COURSEWARE

DATA MODELLING AND DESIGN

BASED ON CDMP

Van Haren PUBLISHING

MICHEL DEKKER

Data Modelling and Design
Based on CDMP

Colophon

Title:	Data Modelling and Design
Subtitle:	Based on CDMP
Authors:	Michel Dekker
Publisher:	Van Haren Publishing, ‘s-Hertogenbosch
ISBN Hard Copy:	978 94 018 1378 5
Edition:	First edition, first print, October, 2025
Design:	Van Haren Publishing, ‘s-Hertogenbosch
Copyright:	© Van Haren Publishing 2025

For further information about Van Haren Publishing please e-mail us at: info@vanharen.net or visit our website: www.vanharen.net

Publisher about the Courseware

The Courseware was created by experts from the industry who served as the author(s) for this publication. The input for the material is based on existing publications and the experience and expertise of the author(s). The material has been revised by trainers who also have experience working with the material. Close attention was also paid to the key learning points to ensure what needs to be mastered.

The objective of the courseware is to provide maximum support to the trainer and to the student, during his or her training. The material has a modular structure and according to the author(s) has the highest success rate should the student opt for examination. The Courseware is also accredited for this reason, wherever applicable.

In order to satisfy the requirements for accreditation the material must meet certain quality standards. The structure, the use of certain terms, diagrams and references are all part of this accreditation. Additionally, the material must be made available to each student in order to obtain full accreditation. To optimally support the trainer and the participant of the training assignments, practice exams and results are provided with the material.

Direct reference to advised literature is also regularly covered in the sheets so that students can find additional information concerning a particular topic. The decision to leave out notes pages from the Courseware was to encourage students to take notes throughout the material. Although the courseware is complete, the possibility that the trainer deviates from the structure of the sheets or chooses to not refer to all the sheets or commands does exist. The student always has the possibility to cover these topics and go through them on their own time. It is recommended to follow the structure of the courseware and publications for maximum exam preparation.

The courseware and the recommended literature are the perfect combination to learn and understand the theory.

-- Van Haren Publishing

Other publications BY Van Haren Publishing

Van Haren Publishing (VHP) specializes in titles on Best Practices, methods, frameworks and standards within four domains:
- IT Management
- Architecture (Enterprise and IT)
- Business Management and
- Project Management

Van Haren Publishing is publishing on behalf of leading organizations and companies: Agile Consortium, World Commerce and Contracting, IAOP, IPMA World, KNVI, PMI-NL, NLAIC and The Open Group.

Van Haren Publishing is part of the Van Haren Group and additional to the book publishing also provides the following services: accredited training materials and e-learning through Van Haren Learning Solutions, as well as independent professional certification via examination through Van Haren Certify.

Topics are (per domain):

IT Management	IT Service Management	FitSM, ISM®, ISO/IEC20000, IT4IT®, ITIL®, VerISM®, SAF, TRIM, XLA®
	Data Management	Data literacy, Data visualization, DMBOK
	IT Asset Management	HAM, ITAM, SAM
	IT Security Management	BIO, ISO/IEC27001, NIS2
	Test Management	CTAP
	Application Management	ASL
	Other	eCF, IT-CMF, Scrum
Project Management	Project Management	Half Double, ICB, ISO/IEC21500, P3.express, PM2, PMBOK Guide, Praxis, PRINCE2
	Agile	Agile, Agile PM
	Other	PMO
Business Management	Operations Management	Lean, Lean Six Sigma, OBM, OMC, RASCI
	Contract Management	CATS CM, CATS RVM, IACCM World
	Business Information Management	BiSL, DID
	Artificial Intelligence	AI, Generative AI
	Outsourcing	OPBOK
Enterprise Architecture	Enterprise Architecture	BIAN, TOGAF
	Modeling	ArchiMate, BPMN
	Software Architecture	ISAQB
	Other	Open Agile Architecture

For the latest information on VHP publications, visit our website: www.vanharen.net.

Table of Content

Self-Reflection of understanding Diagram

'What you do not measure, you cannot control.'' – Tom Peters

Fill in this diagram to self-evaluate your understanding of the material. This is an evaluation of how well you know the material and how well you understand it. In order to pass the exam successfully you should be aiming to reach the higher end of Level 3. If you really want to become a pro, then you should be aiming for Level 4. Your overall level of understanding will naturally follow the learning curve. So, it's important to keep track of where you are at each point of the training and address any areas of difficulty.

Based on where you are within the Self-Reflection of Understanding diagram you can evaluate the progress of your own training.

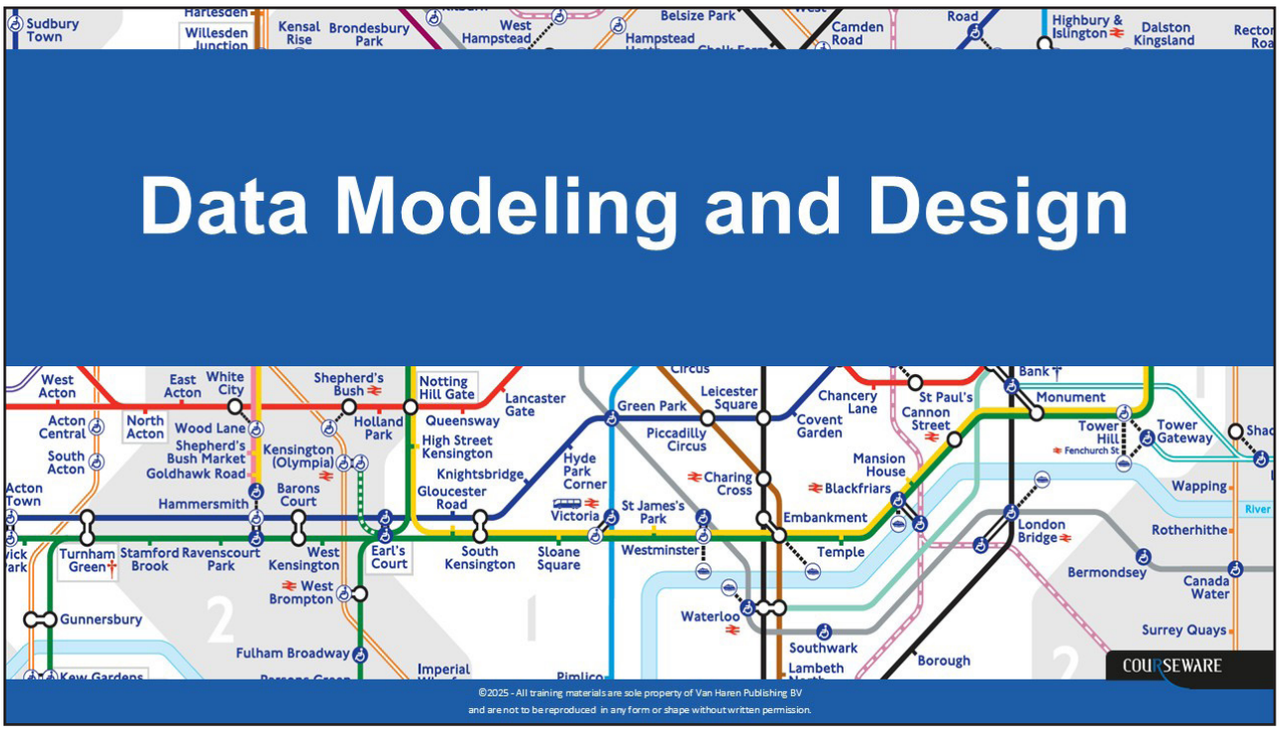
Level of Understanding	Before Training (Pre-knowledge)	Training Part 1 (1st Half)	raining Part 2 (2nd Half)	After studying / reading the book	After exercises and the Practice exam
Level 4 I can explain the content and apply it .					
Level 3 I get it! I am right where I am supposed to be.					Ready for the exam!
Level 2 I almost have it but could use more practice.					
Level 1 I am learning but don't quite get it yet.					

(Self-Reflection of Understanding Diagram)

Write down the problem areas that you are still having difficulty with so that you can consolidate them yourself, or with your trainer. After you have had a look at these, then you should evaluate to see if you now have a better understanding of where you actually are on the learning curve.

Troubleshooting

Problem areas:	*Topic:*
Part 1	
Part 2	
You have gone through the book and studied.	
You have answered the questions and done the practice exam.	



