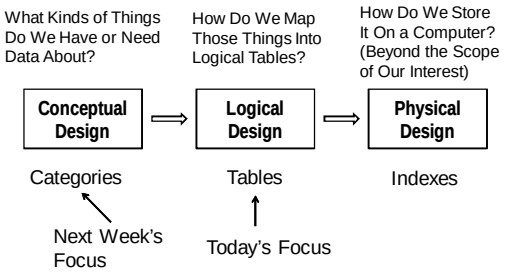
Logical Database Design

Dr. John Artz

Overview

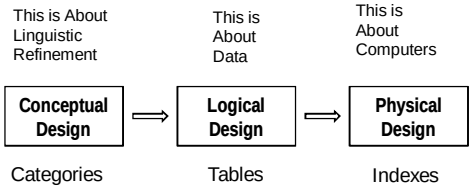
- Relational Table Design is Based on Functional Dependency
- 1st NF: All Data Items Must Be Atomic Facts
- 2nd NF: Full Functional Dependencies
- 3rd NF: No Transitive Dependencies
- Normalization Leads to Smaller Tables
- Joins Allow Us to Put Them Back Together Again

Three Level Database Design

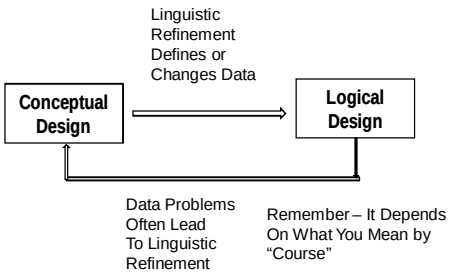


© 2018 Dr. John M. Artz

Three Level Distinctions



More Realistically



Recall: What Do You Mean By Course?

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

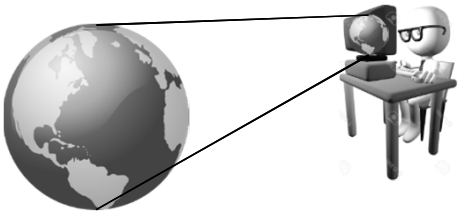
How Many Courses Are Offered?

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

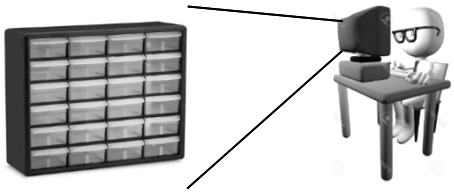
Is There Some Way To Catch This?

© 2018 Dr. John M. Artz

First and Foremost: A Database is a Model



Otherwise, It Is Just Storage



Storage Isn't a Bad Thing

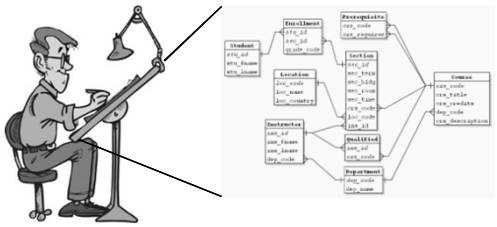
Web Applications
Documents
Images
Videos
Big Data

All Need Storage

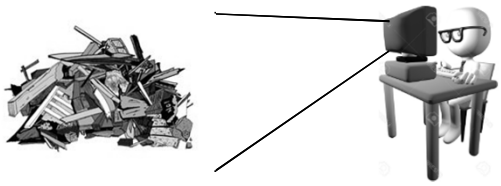
But, If Your Goal is Information, Storage Alone Isn't Enough

