

## Introduction to Conceptual Database Design

Dr. John Artz

---

---

---

---

---

---

---

### Overview

- The Problem With Poorly Defined Categories
- Assuming Well Defined Categories, How to Represent Them
- Entity Relationship Modeling
- Top Down vs Bottom Up Information Modeling

---

---

---

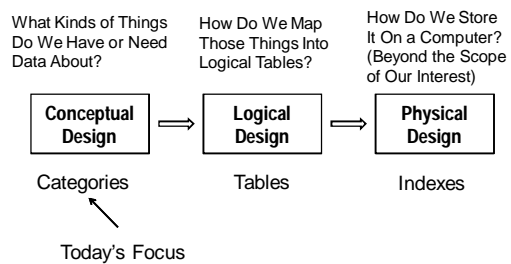
---

---

---

---

### Three Level Database Design



---

---

---

---

---

---

---

### Conceptual Database Design

- Conceptual Database Design is All About Creating Well Defined Categories of Information to be Represented in the Database
- Sometimes This is so Easy That it Seems Obvious
- Sometimes This is so Difficult That it Seems Impossible
- We Will Consider Both Ends of the Spectrum

---

---

---

---

---

---

---

---

### What's Wrong with Unemployment



What Does It Mean to be Unemployed?  
The Concept is so Poorly Defined That  
It is Impossible to do Anything About It

---

---

---

---

---

---

---

---

### Which of These are Unemployed?




---

---

---

---

---

---

---

---

### It All Depends...

- It All Depends on What You Mean By Unemployed
- For Any Concept X, The Question is What Do You Mean By X
- The Meaning of X Can Be:
  - Well Defined and Appropriate for Our Purposes
  - Well Defined and Not Appropriate for Our Purposes
  - Poorly Defined but Fixable
  - Poorly Defined but Not Fixable

---

---

---

---

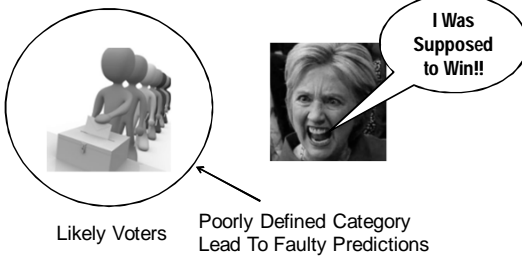
---

---

---

---

### What Went Wrong in 2016 Polling?




---

---

---

---

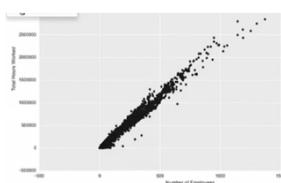
---

---

---

---

### A Good Bad Example



Here is a Scatterplot of The Number of Park Service Employees Vs Productive Hours, I Saw in a YouTube Video

Data Was Downloaded Off the Web and Plotted Using iPython

It Seems to Suggest a Linear Relationship Between Employees and Productivity  
More Employees = More Work Gets Done

---

---

---

---

---

---

---

---

**Why Is It Bad?**

- Lacking a Rigorous Conceptual Design, the Results are Meaningless at Best and Dead Wrong at Worst
- What Do We Mean By Employee?
- What Do We Mean By Productive Hours?
- If Either of These is Unclear, Then the Meaning of the Results is Unclear

---

---

---

---

---

---

---

**What Do We Mean by Employees?**

- Are Employees Full Time, Part Time or Volunteers
- Are Employees New or Experienced
- Are Employees Managed or Self Directed
- Is the Workforce Stable or Turning Over
- Is the Workforce Seasonal or Steady

---

---

---

---

---

---

---

**What Are Productive Hours?**

- The Employee Didn't Call in Sick
- The Employee Showed Up
- The Employee Appeared to be Online Working
- The Employee Was in Their Office
- The Employee Was Making Measurable Progress on Key Tasks Related to the Mission of the Organization