Effective Data Foundation

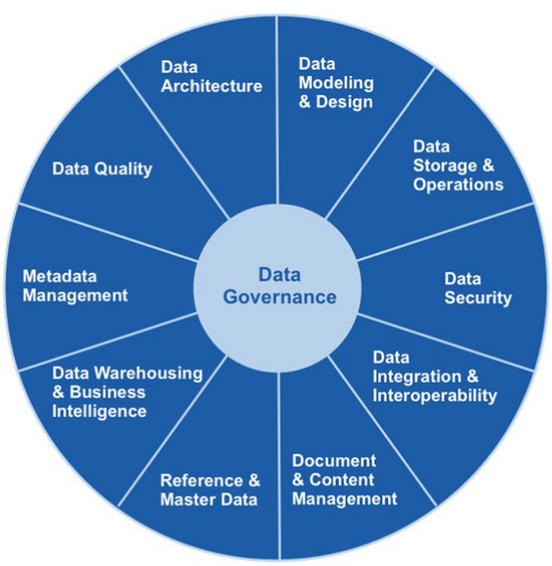
Not-for-profit collective,
which enables **professionals**
to **leverage data** to make
sustainable business decisions

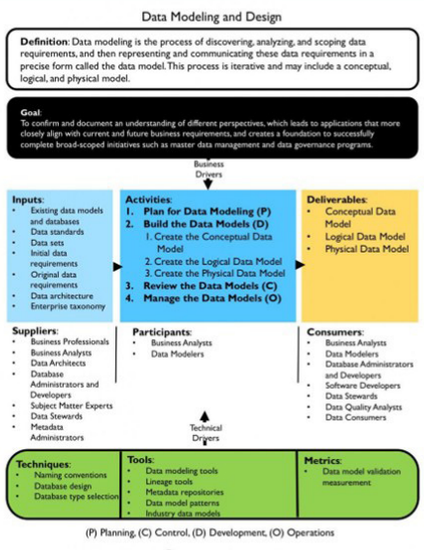
Effective Data Foundation

DATA

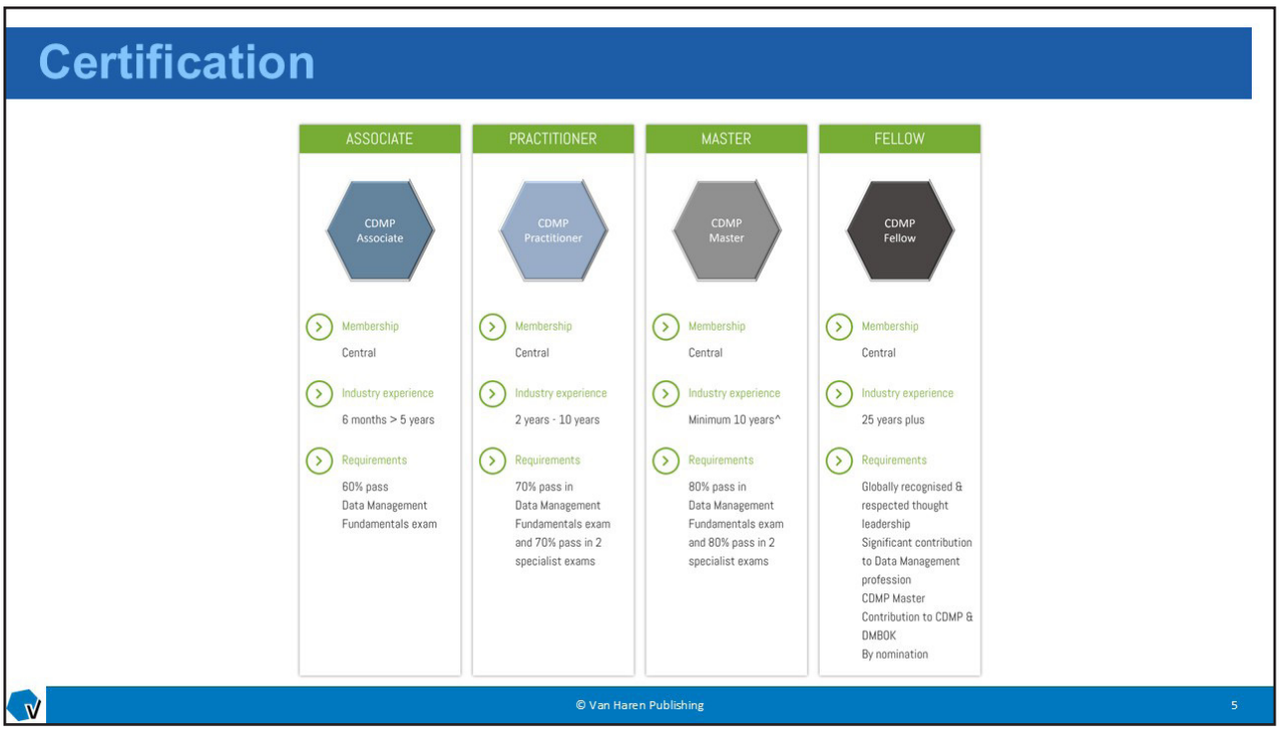
Analysis / Literacy / Management / Visualization

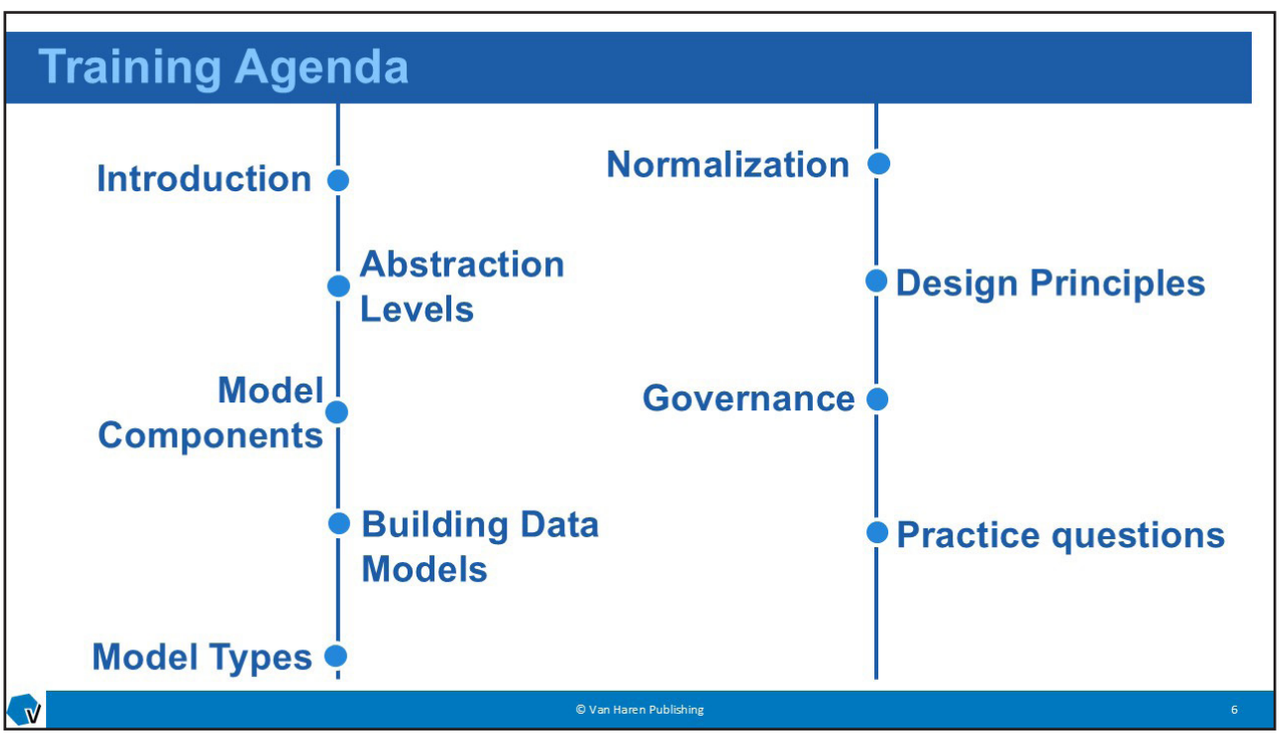
DAMA wheel





Dama.org


Certification
ASSOCIATE
CDMP Associate
Membership
Central
Industry experience
6 months > 5 years
Requirements
60% pass Data Management Fundamentals exam
PRACTITIONER
CDMP Practitioner
Membership
Central
Industry experience
2 years - 10 years
Requirements
70% pass in Data Management Fundamentals exam and 70% pass in 2 specialist exams
MASTER
CDMP Master
Membership
Central
Industry experience
Minimum 10 years^
Requirements
80% pass in Data Management Fundamentals exam and 80% pass in 2 specialist exams
FELLOW
CDMP Fellow
Membership
Central
Industry experience
25 years plus
Requirements
Globally recognised & respected thought leadership
Significant contribution to Data Management profession
CDMP Master
Contribution to CDMP & DMBOK
By nomination
© Van Haren Publishing
5

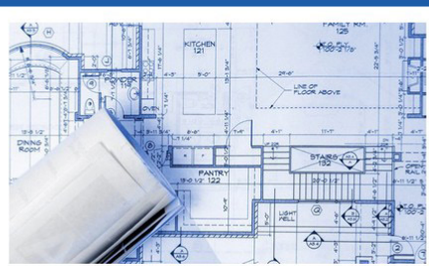

Training Agenda
Introduction
Abstraction Levels
Model Components
Building Data Models
Model Types
Normalization
Design Principles
Governance
Practice questions
© Van Haren Publishing
6

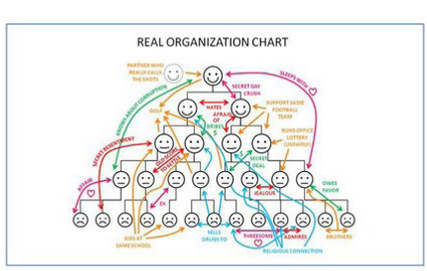
Introduction

The main purpose of a data model is not to design a database, but to ***describe a business***

C. Bradley, CDMP Fellow 2010

What is a Model?





What is Data Modeling?

The process of **discovering, analyzing and scoping** data requirements,

and then **representing and communicating** these requirements

in a **data model**

 9

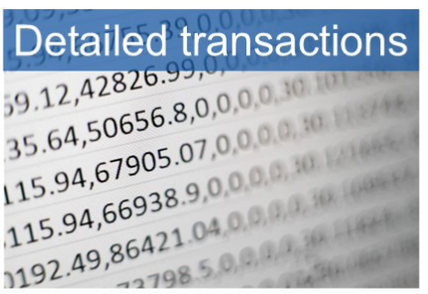
Business Drivers

Return on Modeling Effort

- **R** Common vocabulary
- **O** Capturing knowledge
- **M** Improve communication
- **E** Starting point for changes

 10

Types of data that are modelled



Abstraction levels

If you can't explain it simply, you don't understand it well enough.