© 2018 Dr. John M. Artz

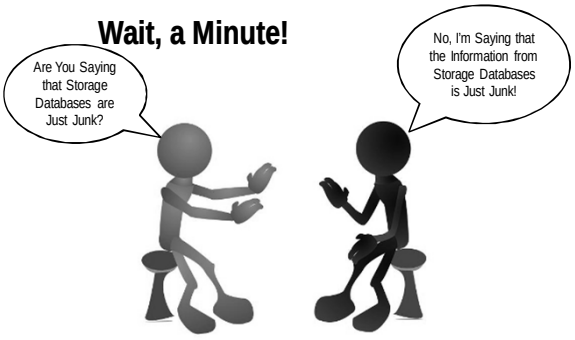
Second, It Must Be Designed Correctly



Otherwise, It Is Just Junk

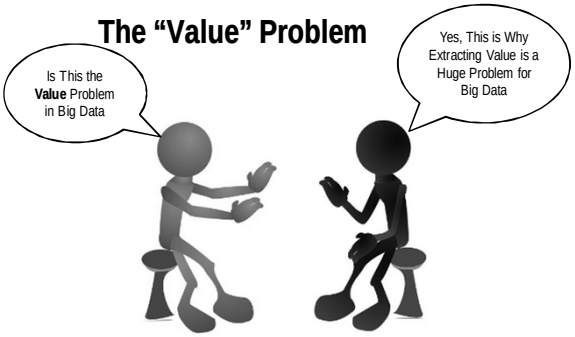


Wait, a Minute!

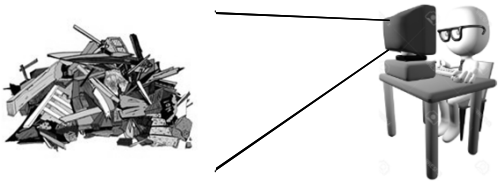


© 2018 Dr. John M. Artz

The "Value" Problem



You Can Extract Value From Junk



But, There Are Better Ways

This is *Designed* to Provide What You Need

Database Design Prevents Junk

This, too, Should be *Designed*
To Provide What You Need

Storage vs. Information Databases

Storage Databases Provide

Storage

Designed Databases Provide

Information

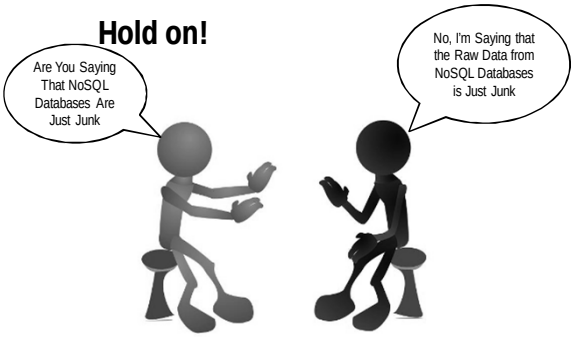
Storage Databases

These Are Referred to as NoSQL Databases

Storage Database Applications

Web Applications

Hold on!



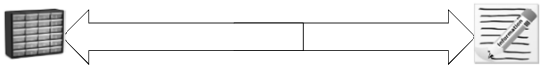
An Analogy for Big Data

Gold Mining
Produces
Gold Dust

It Takes Design And
a Production Process
to Produce Jewelry

The Challenge is Turning Gold Dust into Jewelry
The Same is True With Big Data

Actually, It's a Spectrum



Utility
Junk Information
Lousy ROI

Investment
Good Information
Good ROI

This Raises a Question

How Do You "*Design*" a Database?

The Answer is Normalization

- There Are Rules to Follow Called Normalization
- However, the Rules are Not Enough Without a Deeper Understanding
- So, We Will Sneak Up on It By Providing an Understanding of Foundation Concepts:
  - What is Data?
  - What is an Entity?

© 2018 Dr. John M. Artz

The World is Made Up of Things


People


Activities


Locations


Products


Events

We Call These Things Entities

- An entity is something that exists as itself, as a subject or as an object, actually or potentially, concretely or abstractly, physically or not. It need not be of material existence. In particular, abstractions and legal fictions are usually regarded as entities. – Wikipedia
- OK, Let's Parse This

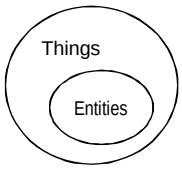
Parsing Entity Definition

- An Entity is a "Thing"
- It Exists As Itself (Day vs. Night, War vs. Peace)
- Subject (Acted Upon) or Object (Actor)
- Actually (Now) or Potentially (In the Future)
- Concrete (e.g. Person) or Abstract (e.g. Teenager)
- Physical (e.g. Product) or Not (e.g. Service)

