Modeling (FBM)
  - 1.3.4.5 Time-Based
  - 1.3.4.6 NoSQL
- 1.3.5 Data Model Levels of Detail
  - 1.3.5.1 Conceptual
  - 1.3.5.2 Logical
  - 1.3.5.3 Physical
- 1.3.6 Normalization
  - **1NF**
  - **2NF**
  - **3NF**
  - **BCNF**
  - **4/5NF**
- 1.3.7 Abstraction
  - Generalization
    - Supertypes
  - Specialization
    - Subtypes

HOLD THAT THOUGHT!

2. Activities

Introduction

Activities

Tools

Best Practices

Governance

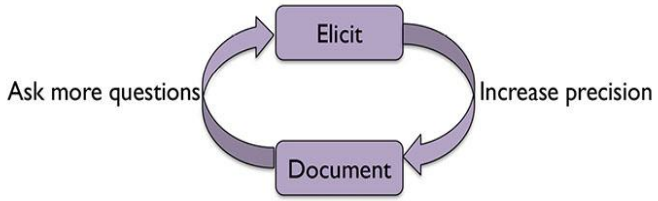


Figure 53 Modeling is Iterative

2.1 Plan for Data Modeling

- Diagrams
- Definitions
- Issues and questions
- Lineage

2.2 Build the Data Model

- 2.2.1 Forward Engineering
  - **CDM, LDM, PDM**
- 2.2.2 Reverse Engineering

2.3 Review the Data Models

- Quality Control
- Continuous Improvement

2.4 Maintain the Data Models

- Update when change occurs
- Sync physical to logical models

Suppliers → Participants → Consumers → Governance

MINI-DIVE

2.2.1.1 Conceptual Data Modeling (CDM)

Introduction

Activities

Tools

Best Practices

Governance

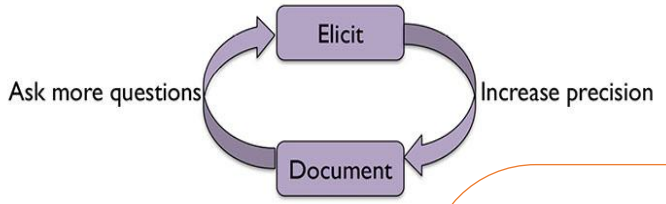


Figure 53 Modeling is Iterative

Enterprise Taxonomy

Select Scheme

- relational, dimensional, OO, fact-based, time-based, NoSQL

Select Notation

- Standards
- User familiarity

Complete Initial CDM

- concepts (nouns) **=Entities**
- activities (verbs) **=Relationships**

Incorporate Enterprise Terminology

Obtain Sign-off

Successful projects will iterate/change back up to this level during LDM and PDM – managing expectations is key

This might not look like a model yet: it often starts as a glossary

This never stops

2.2.1.2 Logical Data Modeling (LDM)

Introduction

Activities

Tools

Best Practices

Governance

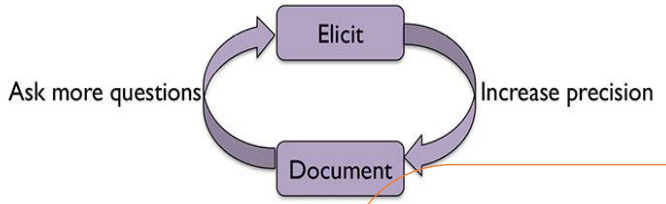


Figure 53 Modeling is Iterative

Entity-Relationship Diagram

2.2.1.2.1 Analyze Information Requirements

2.2.1.2.2 Analyze Existing Documentation

2.2.1.2.3 Add Associative **Entities**

2.2.1.2.4 Add **Attributes**

2.2.1.2.5 Assign **Domains**

2.2.1.2.6 Assign **Keys (Relation-ships)**

For effective transformation, stay focused on future state and key gaps

Domain in this context means the complete set of assignable values

2.2.1.3 Physical Data Modeling (PDM)

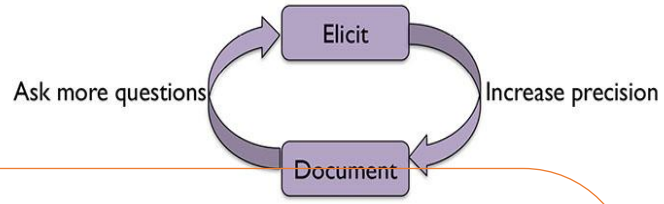
Introduction

Activities

Tools

Best Practices

Governance



Development has often started by now

Hybrid cloud platforms are disruptive to PDM

2.2.1.3.1 Resolve Logical Abstrac-tions

2.2.1.3.2 Add Attribute Details

2.2.1.3.3 Add Reference Data Objects

2.2.1.3.4 Assign Surrogate Keys

2.2.1.3.5 Denor-malize for Perfor-mance

2.2.1.3.6 Index for Perfor-mance

2.2.1.3.7 Partition for Perfor-mance

2.2.1.3.8 Create Views

DevOps teams need architectural guidance to execute transformation during continuous delivery