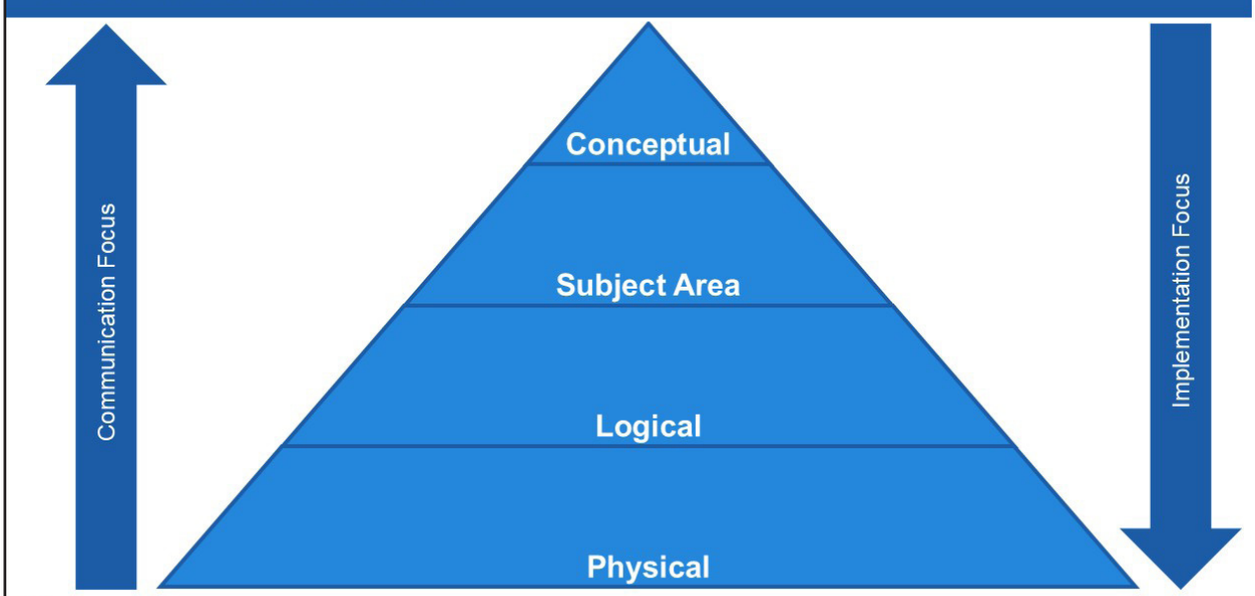
(probably not) Albert Einstein

Zachman

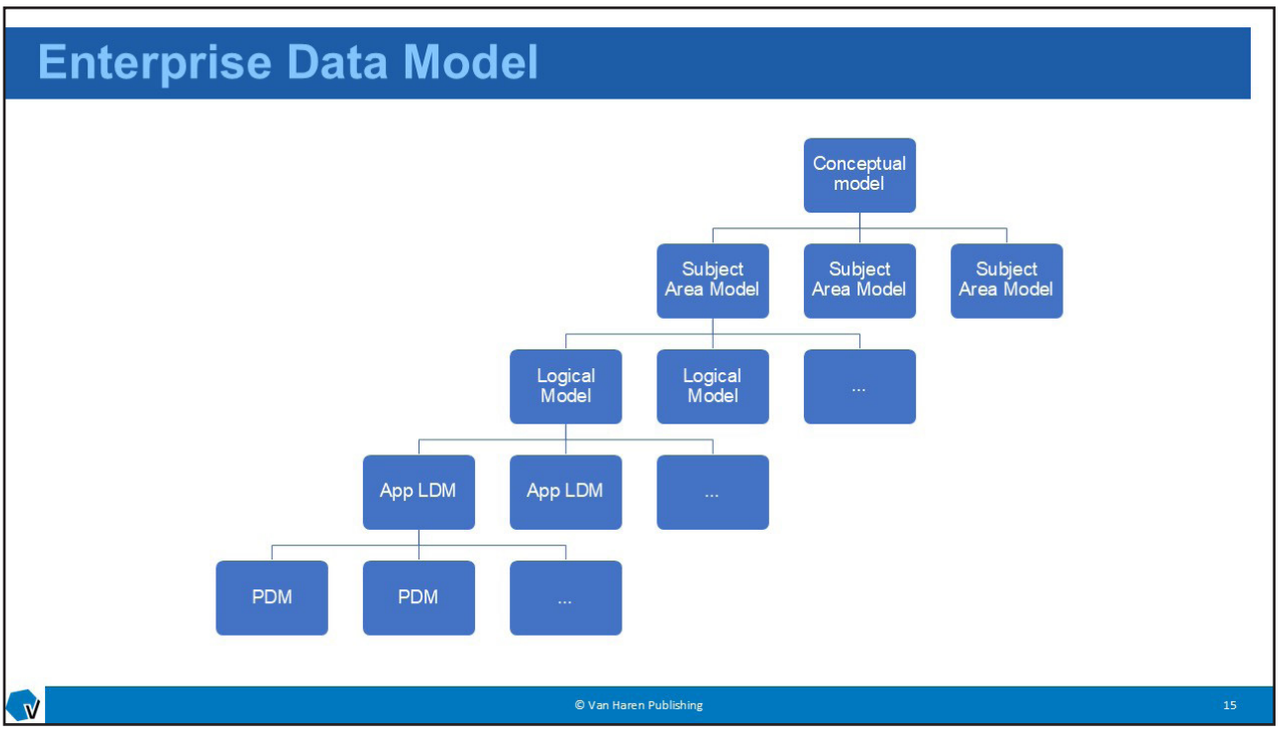
	Data What	Process How	Network Where	Role Who	Timing When	Motivation Why
Scope (contextual) Planner	List of things important to the business	List of processes the business performs	List of locations in which the business operates	List of business responsibilities	List of events significant to the business	List of business goals/strategy
Business model (conceptual) Owner	e.g., Semantic Model	e.g., Business Process Model	e.g., Logistics Networks	e.g., Workflow Model	e.g., Master Schedule	e.g., Business Plan
System model (logical) Designer	e.g., Logical Data Model	e.g., Application Architecture	e.g., Distributed System Architecture	e.g., Human Interface Architecture	e.g., Processing Structure	e.g., Business Rule Model
Technology model (physical) Builder	e.g., Physical Data Model	e.g., System Design	e.g., Configuration Design	e.g., Presentation Architecture	e.g., Control Structure	e.g., Rule Design
Detailed representations (out-of-context) Subcontractor	e.g., Data Definition	e.g., Program	e.g., Network Architecture	e.g., Security Architecture	e.g., Timing Definition	e.g., Rule Specification
Functioning enterprise	e.g., Data	e.g., Process	e.g., Network	e.g., Organization	e.g., Schedule	e.g., Strategy