---

---

---

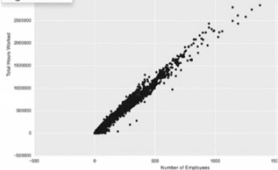
---

---

---


---

### Back to the Example



If Work is Measured in Hours, Then, Of Course, More People Means More Work Done

The Only Thing We Can Learn From This Graph Is That Definitions Matter



---

---

---

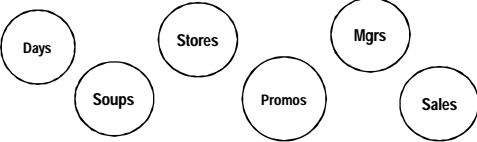
---

---

---

---

### Soup Sales Categories



We Have Some Categories From Soup Sales Which Do Not Appear to be Problematic, So We Will Use Them to Explore Information Modeling

---

---

---

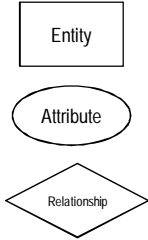
---

---

---

---

### Entity Relationship Modeling



---

---

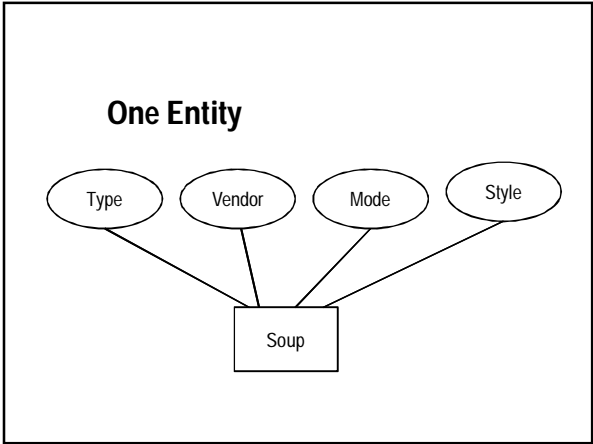
---

---

---

---

---



---

---

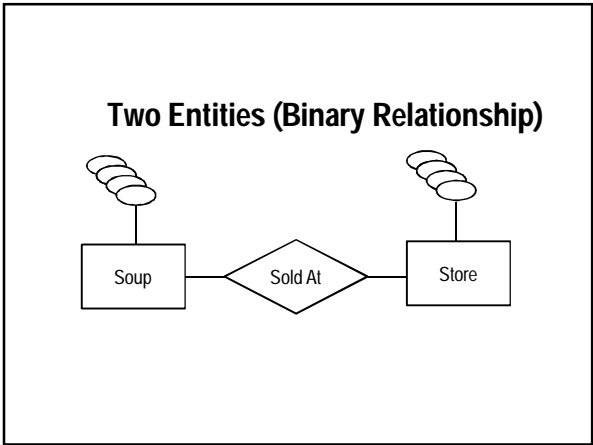
---

---

---

---

---



---

---

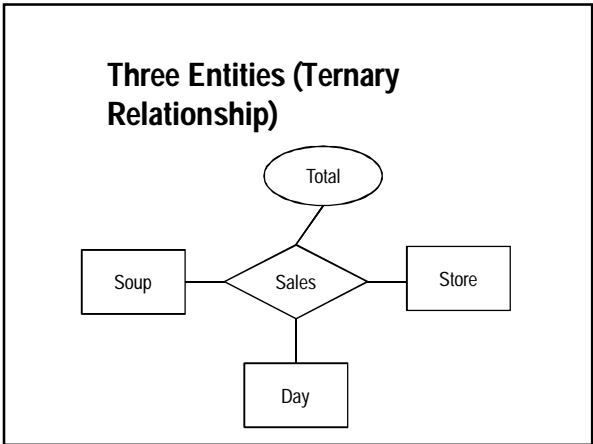
---

---

---

---

---



---

---

---

---

---

---

---

**Relationship Degree**

- The Number of Entities That Participate in a Relationship is Known as the Degree
  - Unary – One Entity
  - Binary – Two Entities
  - Ternary – Three Entities
- What About Four, Five, Six or More?