Productized data models are disruptive to PDM

This will stop scaling

2. Activities

DATA MODELING ROLE	Supply	Participate	Consume
Business Professionals/SMEs	Y		
Business Analysts	Y	Y	Y
Data Architects	Y		
DBAs/Developers	Y		Y
Data Stewards	Y		Y
Metadata Admins	Y		
Data Modelers		Y	Y
Software Developers			Y
Data Quality Analysts			Y
Data Consumers			Y
10 ROLES	**6**	**2**	**7**

This is not "continuous delivery of working software," but is critical to it

2.3 Review the Data Models

- Quality Control
- Continuous Improvement

GOVERNANCE

2.4 Maintain the Data Models

- Update when change occurs
- Sync physical to logical models

BEST PRACTICE

3. Tools

- Introduction
- Activities
- Tools
- Best Practices
- Governance

Data Catalogs

3.1 Data Modeling Tools
- Drawing
- Rubber-banding
- Forward-engineering with DDL
- Reverse engineering
- Metadata
- Sharing

3.2 Lineage Tools
- Attribute source structures
- Impact analysis
- Integration Tools

3.3 Data Profiling Tools
- Data exploration
- Metadata validation
- Data quality analysis
- Data model validation

3.4 Metadata Repositories
- Stores descriptive model
- Sharing
- Viewing and navigation

Reference Models

3.5 Data Model Patterns
- Reusable
- Elementary patterns
- Assembly patterns
- Integration patterns

3.6 Industry Data Models
- Prebuilt for an industry domain
- Broad and detailed
- Reference model for customization

Data catalogs are “toolboxes” that include varying bundles of these tools

4. Best Practices

Introduction

Activities

Tools

Best Practices

Governance

4.1 Best Practices in Naming Conventions

- ISO 11179 Metadata Registry
- Complement related IT standards
- Meaningful to business users
- Minimize name changes across environments

4.2 Best Practices in Database Design (PRISM)

- Performance and ease of use
- Reusability
- Integrity
- Security
- Maintainability

5.1 Data Model and Design Quality Management

Introduction

Activities

Tools

Best Practices

Governance

5.1.1 **Develop** Data Modeling and Design **Standards**

5.1.2 **Review** Data Model and Database Design Quality

5.1.3 **Manage** Data Model Versioning and Integration

Governance teams can get bogged down here – adopt a scorecard

5.2 Data Modeling Metrics

Introduction

Activities

Tools

Best Practices

Governance

Table 11 Data Model Scorecard® Template

#	Data governance teams will usually need training and practice to answer model scoring questions	Total score	Model score	%	Comments
1	How well does the model capture the requirements?	15			
2	How complete is the model?	15			
3	How well does the model match its scheme?	10			
4	How structurally sound is the model?	15			
5	How well does the model leverage generic structures?	10			
6	How well does the model follow naming standards?	5			
7	How well has the model been arranged for readability?	5			
8	How good are the definitions?	10			
9	How consistent is the model with the enterprise?	5			
10	How well does the metadata match the data?	10			
	TOTAL SCORE	100			

- How well does the model capture the current state vs. the future state?
- Can the model measure progress from present to future state?
- Does the model illustrate requirements gaps?
- Does the model include data in motion as well as at rest?
- Does the model document known variations and inconsistencies across the enterprise?
- Can the model be used as a data roadmap for digital transformation?

(Data Modeling for Digital Transformation with Data Vault, Noel 2020 DVA)