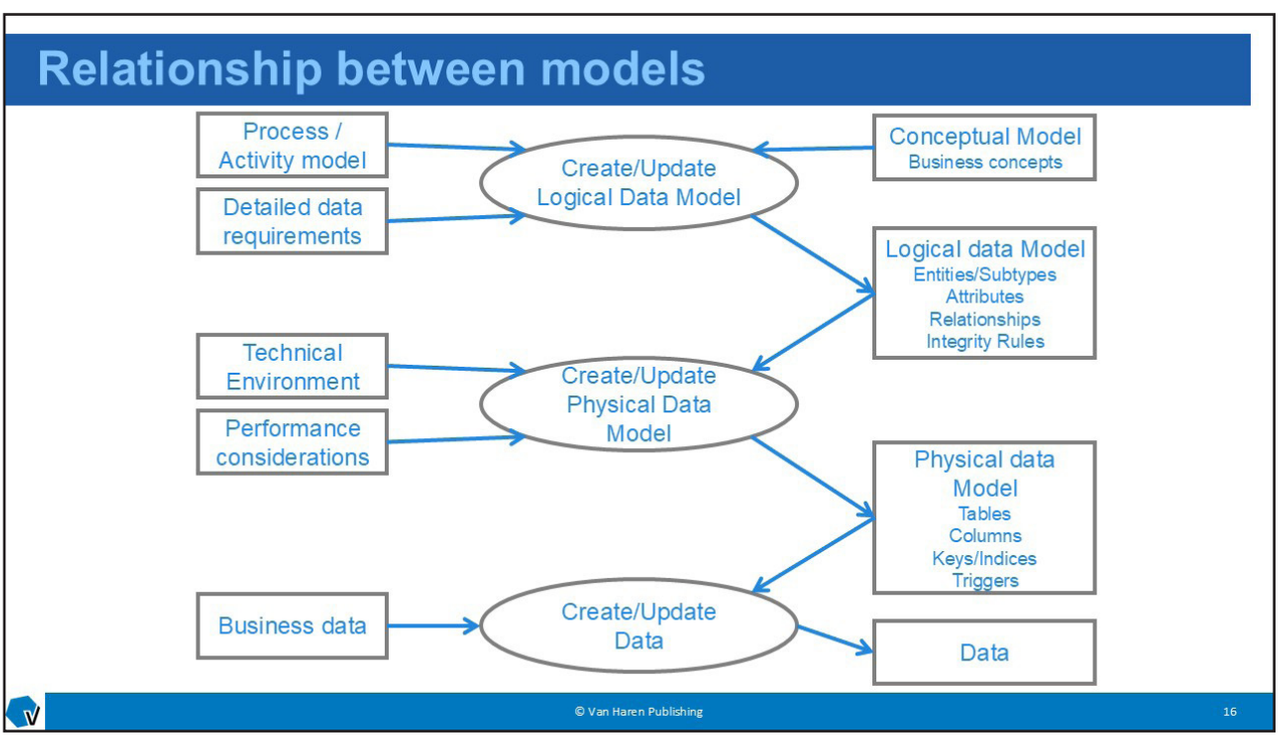
 13

Data Model Levels

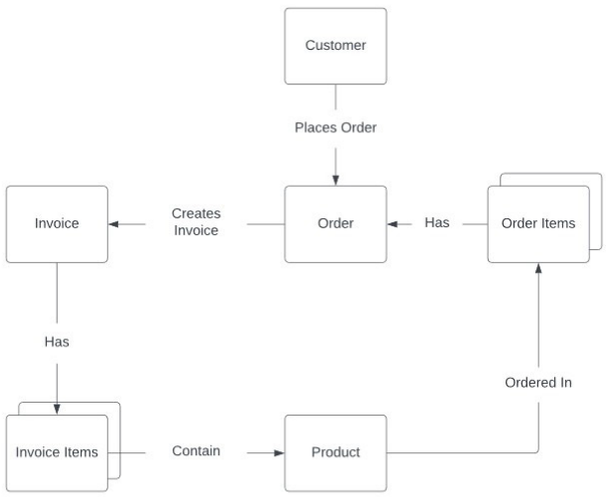


 14


Enterprise Data Model
Conceptual model
Subject Area Model
Subject Area Model
Subject Area Model
Logical Model
Logical Model
...
App LDM
App LDM
...
PDM
PDM
...
© Van Haren Publishing
15

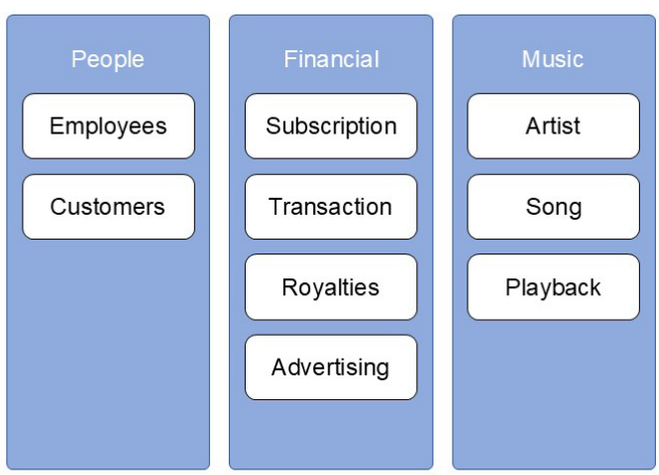

Relationship between models
Process / Activity model
Detailed data requirements
Create/Update Logical Data Model
Conceptual Model
Business concepts
Logical data Model
Entities/Subtypes
Attributes
Relationships
Integrity Rules
Technical Environment
Performance considerations
Create/Update Physical Data Model
Physical data Model
Tables
Columns
Keys/Indices
Triggers
Business data
Create/Update Data
Data
© Van Haren Publishing
16

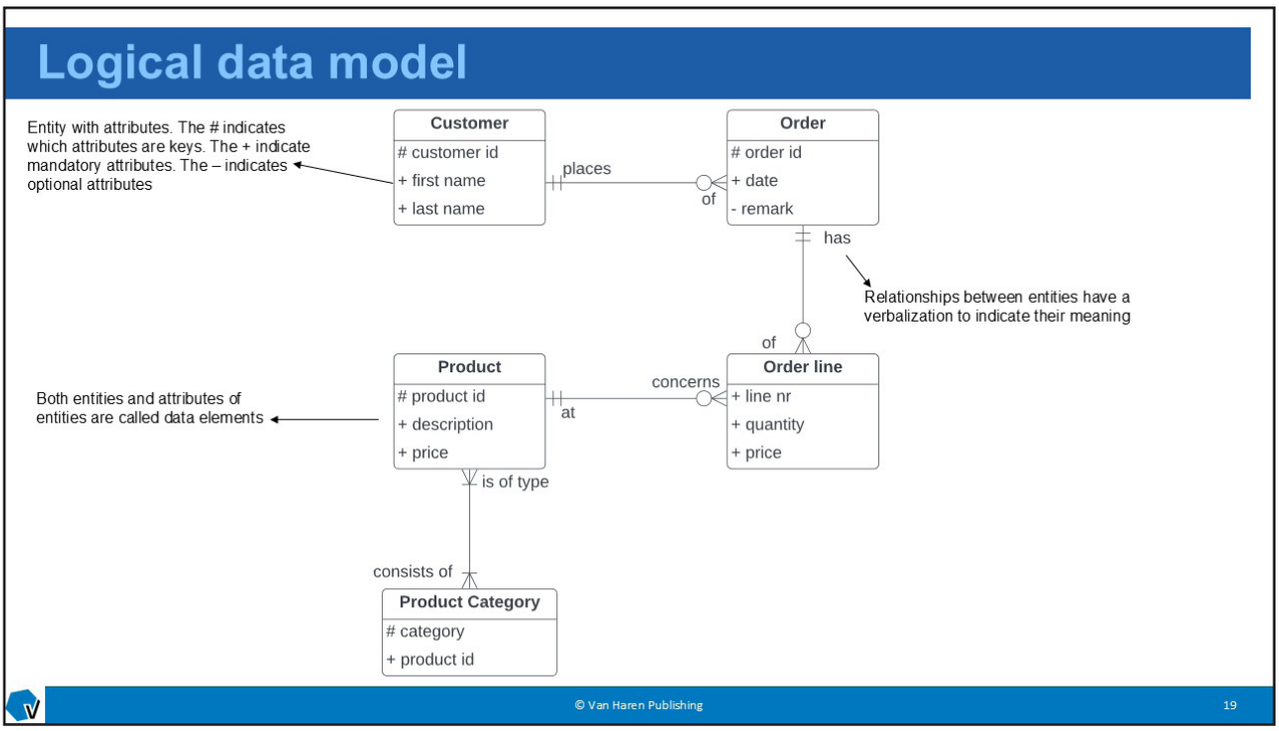
Conceptual data model

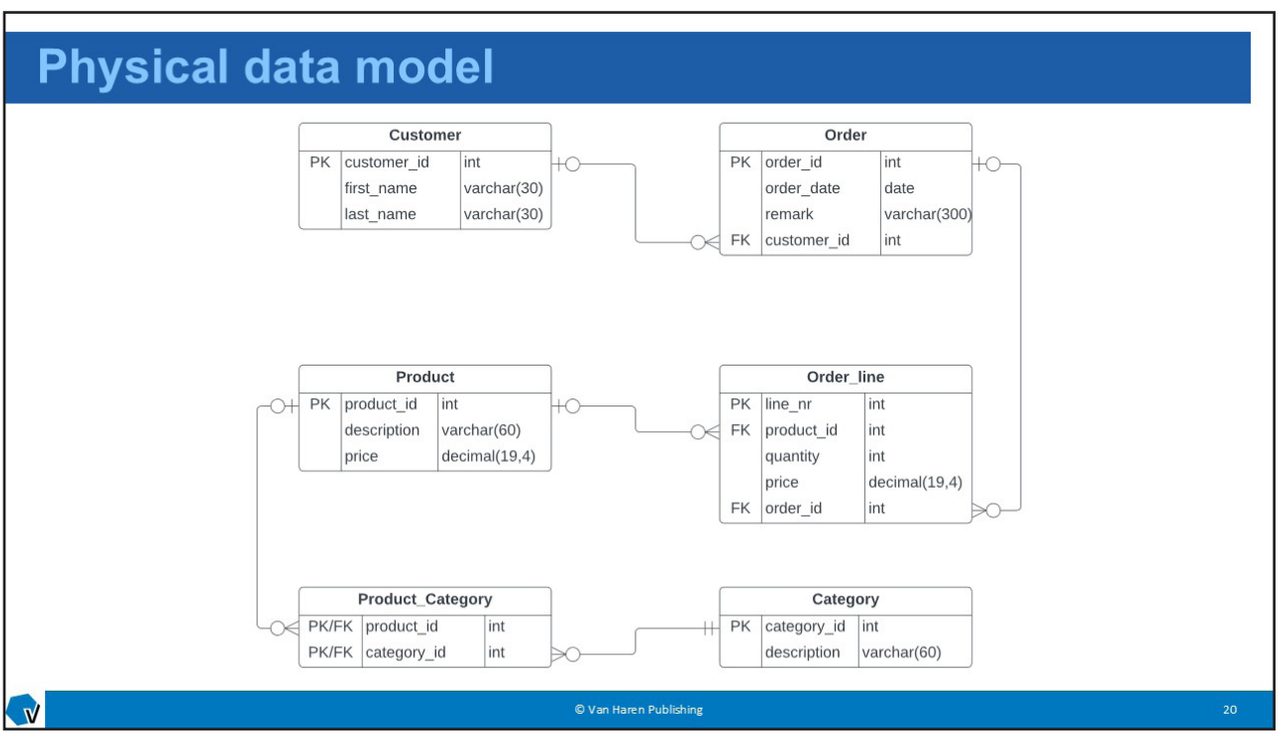


Subject Area Model

- Scope
- Abstraction
- Communication
- Understanding
- Foundation for further modeling
- Consistency