© 2018 Dr. John M. Artz

Thing vs Entity

- Every Entity is a "Thing"
- But, Not Every "Thing" is an Entity



Examples of Entities

Day Soup Store Mgr Promo Sale

These Are All Entities, In Case You Get Confused As We Tease Out Some Subtleties

What Are Some Non Entities

- Air, Water, Outer Space, The Universe
- Love, Justice, Kindness, Well Being
- Evolution, Philosophy, Literature
- Ambition, Motivation, Revenge
- Good, Bad, Right, Wrong, Eternity
- Gravity, Dark Matter, Time, Space

© 2018 Dr. John M. Artz

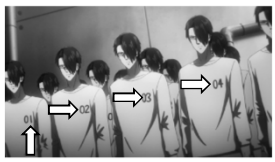
Entities vs. Non Entities

- Some Non Entities Can Become Entities By Redefining Them More Rigorously
- Some Entities Can Become Non Entities Due to Fuzzy Definitions or Sloppy Conversational Usage
- Is Your "Friend" an Entity
- I Was Talking to My "Friend" the Other Day….

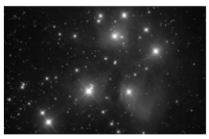
Non Entity – Not Unique

These Are Non Entities Because They Are Not Uniquely Identifiable

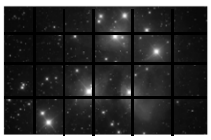
These Are Entities. They are Uniquely Identifiable



Non Entity – Not Distinguishable

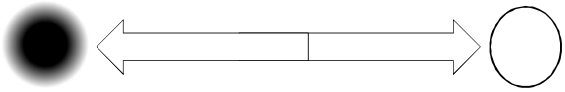


The Night Sky is Not an Entity Although It Is a Thing



Cells Could Be Entities Or the Stars Could Be As Well

It's a Spectrum Here Too



Poorly Defined Entity Leads to Poor Quality Information

Well Defined Entity Leads to Solid Information

Entities Have Facts About Them

Name
Birth Date
Height
Weight
Hair Color
Eye Colr
Cell Phone
Email Addr

What Are Facts?

- Facts Are Verifiable Pieces of Data About an Entity that Can Be Used in Identification, Elaboration, or Summary Regarding the Entity

© 2018 Dr. John M. Artz

Fact vs. Opinions

Facts	Opinions
Name	Cute
Birth Date	Smart
Height	Smells Nice
Weight	Friendly
Hair Color	Too Loud
☑	☒

What is Data?

- Data Is a Set of Values of Qualitative or Quantitative Variables
- Data is Information or Knowledge Represented or Coded in Some Form Suitable for Better Usage or Processing
- A Data Item is a Single Atomic Piece of Data
- Data Items Are Facts About Entities

What is Not Data?

- Images
- Documents
- Spreadsheets
- Newspaper Stories
- Gossip
- Online Reviews

© 2018 Dr. John M. Artz

What Is a Database?

- A Database Is An Organized Collection Of Data
  - Wikipedia
- Notes
  - Organization Implies Purpose
  - Collection Implies Containment
  - Data Implies Discreet Pieces of Information

What Is a Relational Database?

- A relational database is a digital database whose organization is based on the relational model of data, as proposed by E.F. Codd in 1970. This model organizes data into one or more tables (or "relations") of rows and columns, with a unique key for each row.
  - Wikipedia

Relational vs. Not Relational

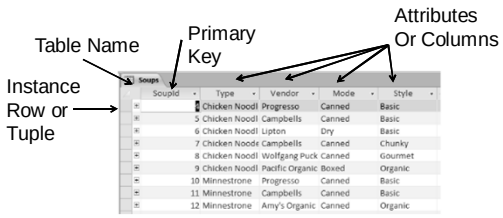
Relational

Not Relational

```
<Soup>
<Id>1</Id>
<Type>Chicken Noodle</Type>
<Vendor>Progresso</Vendor>
        <Mode>Canned</Mode>
<Vendor>Campbells</Vendor>
        <Mode>Canned</Mode>
<Vendor>Lipton</Vendor>
        <Mode>Dried</Mode>
</Soup>
```