Essential Concepts Deep Dive

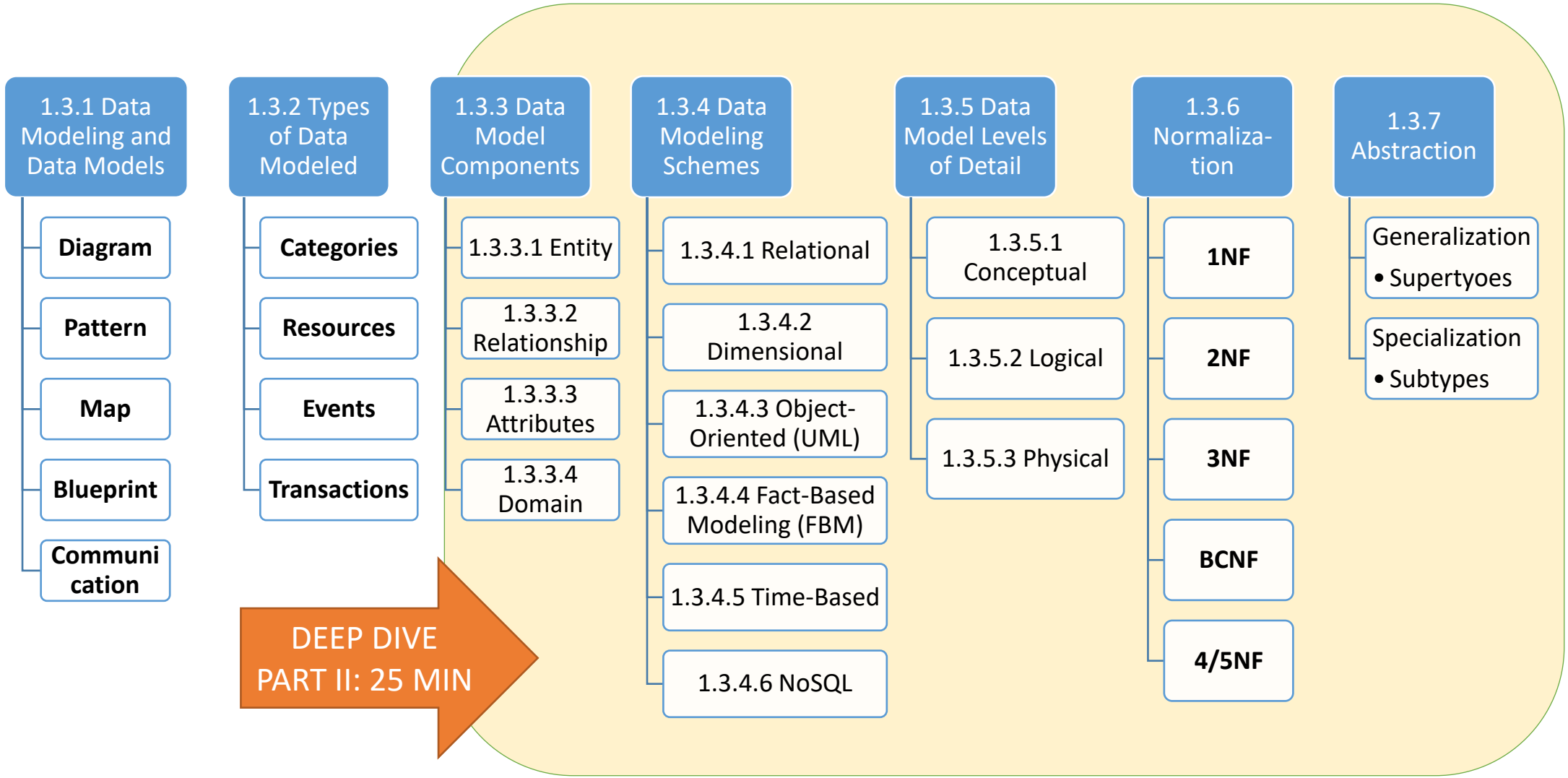
Components

Schemes

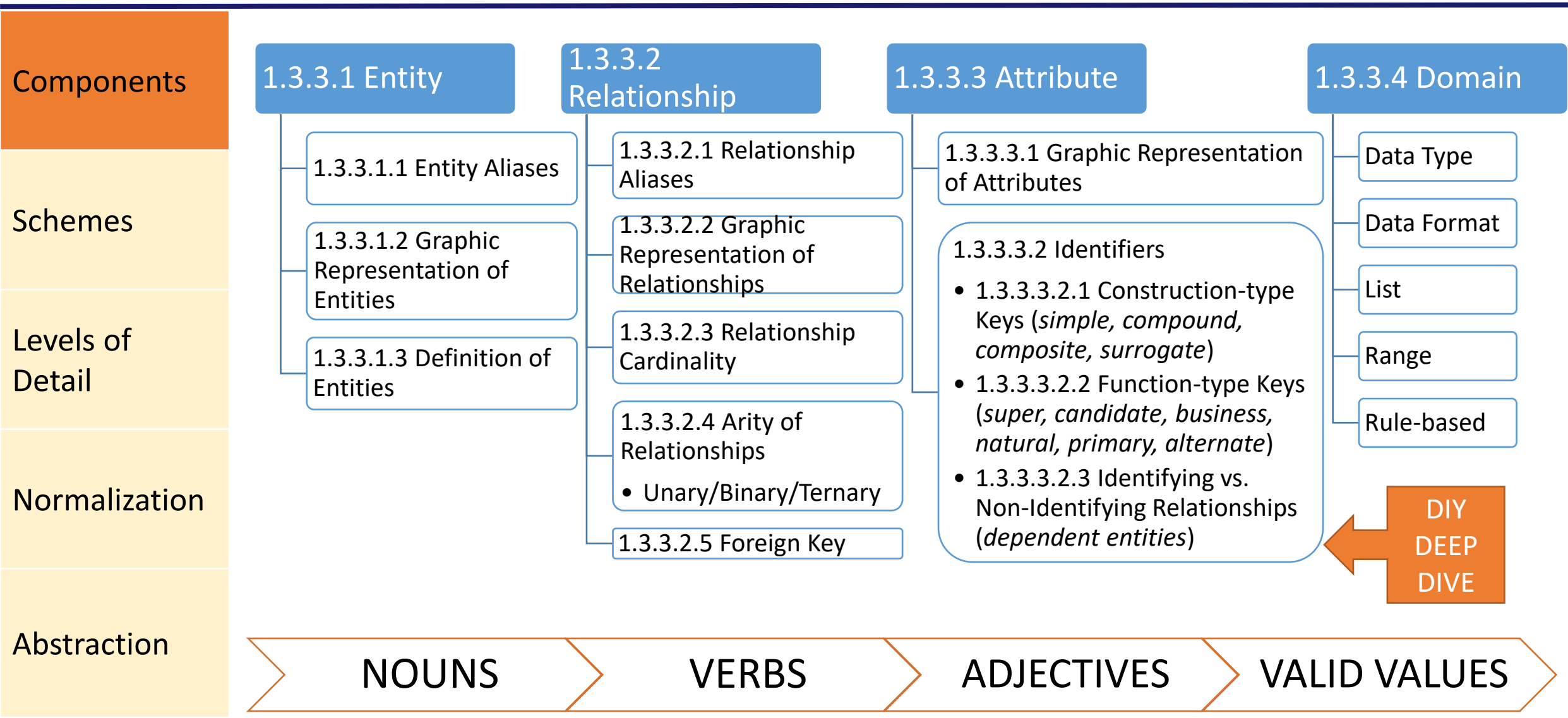
Levels of Detail

Normalization

Abstraction



1.3.3 Data Model Components



1.3.3 Data Model Components – Graphic Representation

- Components
- Schemes
- Levels of Detail
- Normalization
- Abstraction

E/R in ERD

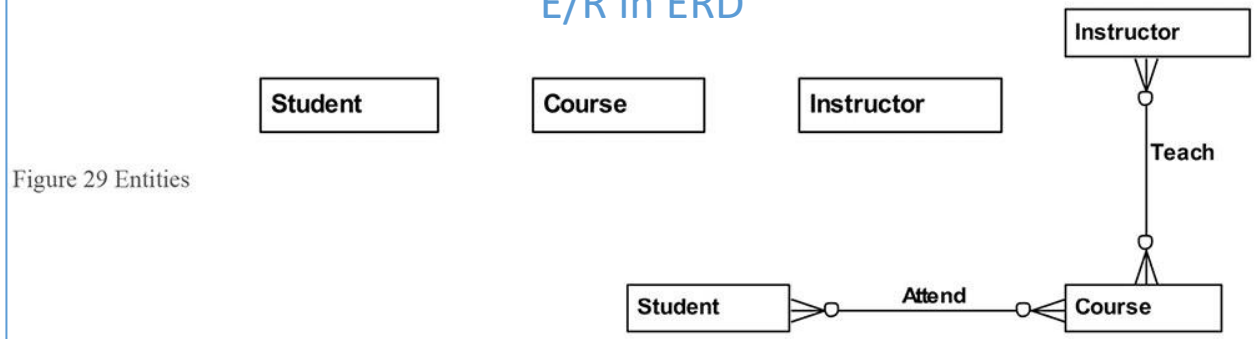


Figure 29 Entities

Figure 30 Relationships

CARDINALITY (HOW MANY)

Figure 31 Cardinality Symbols

ATTRIBUTES & FOREIGN KEYS

Figure 36 Foreign Keys

ASSOCIATIVE ENTITY

ARITY (# ENTITIES)

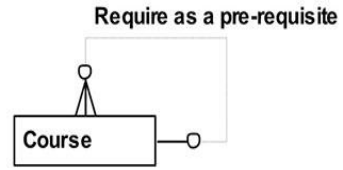
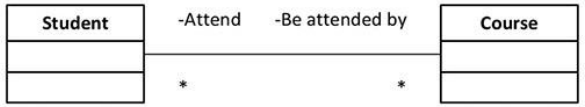


Figure 32 Unary Relationship - Hierarchy



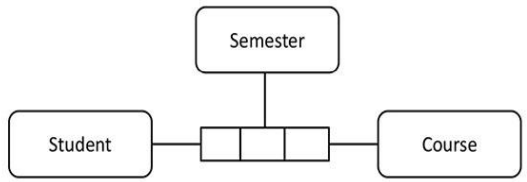


Figure 35 Ternary Relationship

1.3.4 Data Modeling Schemes

Components

Schemes

Levels of Detail

Normalization

Abstraction

1.3.4.1 Relational

- one fact in one place
- Information Engineering (IE)
- Integration Definition for Information Modeling (IDEF1X)

1.3.4.2 Dimensional

- 1.3.4.2.1 Fact Tables
- 1.3.4.2.2 Dimension Tables
- 1.3.4.2.3 Snowflaking
- 1.3.4.2.4 Grain
- 1.3.4.2.5 Conformed Dimensions
- 1.3.4.2.6 Conformed Facts

1.3.4.3 Object-Oriented (UML)

- Classes
- Attributes
- Operations

1.3.4.4 Fact-Based Modeling (FBM)

- 1.3.4.4.1 Object Role Modeling (ORM or ORM2)
- 1.3.4.4.2 Fully Communication Oriented Modeling (FCO-IM)