---

---

---

---

---

---

---

**Degree Debates**

- Some Modelers Advocate Only Binary Relationships to Maximize Simplicity
- Some Modelers Advocate Limiting Relationships to Ternary to Limit Complexity
- However, Most Analytical Applications Require More Than Three Entities in a Relationship So This Issue is Really Unresolvable

---

---

---

---

---

---

---

**Cardinality**

- Cardinality Refers to the Number of Occurrences of One Entity That Can Be Related to an Occurrence of the Other Entity
- Options for Cardinality Are:
  - 1 – 1 One to One
  - 1 – M One to Many
  - M – N Many to Many

---

---

---

---

---

---

---

**1 – 1 Cardinality**

- 1-1 Relationships Require That Only One Occurrence of an Entity Can Be Related to an Occurrence of the Other Entity in the Relationship
- For Example Each Store Has One Manager and Each Manager Manages Only One Store

---

---

---

---

---

---

---

**Representing 1 – 1 Cardinality**

---

---

---

---

---

---

---

**1 – M Cardinality**

- 1 -- M Relationships Allow Many Occurrence of the Second Entity in a Relationship Can Be Related to Each Occurrence of the First Entity, But Only One Occurrence of the First Entity Can Be Related to Each of the First
- A Day May Have Many Sales Amounts But Each Sales Amount Belongs to a Given Day

---

---

---

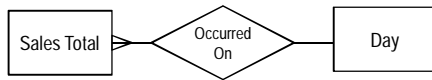
---

---

---

---



**Representing 1 – M Cardinality**

---

---

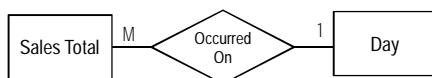
---

---

---

---

---

**Alternative 1 – M Representation**

---

---

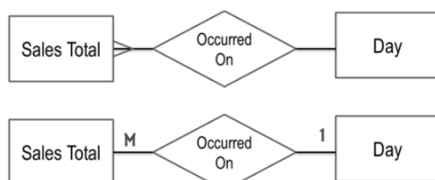
---

---

---

---

---

**Which is Right?**

Models Can Be Represented in a Variety of Ways  
The Best is the One That is Easiest to Understand  
And Most Useful in Facilitating Communication

---

---

---

---

---

---

---

### M – N Cardinality

- M -- N Relationships Allow Many Occurrence of the Second Entity in a Relationship To Be Related to Each Occurrence of the First Entity, And Many Occurrences of the First Entity Can Be Related to Each of the First
- A Soup May Sell at Many Stores and Each Store Can Carry Many Soups

---

---

---

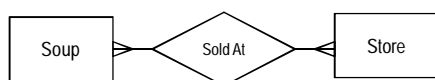
---

---

---

---

### Representing M – N Cardinality




---

---

---

---

---

---

---

### Cardinality, What If



How Do You Know Who Is Responsible for a Store's Productivity  
How Do You Know Which Managers Are Better

These Models Can Be Used to Refine the Business Rules  
And the Semantics of the Domain

---

---

---

---

---

---

---

**Optionality**

- Optionality Addresses Whether or Not Occurrences of the Relationship Are Required
- Required – The Relationship is Required
- Optional – The Relationship is Optional
- We Are Getting Into the Weeds Here and There is a Lot of Disagreement Over Whether or Not More Detail is Better

---

---

---

---

---

---

---

**Required Relationship**

- If a Relationship is Required Then Each Occurrence of One Entity is Required to be Associated With an Occurrence of the Other Entity
- For Example, Each Store Must Have a Manager and Each Manager Must Have a Store

---

---

---

---

---

---

---

**Representing Required Relationships**

---

---

---

---

---

---

---

### Alternate Required Relationship



Which One is Right?