© 2018 Dr. John M. Artz

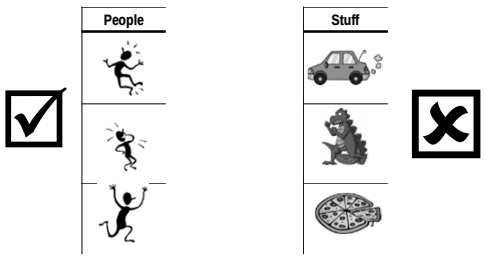
A Relational Table



Relational Rules

- One Kind of Entity Per Table (Entity Integrity)
- Entities Must Be Uniquely Identifiable
- Facts Must Be Atomic

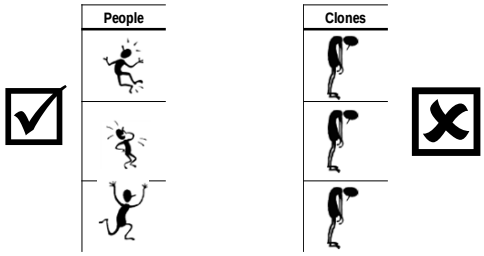
One "Kind" of Entity per Table



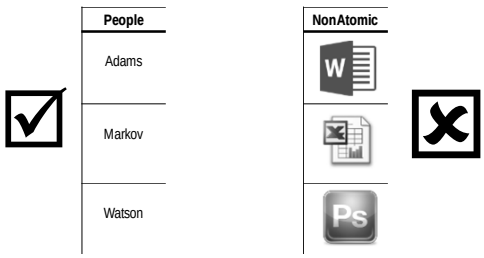
Entity Integrity

- In the Relational Model We Refer to this as Entity Integrity
- This is the Single Most Important Principle in Relational Database Design
- Notice the Slick Linguistic Slight of Hand With "One Kind of Entity"

Uniquely Identifiable Entities



Atomic Values



A Few Intuitive Observations

- Table Name Must Describe a Homogenous Group of Entities
- Rows Must Be Instances of Those Entities
- Instances Must Have Unique Identifiers
- All Instances Must Have the Same Attributes
- Attributes Must Be Single Valued Facts

Functional Dependency

- Relational Database Design is Based on a Relationship Between Data Items (Facts) Called Functional Dependency
- Intuitively, Y is Functionally Dependent on X, If a Given Value of X Determines a Unique Value for Y
- Notationally:
  - X → Y (read X determines Y)

Understanding Functional Dependency

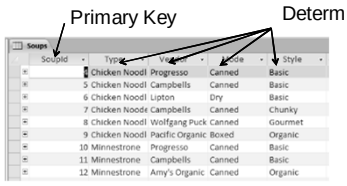
- If I Give You a Name, Can You Give Me a Unique Cell Phone Number?
- If I Give You a Cell Phone Number, Can You Give Me a Unique Name?
- If I Give You an Email Address Can You Give Me a Unique Name

© 2018 Dr. John M. Artz

Functional Dependency Exercise

- What Does a GWID Determine?
- What Determines a GWID?
- What Does a City Name Determine?
- What Determines a City Name?

Intuitive Table Design



Intuitively, A Relational Table has a Determinant Which We Call the Primary Key, Followed By the Facts it Determines

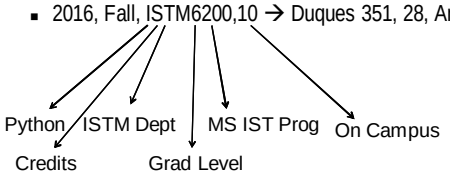
MultiAttribute Dependency

- If X → Y, X may be a Set of Data Items
- For Example, This Class is Determined Uniquely By: Year, Semester, Course, Section
- 2016, Fall, ISTM6200,10 → Duques 351, 28, Artz