1.3.4.5 Time-Based

- 1.3.4.5.1 Data Vault
- 1.3.4.5.2 Anchor Modeling

1.3.4.6 NoSQL

- 1.3.4.6.1 Document
- 1.3.4.6.2 Key-value
- 1.3.4.6.3 Column-oriented
- 1.3.4.6.4 Graph

Mid- to large enterprises usually have an app landscape with multiple schemes and models evolved over time

POPULAR FOR BIG DATA

1.3.4 Data Modeling Schemes and Graphical Notation

- Components
- **Schemes**
- Levels of Detail
- Normalization
- Abstraction

AXIS

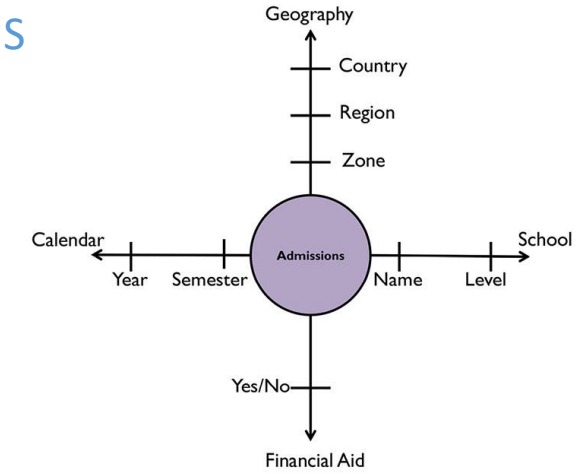
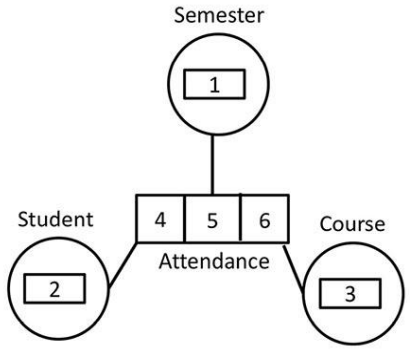