---

---

---

---

---

---

---

---

### Optional Relationships

- If a Relationship is Optional Then an Occurrence of One Entity is Not Required to be Associated With an Occurrence of the Other Entity
- A Soup May Not Be Carried in Every Store and a Store May Not Carry Every Soup
- Note: Optionality Must Be Determined in Both Directions

---

---

---

---

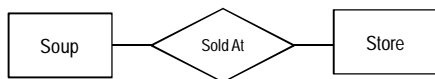
---

---

---

---

### Representing Optional Relationships




---

---

---

---

---

---

---

---

### Partial Optionality

- Since Optionality Must Be Determined in Both Directions, It is Possible that a Relationship is Required in One Direction But Not the Other
- A Store Must Have a Manager, But a Manager May Not Have a Store

---

---

---

---

---

---

---

---

### Partial Optionality Represented




---

---

---

---

---

---

---

---

### Optionality What If



If a Manager Resigns, Does the Store Close?  
 If a Store Closes is the Manager Fired?  
 Models Can Be Used to Refine or Clarify  
 Business Rules

---

---

---

---

---

---

---

---

**Attributes**

- An Attribute Is an Item of Information That Describes an Entity
- It Should Be a Single Fact and Should Be Single Values
- Types: Naming, Referential, or Descriptive
- Optionality: Mandatory or Optional
- Persistence: Essential or Transitory

---

---

---

---

---

---

---

**Attribute Types**

- A Naming Attribute Helps Identify Occurrences of the Entity
- A Referential Attribute is a Naming Attribute of Another Entity and thus Indicates a Relationship Between Entities
- A Descriptive Attribute is a Fact About the Entity

---

---

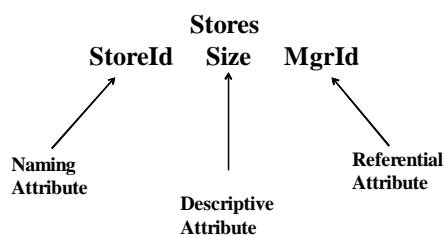
---

---

---

---

---

**Attribute Type Example**

---

---

---

---

---

---

---

**Attribute Optionality**

- An Attribute Can Be Mandatory or Optional
- Warning: Optional Attributes May Indicate Internal Consistency Problems
- Naming Attributes Are Always Mandatory
- Descriptive Attributes May be Optional
- Referential Attributes May be Optional

---

---

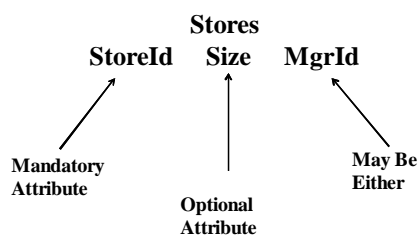
---

---

---

---

---

**Optionality Example**

---

---

---

---

---

---

---

**Attribute Persistence**

- An Attribute Can Be Essential (cannot be changed) or Transitory
- Naming Attributes are usually Essential
- Descriptive and Referential May Be Transitory