© 2018 Dr. John M. Artz

Subset Determinants

- 2016, Fall, ISTM6200,10 → Duques 351, 28, Artz



Subset Determinants Should
Not Be in the Same Table

Full Functional Dependency

- An Attribute Y may be Determined by a Set of Attributes A,B,C (ABC → Y)
- If X is a Set of Attributes Such That X → Y and there is no Subset Z of X so that Z → Y Then Y is Fully Functionally Dependent on X

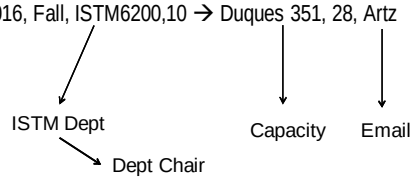
Full FD Example

- 2016, Fall, ISTM6200,10 → Python, 28
- A Functional Dependency Exists:
  - 2016, Fall, ISTM6200,10 → Python
- But, It is Not a Full Functional Dependency As:
  - ISTM6200 → Python
- Course Name Is Determined By a Subset

© 2018 Dr. John M. Artz

Transitive Determinants

- 2016, Fall, ISTM6200,10 → Duques 351, 28, Artz



Subset Determinants Should
Not Be in the Same Table

Transitive Dependency

- Let X, Y, and Z be Attributes such that

  X → Y and

  Y → Z
- Then a Transitive Dependency Exists Between X and Z such that

  X → Z

Normalization

- Normalization Helps Us Achieve Atomicity
- 1st Addresses Attribute Atomicity
- 2nd and 3rd Normal Forms Addresses Entity Kind Atomicity

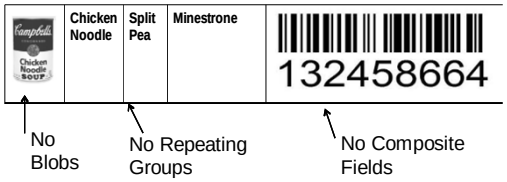
© 2018 Dr. John M. Artz

Higher Normal Forms

- 4th Normal Form Addresses Multivalued Dependencies
- 5th Normal Form Addresses Lossless Joins
- Boyce-Codd Normal Form Defines Normalization in Terms of Alternative Keys
- These Are Beyond the Scope of this Introduction

1st Normal Form

All Fields Must Be Atomic



Non Atomic Key Problem

Date	Leap Year	Day of Week	Season	Weekend
Apr 06 2016	Y	Weds	Spring	N
May 8 2017	N	Mon	Spring	N
July 15 2017	N	Sat	Summer	Y

© 2018 Dr. John M. Artz

Non Atomic Value Problem

CRN	Description	Course
13311	Database	ISTM6202
15166	Python	ISTM6290

What Department is Responsible For the Python Course?
How Many Courses Does That Department Offer?

Structured Numbers

Work Order Charges				
Account	Date	Charger	Amount	Code
195-2436-023	02/06/2015	Jones, Fred	$394	2A7

Account is a Structured Number Consisting Of a Project Id, Task Id and Work Order Number

Code Consists of a Category, Level of Importance and a Labor Level

Why Is This a Problem?

Repeating Groups

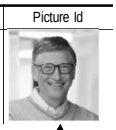
Account Charge												
Account	1	2	3	4	5	6	7	8	9	10	11	12
195-2436-023	23	17	19	26	26	24	33	22	19	17	22	25

This Table Maintains the Account Charges for the Past Twelve Months

Why Is This a Problem?

© 2018 Dr. John M. Artz

Non Text Attributes

Employees		
Employee Number	Name	Picture Id
000000001	Gates, Bill	

This Table Has a Non Text Attribute

Why Is This a Problem?