Logical data model
Entity with attributes. The # indicates which attributes are keys. The + indicate mandatory attributes. The – indicates optional attributes
Customer
customer id
+ first name
+ last name
places
of
Order
order id
+ date
- remark
has
Relationships between entities have a verbalization to indicate their meaning
of
Product
product id
+ description
+ price
concerns
at
Order line
+ line nr
+ quantity
+ price
Both entities and attributes of entities are called data elements
is of type
consists of
Product Category
category
+ product id
© Van Haren Publishing
19


Physical data model
Customer
PK customer_id int
first_name varchar(30)
last_name varchar(30)
Order
PK order_id int
order_date date
remark varchar(300)
FK customer_id int
Product
PK product_id int
description varchar(60)
price decimal(19,4)
Order_line
PK line_nr int
FK product_id int
quantity int
price decimal(19,4)
FK order_id int
Product_Category
PK/FK product_id int
PK/FK category_id int
Category
PK category_id int
description varchar(60)
© Van Haren Publishing
20

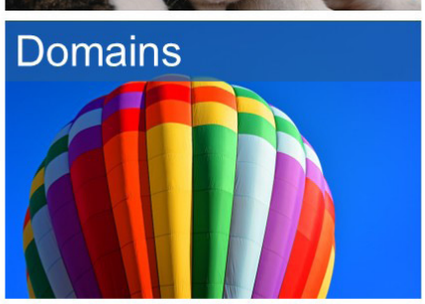
Data Model Components

*All models are **wrong**, some are **useful***

George Box, 1976

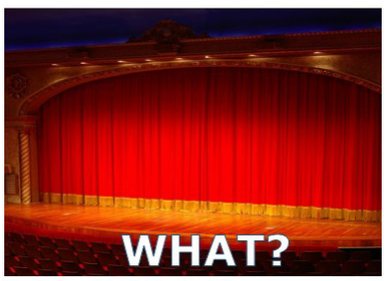
 21

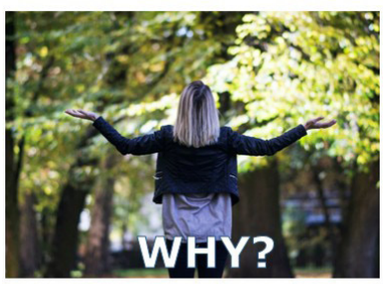
Data model components

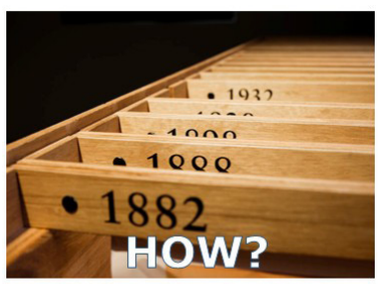


 22

Entity



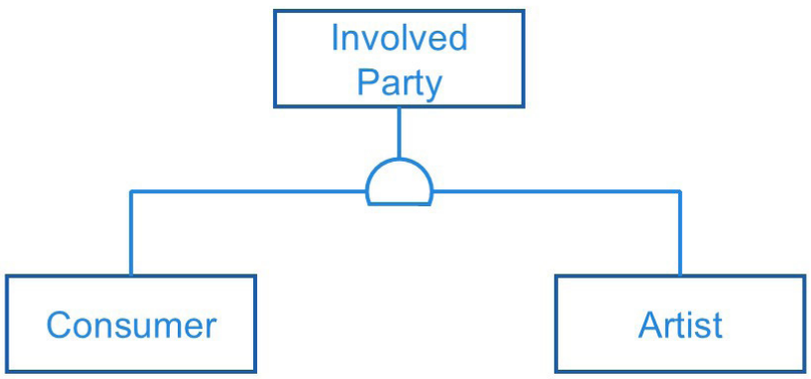




Entity

Artist	A person, or group of persons, who creates, performs or produces music as a profession.
Music Track	A recorded piece of music with melody, rhythm, and instruments.
Consumer	A person who engages with and experiences the auditory and emotional aspect of musical compositions, either by actively choosing to hear them or by passively encountering them. This person has an active *subscription* to a *music platform*.