Figure 40 Axis Notation for Dimensional Models

FCO-IM



DATA VAULT

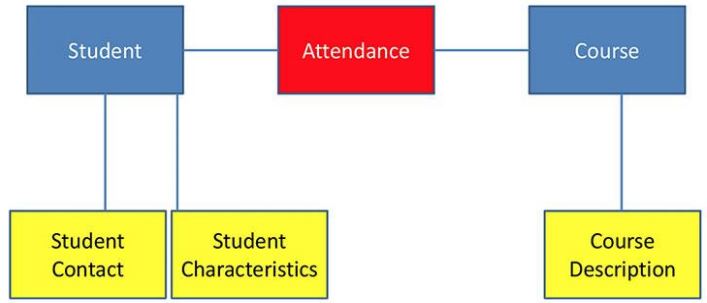


Figure 44 Data Vault Model

UML

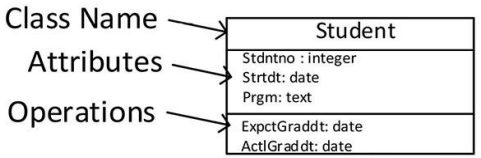


Figure 41 UML Class Model

BPMN is increasingly used as business-readable notation for data modeling

ANCHOR

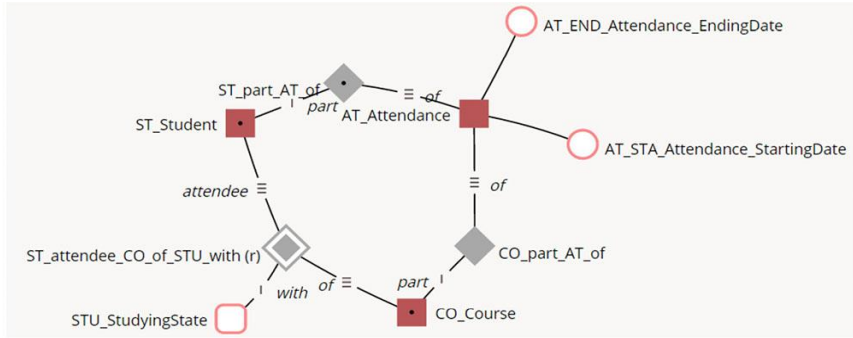


Figure 45 Anchor Model

ORM

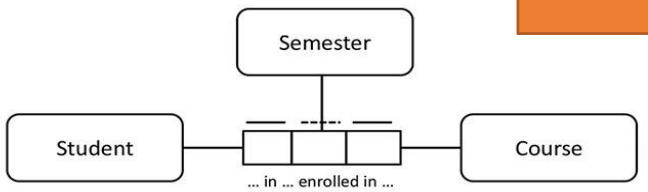
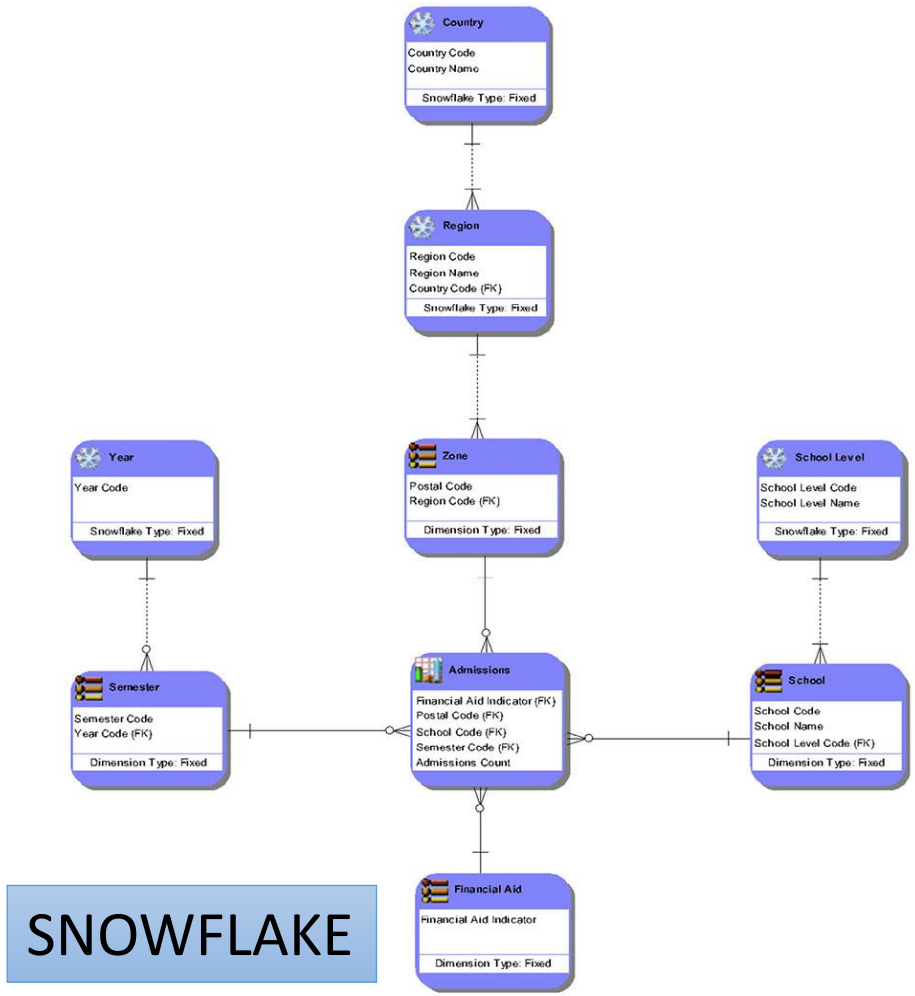


Figure 42 ORM Model

1.3.5 Data Model Levels of Detail – LDM vs. PDM Examples

- Components
- Schemes
- Levels of Detail
- Normalization
- Abstraction



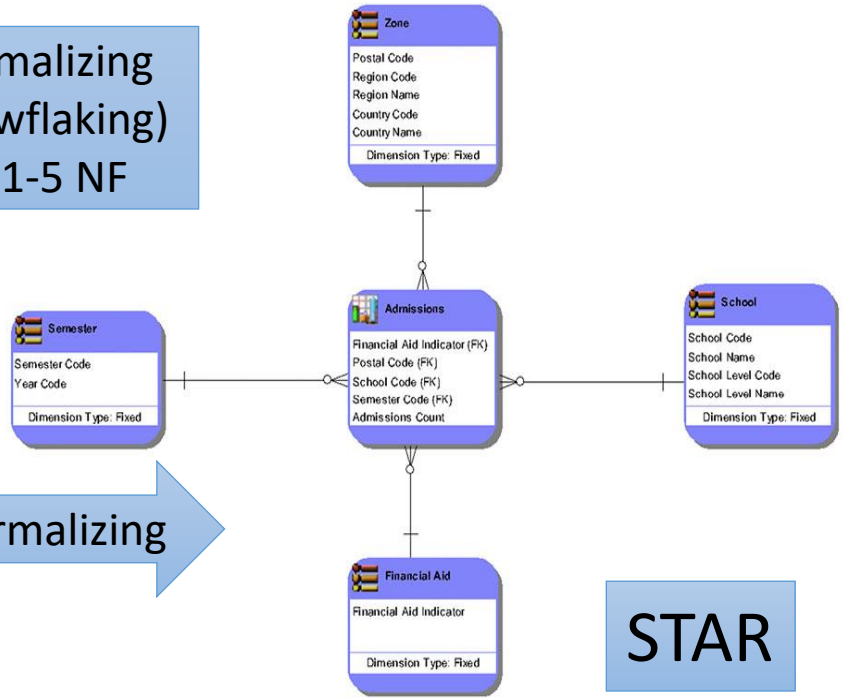
SNOWFLAKE

Figure 49 Dimensional Logical Data Model

- Country, Region, Postal Code included in Zone
- School Level included in School
- Year included in Semester