Violations of First Normal Form

- All Violations of 1NF Must Be Removed in Order for a Table to Be in 1NF
- And a Table Must Be in1NF Before Proceeding to Higher Normal Forms

2nd Normal Form

- All Non Key Attributes Must Be Dependent on the Whole Key

Soup Sales						
Promo Id	Store Id	Soup Id	Store Id	Day Id	Soup Type	Sales
8	5	3	4	17	Split Pea	432.25

Soup Type is a Fact About The Soup Not the Sale

2nd Normal Form Solution

Soup Sales					
Promo Id	Store Id	Soup Id	Store Id	Day Id	Sales
8	5	3	4	17	432.25

Soups	
Soup Id	Soup Type
3	Split Pea

We Decompose the Larger Table into Smaller Atomic Tables

3rd Normal Form

- No Transitive Dependencies

Store

StoreId	Location	elevati	size	mgrname	grade	years
1	Rockville	Basic	Medium	Smith, J.	Junior	5
2	Potomac	Upscale	Medium	Williams, K.	Midlevel	8
3	Gaithersburg	Basic	Large	Costenably	Junior	6
4	Bethesda	Upscale	Medium	Albertson, M.	Senior	15
5	Frederick	Basic	Large	Gershman, L.	Midlevel	9

Manager Grade and Years Are Facts About The Manager, Not About the Store

3rd Normal Form Solution

Stores

StoreId	Location	Elevatio	Size	MgrId
1	Rockville	Basic	Medium	1
2	Potomac	Upscale	Medium	2
3	Gaithersburg	Basic	Large	3
4	Bethesda	Upscale	Medium	4
5	Frederick	Basic	Large	5

Facts About the Store

Facts About the Manager

We Decompose the Larger Table into Smaller Atomic Tables

Managers

MgrId	MgrName	Grade	Years
1	Smith, J.	Junior	5
2	Williams, K.	Midlevel	8
3	Costenably	Junior	6
4	Albertson, M.	Senior	15
5	Gershman, L.	Midlevel	9

© 2018 Dr. John M. Artz

Normalization

- The Process of Logically Designing Tables in a Relational Database is Called Normalization
- We Will Get to That in Week 12
- For Now, We Can Simply Say That Normalization Tends to Break Larger Tables into Smaller Ones

An Un-normalized Table

Day of Sale Info

Store Info

Promo Info

Mgr Info

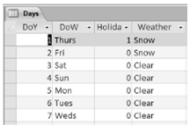
Soup Info

Raw Data Usually Contains Information About a Variety of Things

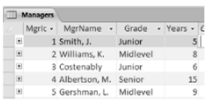
Normalized Data

Normalized Data Usually Contain a Single Table For Each Thing of Interest

StoreID	Location	Size	Elevation	MgrID
1	Rockville	Basic	Medium	1
2	Potomac	Upscale	Medium	2
3	Gaithersburg	Basic	Large	3
4	Bethesda	Upscale	Medium	4
5	Frederick	Basic	Large	5



DoY	DoW	Holida	Weather
1	Thurs		1 Snow
2	Fri		0 Snow
3	Sat		0 Clear
4	Sun		0 Clear
5	Mon		0 Clear
6	Tues		0 Clear
7	Weds		0 Clear



Mgric	MgrName	Grade	Years
1	Smith, J.	Junior	5
2	Williams, K.	Midlevel	8
3	Costenably	Junior	6
4	Albertson, M.	Senior	15
5	Gershman, L.	Midlevel	9

This Creates Unit of Decomposition That Can Be Reassembled to Provide a Wide Variety of Information

© 2018 Dr. John M. Artz

The Problem With Normalizing

Stores

StoreId	Location	Size	Elevation	MgrID
1	Rockville	Basic	Medium	1
2	Potomac	Upscale	Medium	2
3	Gaithersburg	Basic	Large	3
4	Bethesda	Upscale	Medium	4
5	Frederick	Basic	Large	5

Managers

Mgrlc	MgrName	Grade	Years
1	Smith, J.	Junior	5
2	Williams, K.	Midlevel	8
3	Costenably	Junior	6
4	Albertson, M.	Senior	15
5	Gershman, L.	Midlevel	9

What If We Wanted Information About a Store Including the Manager's Name?