Supertypes and Subtypes

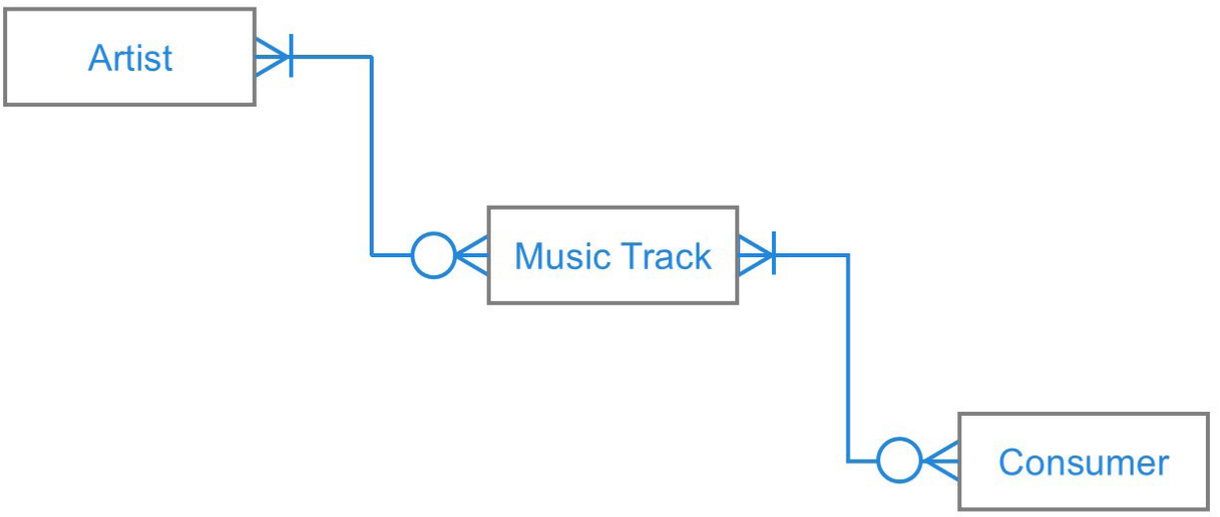


 25

Relationship

 26

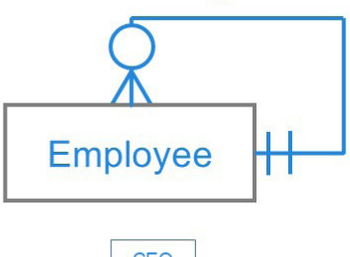
Relationship

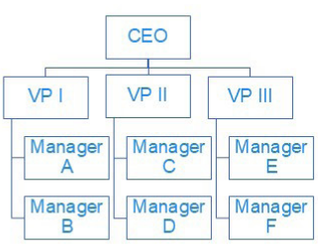


 27

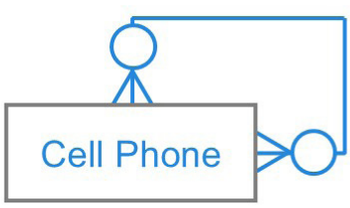
Relationship arity - Unary

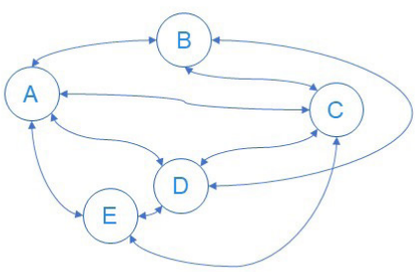
Hierarchy





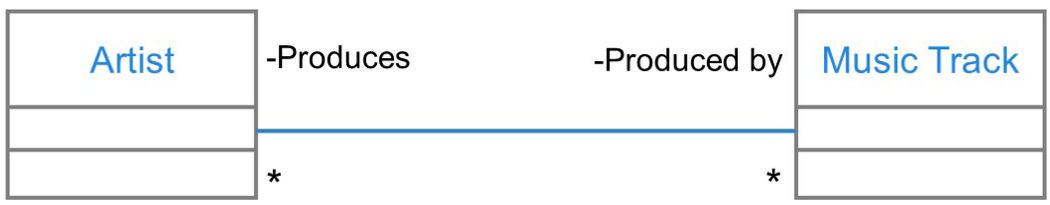
Network





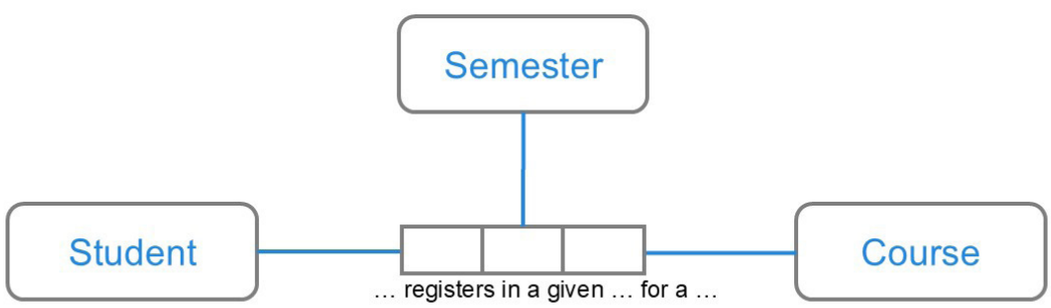
 28

Relationship arity - Binary



 29

Relationship arity - Ternary

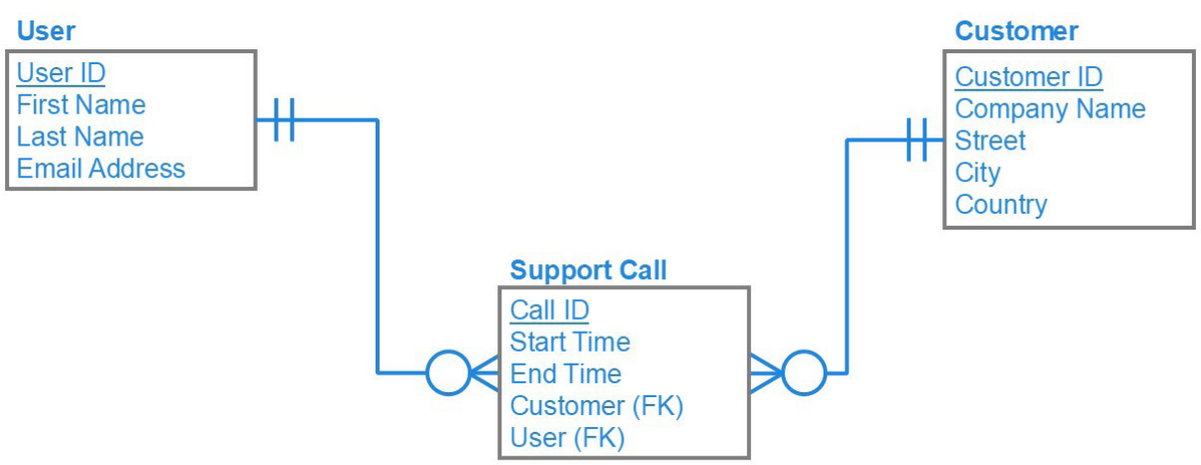


 30

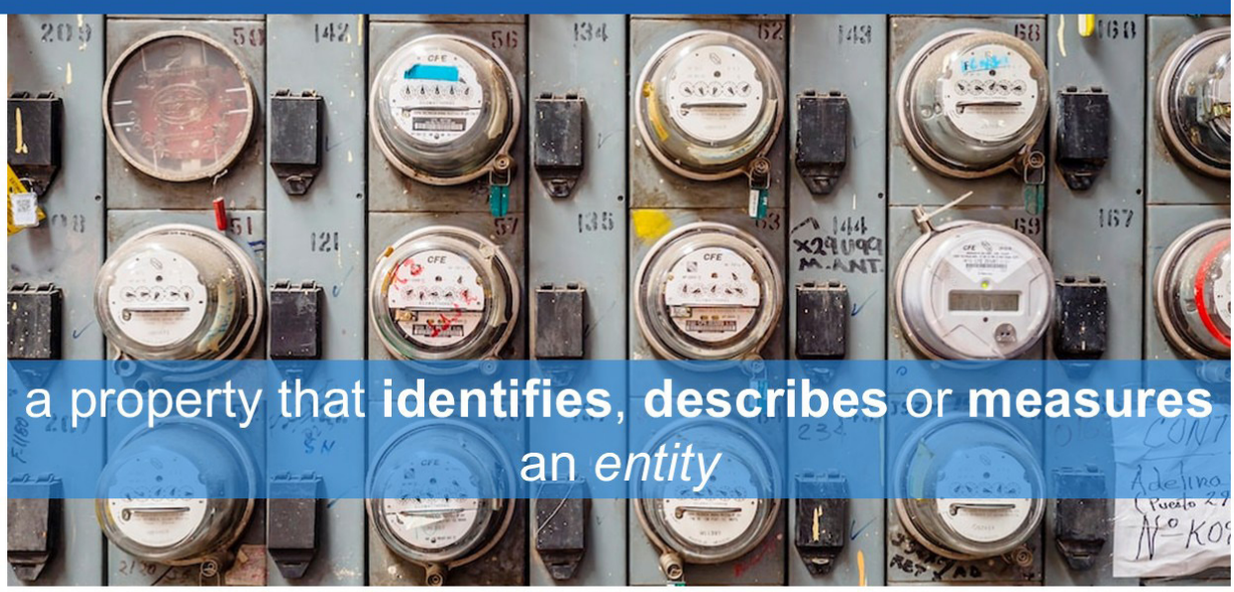
Foreign key

used in physical and sometimes logical relational data modeling schemes to represent a *relationship*

Foreign key



Attribute



 33

Attribute

User

User ID
First Name
Last Name
Email Address
Login attempts

 34

Identifier or Key

set of one or more *attributes* that uniquely defines an instance of an *entity*

Keys by construction and function

Simple Key: one attribute that uniquely identifies an entity instance

Composite Key: set of two or more attributes that together uniquely identifies an entity instance

Compound Key: *composite key* only consisting of *foreign keys*

Super Key: any set of attributes that uniquely identifies an entity instance

Candidate Key: minimal set of attributes that uniquely identifies an entity instance

Business Key: one or more attributes that a business professional would use to retrieve a single entity instance

Surrogate Key: simple key whose meaning is unrelated to its face value

Keys by construction and function

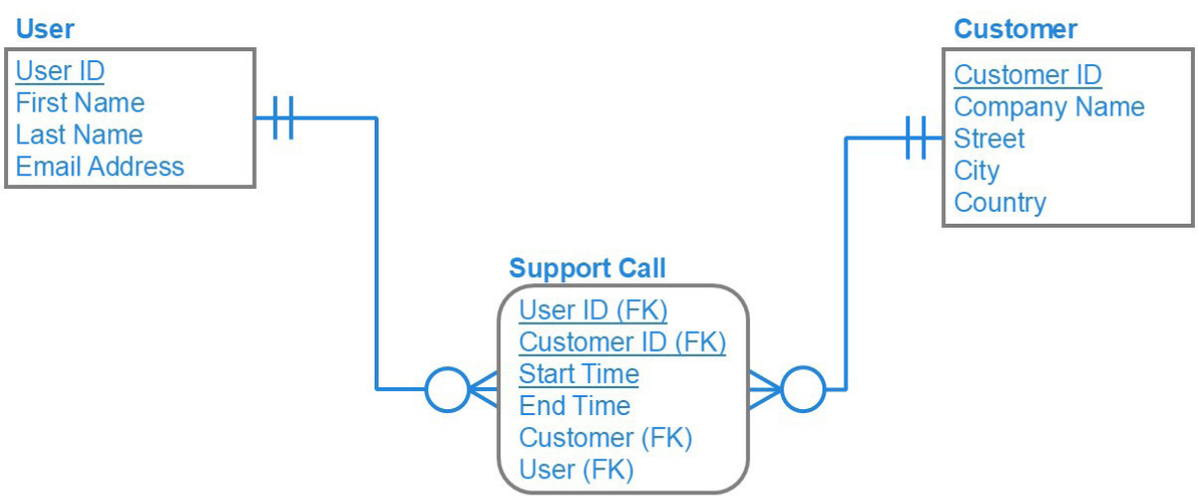
Candidate Key: minimal set of attributes that uniquely identifies an entity instance

PK **Primary Key:** *candidate key* that is chosen to be **THE** unique identifier for an entity

AK **Alternate Key:** *candidate key that* although unique, was **NOT** chosen as the *primary key*

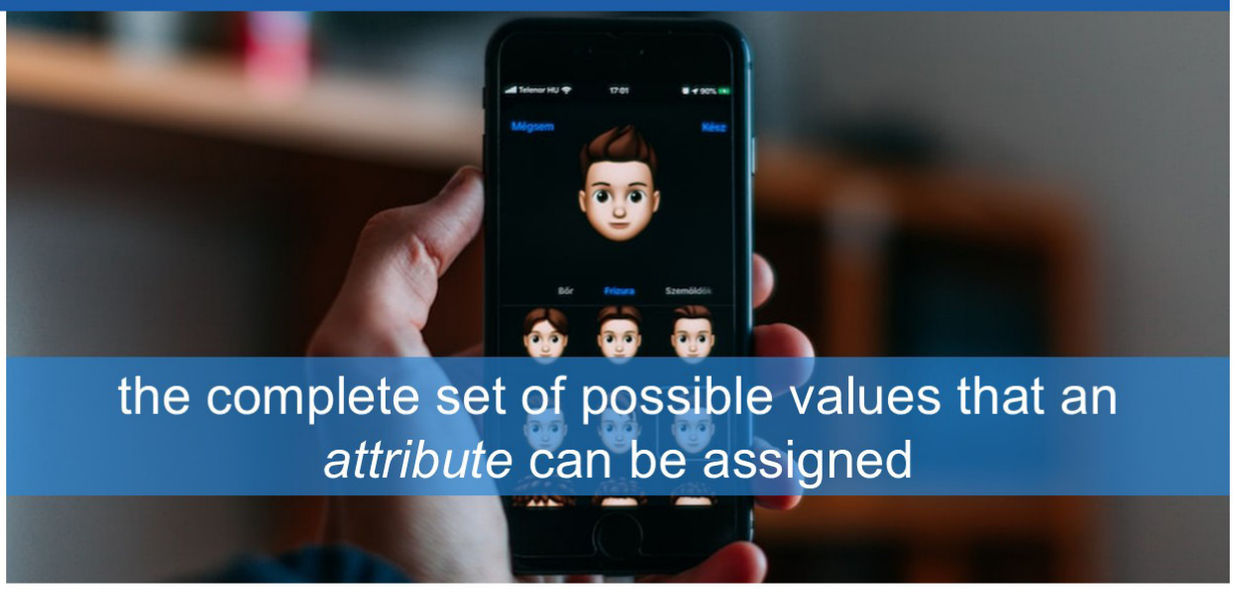
 37

Identifying vs. non-identifying relationships



 38

Domain



the complete set of possible values that an *attribute* can be assigned

 39

Data types

	SQL Server	Oracle DB
Numeric	INT BIGINT SMALLINT TINYINT DECIMAL FLOAT REAL MONEY SMALLMONEY	NUMBER INTEGER FLOAT BINARY_FLOAT BINARY_DOUBLE
Character Strings	CHAR VARCHAR TEXT NCHAR NVARCHAR NTEXT	CHAR VARCHAR2 NCHAR NVARCHAR2 CLOB NCLOB
Date and Time	DATE TIME DATETIME SMALLDATETIME DATETIME2 DATETIMEOFFSET	DATE TIMESTAMP TIMESTAMP WITH TIME ZONE TIMESTAMP WITH LOCAL TIME ZONE INTERVAL YEAR TO MONTH INTERVAL DAY TO SECOND
Binary Data	BINARY VARBINARY IMAGE	RAW BLOB
Boolean	BIT	BOOLEAN
Other	UNIQUEIDENTIFIER XML	BINARY_INTEGER