---

---

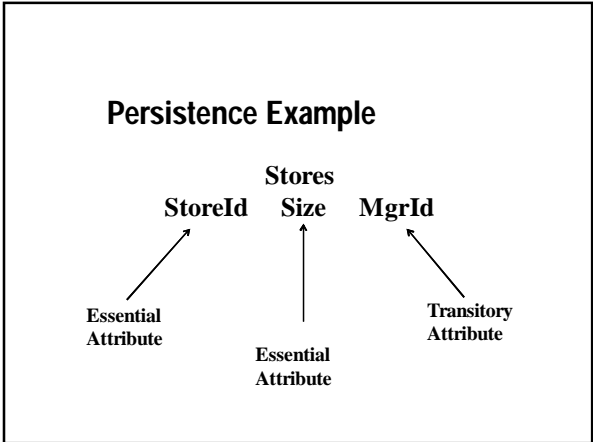
---

---

---

---

---



---

---

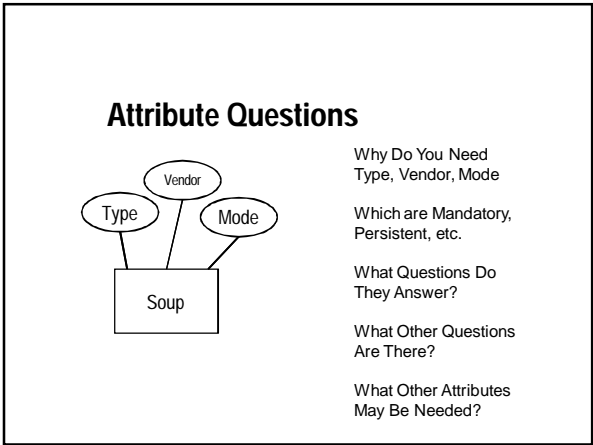
---

---

---

---

---



---

---

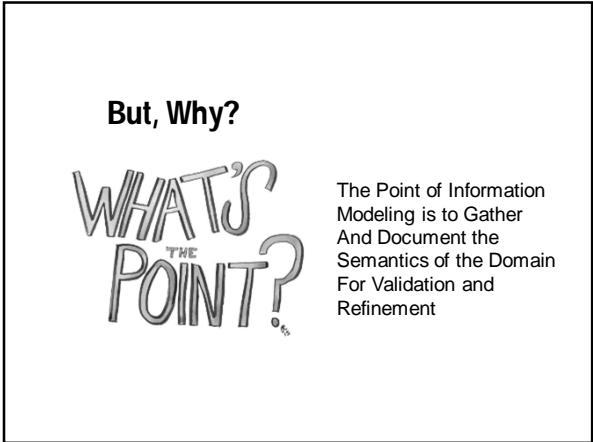
---

---

---

---

---



---

---

---

---

---

---

---



**The Myth of Canonical Models**

- Any Given Domain May Be Modeled in Many Different Ways
- E/R Models May Be Wrong But Cannot Be the Only Right Way to Model a Domain
- It Is Best to Think of E/R Models Not as Right or Wrong But as More or Less Useful

---

---

---

---

---

---

---

**Useful for What?**

- Information Models Have Two Initial Purposes:
- First, They Serve as a Means of Communication Between Modelers and Domain Experts as They Attempt to Get the Domain Modeled as Correctly as Possible
- Second, Once the Model is Correct, It Can Be Used for Documentation

---

---

---

---

---

---

---

**The Debate Over Level of Detail**

- If the Model Is Incorrect, More Detail Creates an Illusion that the Model is Correct Which Can Be Misleading and Make It Difficult to Change
- If the Model Is Correct, More Detail Provides Valuable Documentation and Useful Input For Computer Generated Database Schemas
- So Get It Right First, and Then Add Detail

---

---

---

---

---

---

---

### A Moving Target

- Domain Semantics Are Often Messy and the Modeling Process Highlights This Messiness Causing the Domain Semantics to Change, Which, In Turn, Causes the Model to Change
- Conceptual Database Design is Often an Evolving Process of Clarifying and Improving Domain Semantics Through Guided Consensus

---

---

---

---

---

---

---

### Guided Consensus?



---

---

---

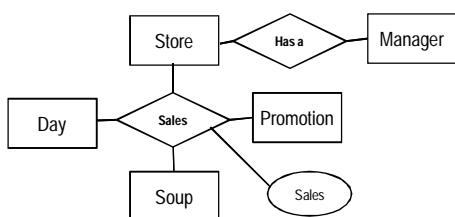
---

---

---

---

### A Soup Sales Information Model



---

---

---

---

---

---

---

**Validating With Business Rules**

- Does a Store Have to Have a Manager
- Is Every Soup Offered at Every Store
- Are Some Promotions Limited to Specific Soups
- Are Some Promotions Limited to Specific Days
- Are There Days When No Promotions Are Used
- Are There Stores That Don't Use Specific Promotions

---

---

---

---

---

---

---

**Model Correctness**

- People Obsess Needlessly Over Whether or Not a Model is Correct
- Much Ink Has Been Spilled Over Crow's Feet versus Numbers in Many to Many Relationships
- This Misses the Point as the Diagram is a Model of Something That Doesn't Exist Until the Model and Hence the Diagram are Completed

---

---

---

---

---

---

---

**Model Usefulness**

- The Real Purpose of an Information Model is to Server as a Vehicle for Communication Between Developers and Users
- If You Obsess Over the Correctness of the Model or the Diagramming Technique, It May Interfere With the Model's Usefulness as a Vehicle of Communication
- So, Keep It Simple