Normalizing (Snowflaking) To 1-5 NF

Denormalizing



STAR

Figure 51 Dimensional Physical Data Model

1.3.5 Data Model Levels of Detail – CDM, LDM, PDM Deep Dive

- Components
- Schemes
- **Levels of Detail**
- Normalization
- Abstraction

1.3.5.1 Conceptual

- Scheme
- Notation
- Nouns & Verbs
- Terminology
- Signoff

1.3.5.2 Logical

- Requirements
- Documentation
- Entities
- Attributes
- Domains
- Keys

1.3.5.3 Physical

1.3.5.3.1 Canonical
- ESB/EAI for data in motion

→ INTEGRATION

1.3.5.3.2 Views
- Virtual tables

→ SECURITY & ANALYTICS

1.3.5.3.3 Partitioning
- Vertical or horizontal

→ STORAGE & OPERATIONS

1.3.5.3.4 Denormalization
- Reduce runtime joins and/or calculations

→ STORAGE & OPERATIONS

1.3.6 Normalization Rules

Each level comprises a separate normal form, and each successive level does not need to include previous levels.

- Components
- Schemes
- Levels of Detail
- Normalization
- Abstraction

First normal form (1NF)	Second normal form (2NF)	Third normal form (3NF)	Boyce / Codd normal form (BCNF)	Fourth normal form (4NF)	Fifth normal form (5NF)
• each entity has a **valid** primary key • every attribute depends on the primary key • removes repeating groups • each attribute is atomic (not multi-valued) • resolution of many-to-many relationships with **associative entities**	• each entity has the **minimal** primary key • every attribute depends on the **complete** primary key	• each entity has no hidden primary keys • each attribute depends on no attributes outside the key • **"the key, the whole key and nothing but the key"**	• Resolves **overlapping composite** candidate keys. • A **candidate** key is either a primary or an alternate key. • **Composite** means more than one • **Overlapping** means there are hidden business rules	• Resolves all **many-to-many-to-many relationships** (and beyond) in pairs until they cannot be broken down into any smaller pieces.	• Resolves **inter-entity dependencies** into basic pairs • all join dependencies use parts of primary keys.

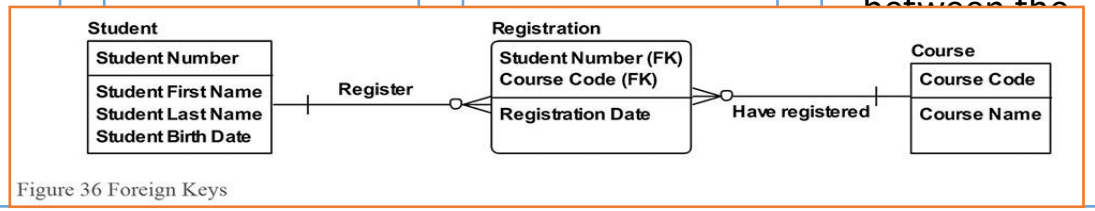
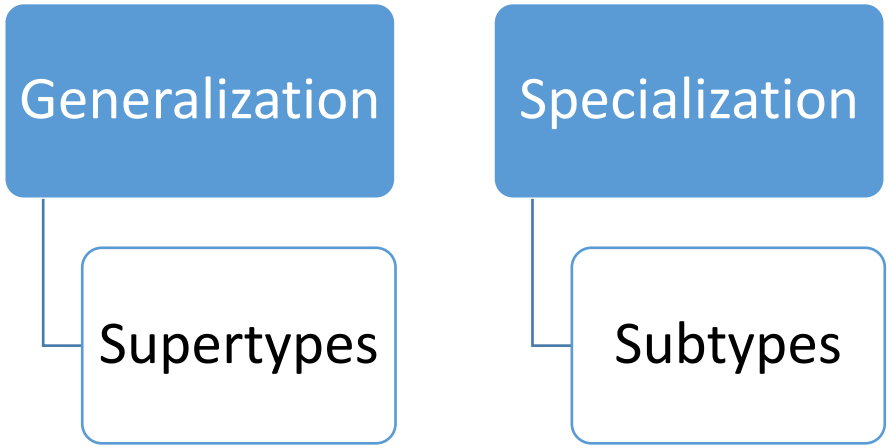


Figure 36 Foreign Keys

Situations requiring BCNF, 4NF, and 5NF occur rarely.

1.3.7 Abstraction

- Components
- Schemes
- Levels of Detail
- Normalization
- Abstraction



Not all modelers or developers are comfortable with, or have the ability to work with abstraction. The modeler needs to weigh the cost of developing and maintaining an abstract structure versus the amount of rework required if the unabstracted structure needs to be modified in the future (Giles 2011).