 40

Cardinality & notations

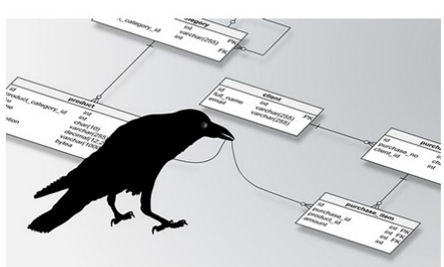
describes a fundamental relationship between two entities. There are three types: ***one-to-one****,* ***one-to-many****, and* ***many-to-many***

Leigh van der Veen

 41

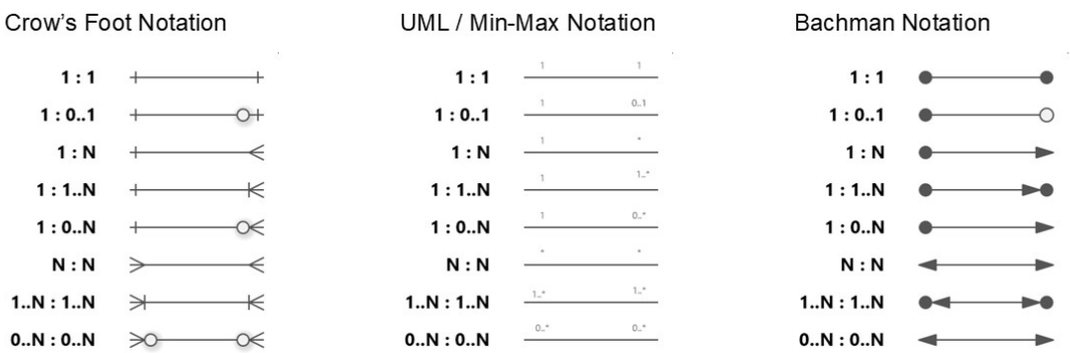
Entity Relationship Symbols

Data Model Notation	Representation	Description
Boxes		Entity
Lines		Relationship between two entities
Dotted Lines		Non-identifying relationship
Solid Lines		Identifying relationship
Crow's feet		"Many" (>1) in an N-to-many relationship
Bar		Mandatory relationship (1 or more instances)
Option		Optional relationship (zero or more instances)



 42

Entity Relationship Symbols



 43

Creating data models

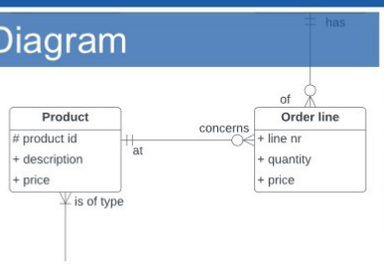
Data is a tool for ***enhancing intuition***

Hilary Mason, data scientist and founder of Fast Forward Labs

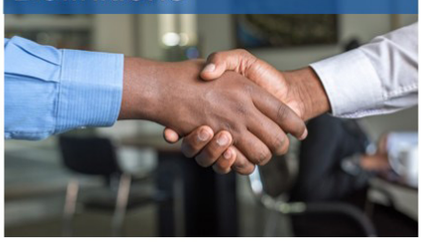
 44

Plan for Data Modeling

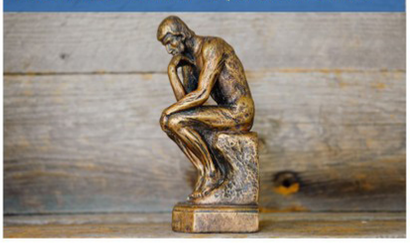
Diagram



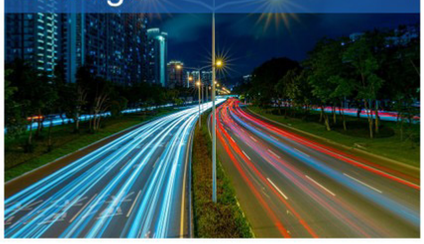
Definitions



Issues and Questions



Lineage



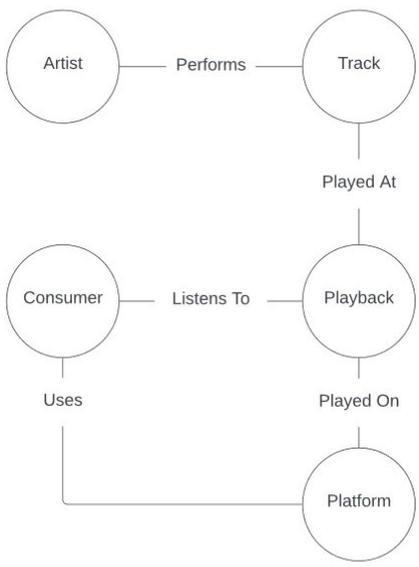
 45

Exercise: Conceptual Data Model

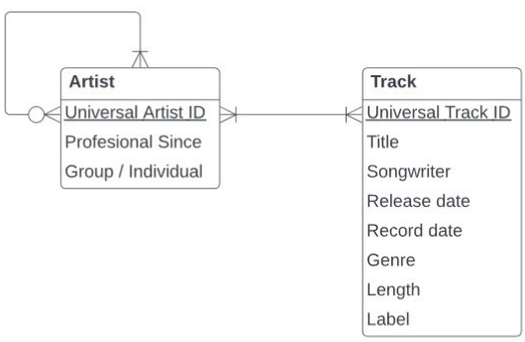
Artist	A person, or group of persons, who creates, performs or produces music as a profession.
Music Track	A recorded piece of music with melody, rhythm, and instruments.
Consumer	A person who engages with and experiences the auditory and emotional aspect of musical compositions, either by actively choosing to hear them or by passively encountering them. This person has an active *subscription* to a *music platform*.
Playback	The event in which a *Music Track* is played by a *Consumer* on a *Platform*.
Platform	A software service on which *Music Tracks* are hosted and that allows for the playback of *Music Tracks*.