---

---

---

---

---

---

---

**Digging a Little Deeper**

- A Relational Database Models Categories of Information
- We Have Been a Little Sloppy in Our Terminology
- We Have Been Using the Term “Entity” When We Should Have Been Using the Term “Entity Class”
- Does This Really Matter?
- Yes!! It Matters a Lot

---

---

---

---

---

---

---

**Entities vs. Entity Classes**

- An Entity Is a Thing of Interest Within the UoD About Which Information May Be Collected, Manipulated, and Stored
- An Entity Type (or Class) is a Grouping of Entities
- An Entity is an Occurrence of an Entity Type
- Where Do These Groupings Come From
- This Will Become a HUGE Philosophical Problem

---

---

---

---

---

---

---

**A Word About Vocabulary**

- Entities Are Individual Occurrences of Things in the Domain
- Entity Types Are the Groups into Which We Organize or Categorize Those Occurrences
- Entity Types, Entity Classes and Categories All Refer to the Groupings
- How Do You Figure Out the Groupings

---

---

---

---

---

---

---

**Are Students Real?**

- To Get a Glimpse at the Philosophical Problem, We Can Ask – Are Students Real?
- “Student” is a Category
- Is that Category Real or is It a Construct That We Made Up
- If it is a Construct Are Other, Perhaps Better, Constructs Possible?

---

---

---

---

---

---

---

**Where Do the Categories Come From?**

- This is a Vexing Question That We Have Been Ignoring and Will Continue to Ignore
- However, the Way in Which One Answers it Will Determine Whether or Not They Can Design Databases Correctly

---

---

---

---

---

---

---

**Two Approaches**

- Given a Well Defined Problem and Well Defined Objectives, We Can Approach Conceptual Database Design From Two Perspectives
  - Top Down Conceptual Design Views the Application Domain as a Universe of Discourse
  - Bottom Up Conceptual Design Looks at Available Data

---

---

---

---


---

---

---

### Top Down Conceptual Design

Top Down Designers  
Examine the Application  
Domain and Identify  
Entity Types Based on  
Word Usage



---

---

---

---


---

---

---

### Bottom Up Conceptual Design

Bottom Up Designers  
Examine the Existing Data  
and Identify Entity Types  
Based on What They Find  
In the Data



---

---

---

---

---

---

---

### How Many Courses Are Offered?

Course	Section	Description	Day	Time
ISTM6202	10	Database	F	4-6
ISTM6202	11	Database	M	6-8
ISTM6202	12	Database	R	6-8
ISTM6203	10	Telecom	W	6-8
ISTM6203	11	Telecom	M	8-10
ISTM6204	10	Proj. Mgmt	R	8-10
ISTM6207	10	IRM	T	6-8

The Answer Could  
Be 4 or 7 Depending  
On What You Mean  
By "Course"

---

---

---

---

---

---

---

**An Important Perspective**

- Information Modelers Often Mistakenly Believe That Their Goal is to Model the Domain as Accurately as Possible
- This is Naive and Wrong
- The Goal of Information Modeling is to Refine the Semantics of the Domain in Order to Answer Important Questions

---

---

---

---

---

---

---

**The Goal of Information Models**

- Information Models Provide Information Modelers With a Visual Vehicle For Organizing Their Thoughts
- They Also Provide a Communication Vehicle to be Used in the Process of Refining the Semantics of the Domain

---

---

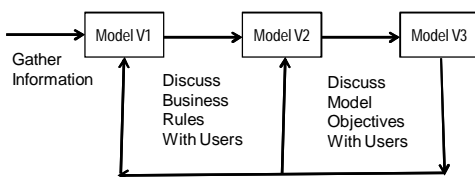
---

---

---

---

---

**The Modeling Process**

---

---

---

---

---

---

---

**Recap**

- The Problem With Poorly Defined Categories
- Assuming Well Defined Categories, How to Represent Them
- Entity Relationship Modeling
- Top Down vs Bottom Up Information Modeling

---

---

---

---

---

---

---