Sales

TrxId	DoY	StoreId	SoupId	PromoId	Sales
1	1	1	1	1	405.18
2	1	1	2	1	453.3
3	1	1	3	1	868.81
4	1	1	4	1	425.61
5	1	1	5	1	267.48
6	1	1	6	1	719.81

Or Information About a Sale With Attributes We Recognize?

Joining Tables

- Joining Tables Allows Us To Denormalize Tables on the Fly By Matching Foreign Keys With The Corresponding Primary Keys In Other Tables
- In Order To Do This We Must Specify The Tables Involved and the Columns With Matching Values

Without Joins

```
Select SoupId, Sales from Sales
Where Sales > 1999.50;
```

We Get →

Soup Id	Sales
10	1999.53
11	1999.73
15	1999.96

But, Which Soups Are They?

The Soup Name is In the Soups Table

© 2018 Dr. John M. Artz

With a Join

```
Select Type, Sales
from Sales, Soups
Where Soups.SoupId = Sales.SoupId
And Sales > 1999.50;
```

←This Is The Join Criteria

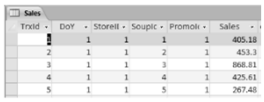
Now We Get →

Type	Sales
Split Pea	1999.53
Clam Chowder	1999.73
Minnestrone	1999.96

Multiple Table Joins

If We Want to Turn All of These Key Values Into Recognizable Values, We Will Have to Join Multiple Tables

⇩ ⇩ ⇩ ⇩



Multiple Table Example

```
Select DoW, Medium, Location, MgrName, Type, Sales
from Days, Promotions, Stores, Managers, Soups, Sales
Where Days.DoY = Sales.DoY
And Promotions.PromoId = Sales.PromoId
And Stores.StoreId = Sales.StoreId
And Managers.MgrId = Stores.MgrId
And Soups.SoupId = Sales.SoupId
And Sales > 1999.50;
```

© 2018 Dr. John M. Artz

Ta Da !!

DoW	Medium	Location	MgrName	Type	Sales
Fri	Newsppr	Gaithersburg	Costenably	Split Pea	1999.53
Fri	Radio	Frederick	Gershman, L.	Clam Chowder	1999.73
Tues	Instore	Potomac	Williams, K.	Minnestrone	1999.96

Renaming Tables

```
Select DoW, Medium, Location, MgrName, Type, Sales
from Days D, Promotions P, Stores St, Managers M,
Soups So, Sales Sa
Where D.DoY = Sa.DoY
And P.PromoId = Sa.PromoId
And St.StoreId = Sa.StoreId
And M.MgrId = St.MgrId
And So.SoupId = Sa.SoupId
And Sales > 1999.50;
```

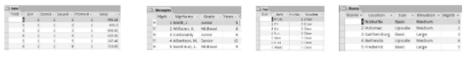
Multi-Table Selection Criteria

```
Select DoW, Type, Sales
from Days, Promotions, Soups, Sales
Where Days.DoY = Sales.DoY
And Promotions.PromoId = Sales.PromoId
And Soups.SoupId = Sales.SoupId
And Soups.Type = “Chicken Noodle”
And Days.Weather = “Clear”
And Promotions.Medium = “Radio”
And Sales > 1400;
```

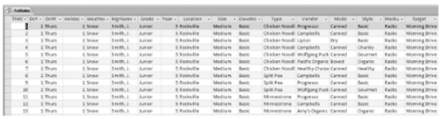
© 2018 Dr. John M. Artz

Joining = Denormalizing

Join



=



Where Do the Tables Come From

- We Have Been Assuming That We Have Tables to Normalize
  - What If We Don't Have Tables
  - Where Do We Start
- Answer: We Define Categories and Map Them Into Tables

Summary

- Relational Table Design is Based on Functional Dependency
- 1st NF: All Data Items Must Be Atomic Facts
- 2nd NF: Full Functional Dependencies
- 3rd NF: No Transitive Dependencies
- Normalization Leads to Smaller Tables
- Joins Allow Us to Put Them Back Together Again

© 2018 Dr. John M. Artz