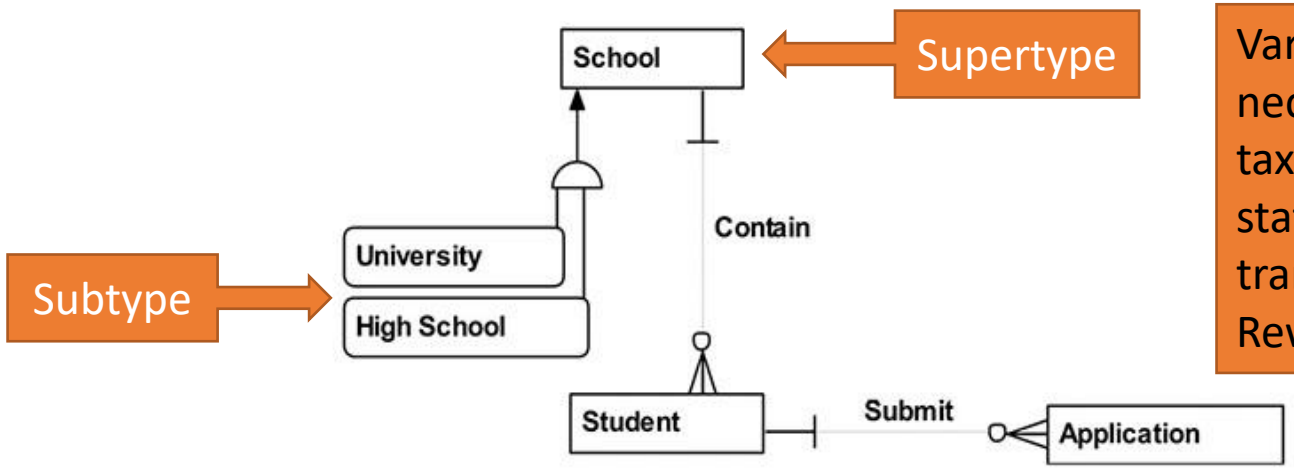


Figure 52 Supertype and Subtype Relationships

Varying degrees of abstraction are necessary for developing an enterprise taxonomy, reconciling legacy and future-state models, and executing digital transformation at speed and scale. Rework is a byproduct of Agile methods.

Chapter Recap

PART I

- **1. Introduction**
  - 1.1 Business Drivers
  - 1.2 Goals & Principles
  - 1.3 Essential Concepts
    - 1.3.1 Data Modeling and Data Models
    - 1.3.2 Types of Data Modeled
    - **1.3.3 Data Model Components**
    - 1.3.4 Data Modeling Schemes
    - **1.3.5 Data Model Levels of Detail**
    - **1.3.6 Normalization**
    - 1.3.7 Abstraction
- **2. Activities**
  - 2.1 Plan for Data Modeling
  - 2.2 Build the Data Model
  - 2.3 Review the Data Models
  - 2.4 Maintain the Data Models
- **3. Tools**
  - 3.1 Data Modeling Tools
  - 3.2 Lineage Tools
  - 3.3 Data Profiling Tools
  - 3.4 Metadata Repositories
  - 3.5 Data Model Patterns
  - 3.6 Industry Data Models
- **4. Best Practices**
  - 4.1 Best Practices in Naming Conventions
  - 4.2 Best Practices in Database Design
- **5. Data Model Governance**
  - 5.1 Data Model and Design Quality Management
  - 5.2 Data Modeling Metrics

DEEP DIVE PART II

"Machine automation, platform apps and data catalogs, and collaborative crowdsourcing are disrupting and reinventing data modeling and design as essential to digital transformation." (Noel 2020)

DAMA-NE 5/27: Aligning Data Management With Digital Transformation

Q & A

NEXT SESSION

Date	Topic and Links to Materials	Presenter(s)
February 2nd	Overview & Introduction & Chapter 1: Data Management	Agnes Vega &Laura Sebastian-Coleman
February 16th	Chapter 2: Data Handling Ethics	Lynn Noel
March 2nd	Chapter 3: Data Governance	Laura Sebastian-Coleman
March 16th	Chapter 4: Data Architecture	Laura Sebastian-Coleman
March 30th	Chapter 6: Data Storage & Operations	Karen Sheridan
April 13th	Chapter 5: Data Modeling & Design	Lynn Noel
April 27th	Chapter 7: Data Security	Laura Sebastian-Coleman
May 11th	Chapter 8: Data Integration & Interoperability	TBD
May 25th	Chapter 9: Document & Content Management	Sandi Perillo-Simmons
June 8th	Chapter 10: Reference & Master Data	TBD
June 22nd	Chapter 11: Data Warehousing & Business Intelligence	Mukta Mohindra
Summer Break	**Enjoy**	
August 10th	Chapter 12: Metadata Management	Karen Sheridan
August 24th	Chapter 13: Data Quality	Laura Sebastian-Coleman
September 7th	Chapter 14: Big Data & Data Science	Nupur Gandhi
September 21st	Chapter 15: Data Management Maturity Assessment	Sandi Perillo-Simmons
October 5th	Chapter 16: Data Management Organization & Role Expectations	Agnes Vega
October 19th	Chapter 17: Data Management & Organizational Change Management	TBD
November 2nd	Final Review	TBD