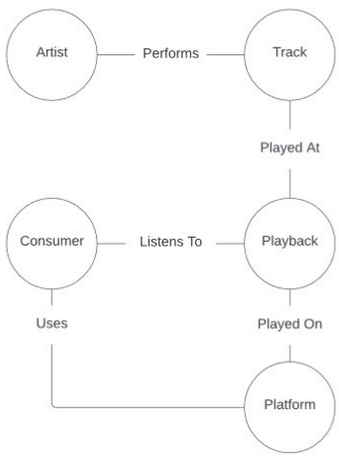
 46

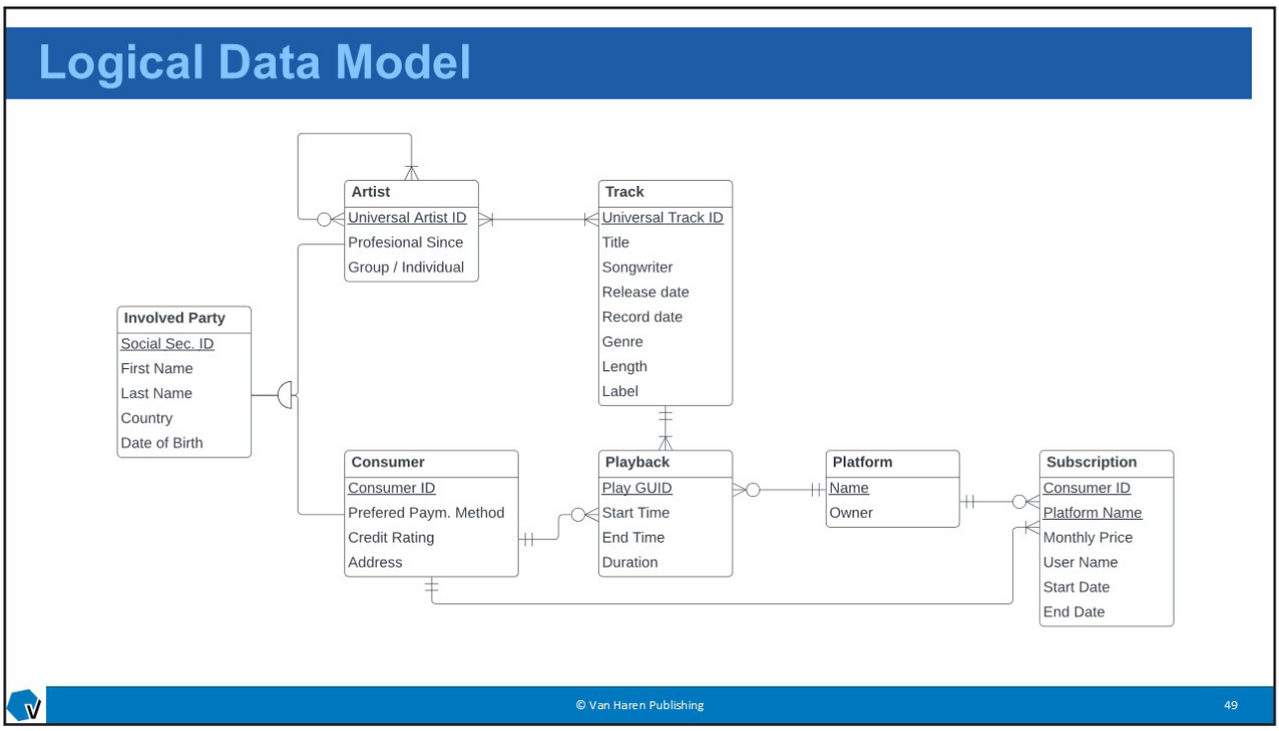
Conceptual Data Model

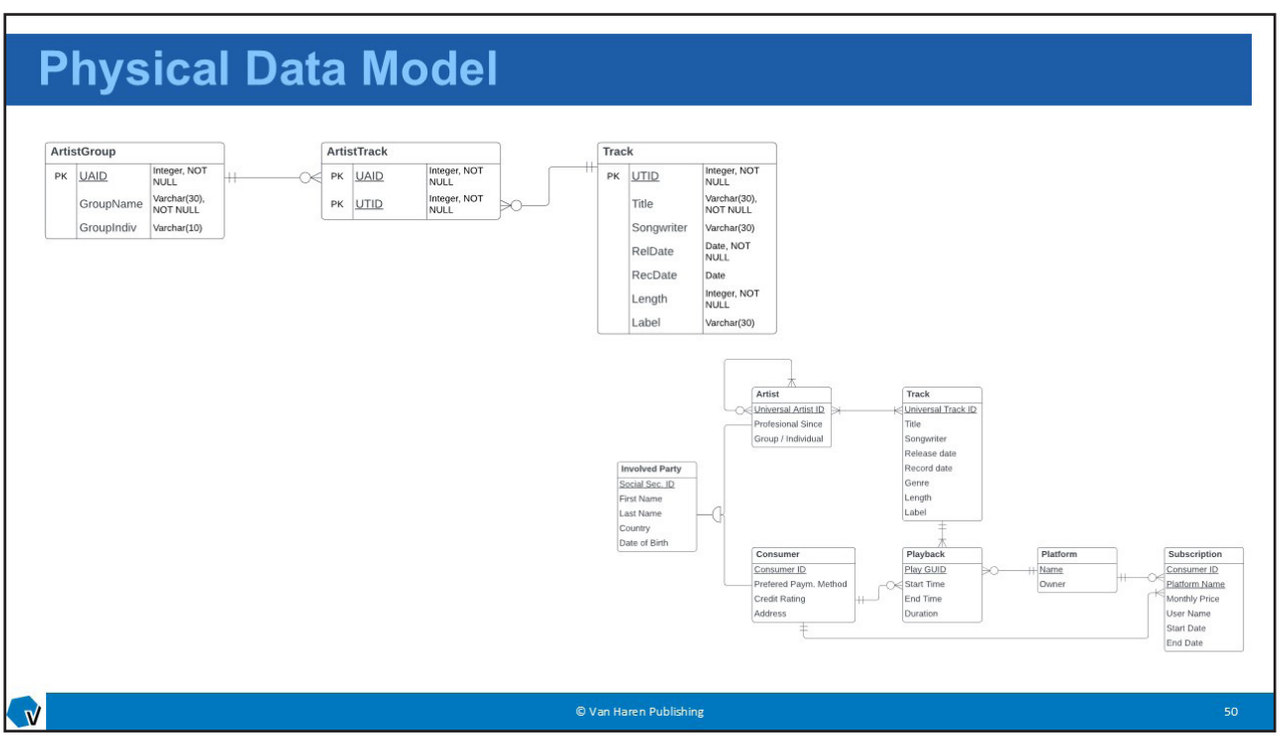


Logical Data Model

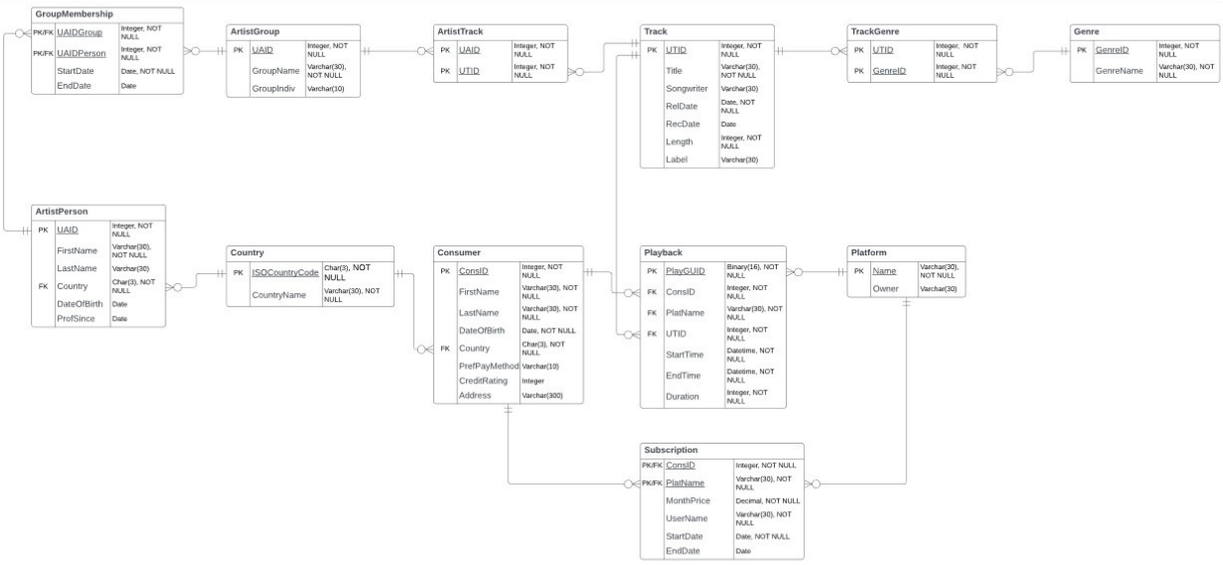





Logical Data Model
Artist
Universal Artist ID
Profesional Since
Group / Individual
Track
Universal Track ID
Title
Songwriter
Release date
Record date
Genre
Length
Label
Involved Party
Social Sec. ID
First Name
Last Name
Country
Date of Birth
Consumer
Consumer ID
Prefered Paym. Method
Credit Rating
Address
Playback
Play GUID
Start Time
End Time
Duration
Platform
Name
Owner
Subscription
Consumer ID
Platform Name
Monthly Price
User Name
Start Date
End Date
© Van Haren Publishing
49


Physical Data Model
ArtistGroup
PK UAID Integer, NOT NULL
GroupName Varchar(30), NOT NULL
GroupIndiv Varchar(10)
ArtistTrack
PK UAID Integer, NOT NULL
PK UTID Integer, NOT NULL
Track
PK UTID Integer, NOT NULL
Title Varchar(30), NOT NULL
Songwriter Varchar(30)
RelDate Date, NOT NULL
RecDate Date
Length Integer, NOT NULL
Label Varchar(30)
Artist
Universal Artist ID
Profesional Since
Group / Individual
Track
Universal Track ID
Title
Songwriter
Release date
Record date
Genre
Length
Label
Involved Party
Social Sec. ID
First Name
Last Name
Country
Date of Birth
Consumer
Consumer ID
Prefered Paym. Method
Credit Rating
Address
Playback
Play GUID
Start Time
End Time
Duration
Platform
Name
Owner
Subscription
Consumer ID
Platform Name
Monthly Price
User Name
Start Date
End Date
© Van Haren Publishing
50

Physical Data Model



Review Data Models

The Data Modeling Scorecard

#	Category	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 wel does the metadata match the data?	10			
	TOTAL SCORE	100			

S. Hoberman, Data Modeling Scorecard, www.stevenhoberman.com

Data Model Types

There's no such thing as Free Lunch

…

 53

Schema Types

Different goals, lead to different schema types.
goals: a) understand domain, b) design transaction system, c) design BI-system

Entity Relation Diagram

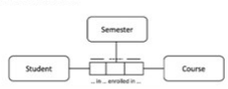
Primary for relation models.

Object Oriented (UML)

Primary for software systems.
The class diagram is widely used.

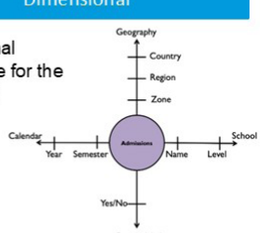
Fact based (ORM2, FCO/IM)

Has a strong connection with the business rules.



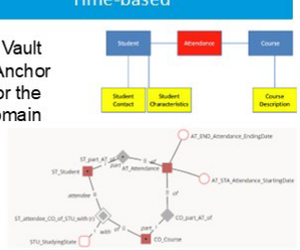
Dimensional

Dimensional models are for the BI-domain



Time-based

Data Vault and Anchor are for the BI-domain



NoSQL

No (standard) visual notation known.
NoSQL is about other forms of databases:
- key/value
- Wide column
- Graph
- Etc.

The figures for the dimensional model, data vault model, and anchor model are taken from DAMA DMBOK v2

 54