Class Overview

- For the past 40 years, Relational Databases have formed the foundation of modern automated Data Processing systems

- This class provides students with an overview of Relational Database concepts and principles

Topics

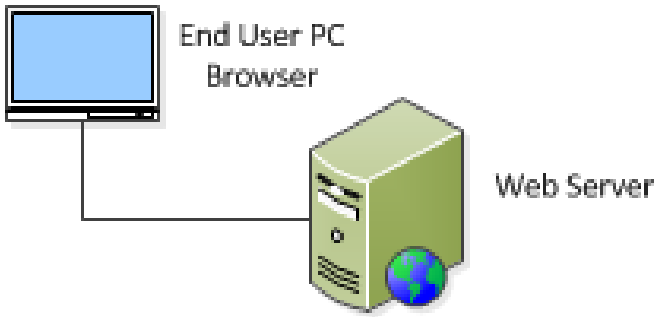
- Basic Database Systems Architecture
- DBMS Software
- The relational model, Relational Terms and Concepts:
  - Entity
  - Relation
  - Row, Column
  - Keys
  - Referential Integrity
  - Nulls
- Designing a Relational Database:
  - Normalization
  - Data Modeling basics
    - Entity, Attribute, Relationship
    - Diagramming Data Models

Applications, the DBMS, and SQL

- Applications are the computer programs that users work with.
  - Example: a Payroll system, a Registrar's system
- The Database Management System (DBMS) is very complex, powerful software that allows us to create databases, load data into databases, retrieve data from databases.
  - Example: MySQL, Oracle, Microsoft SQL Server
- Structured Query Language (SQL) is an internationally recognized standard database language that is used by all commercial DBMSs.

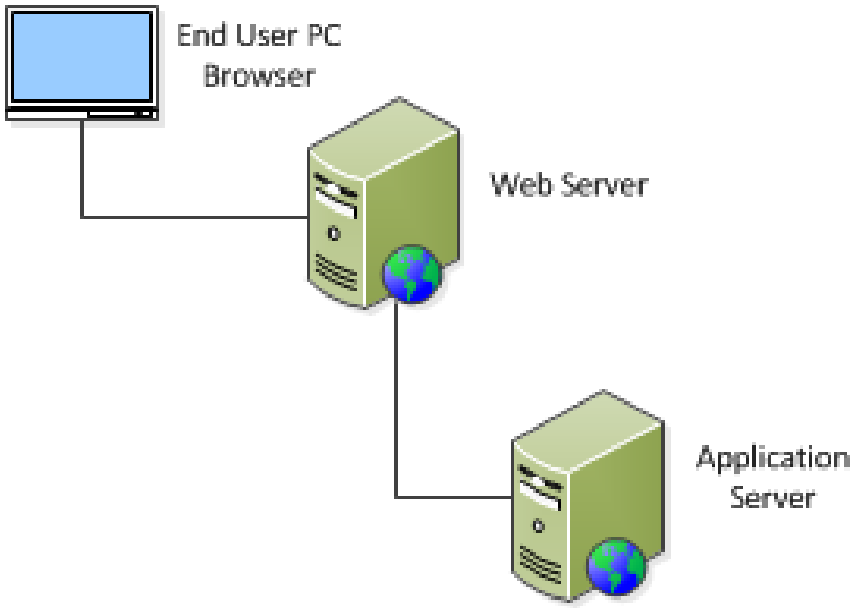
Application Architecture

- Let’s say you work in Human Resources at a small company
- You use the browser on your PC to access the Payroll System



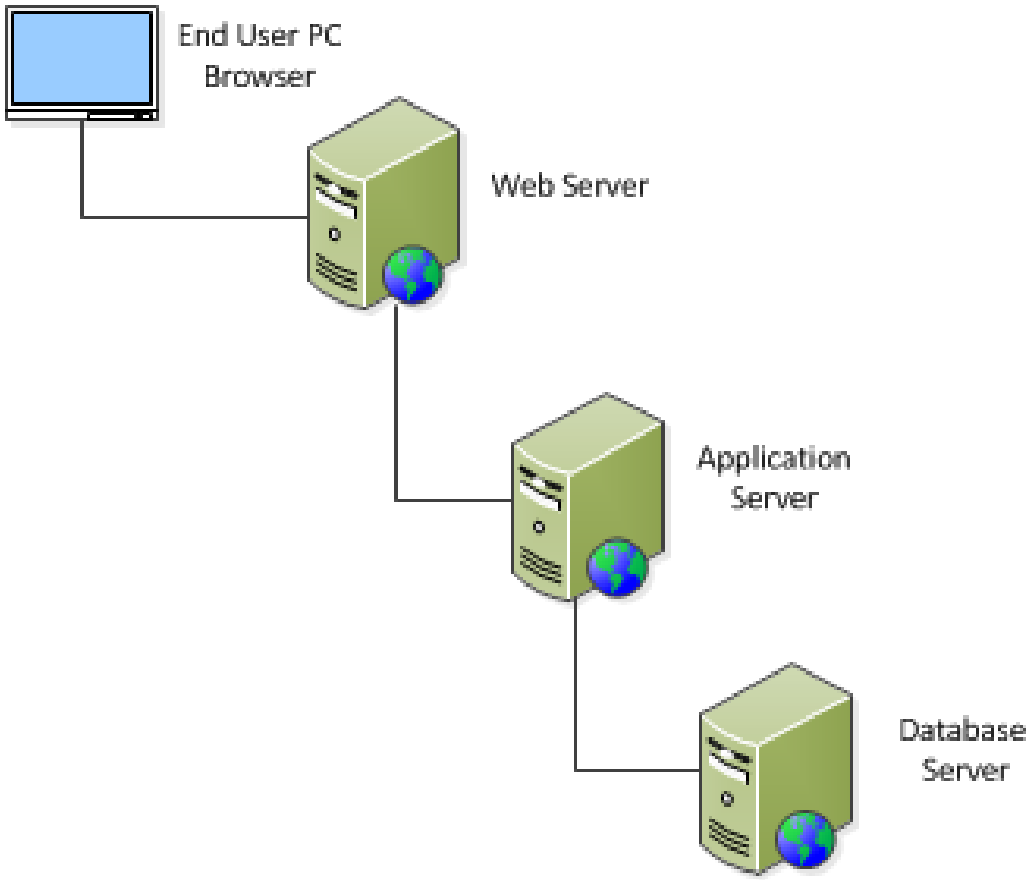
Application Architecture

- Let’s say you work in Human Resources at a small company
- You use the browser on your PC to access the Payroll System



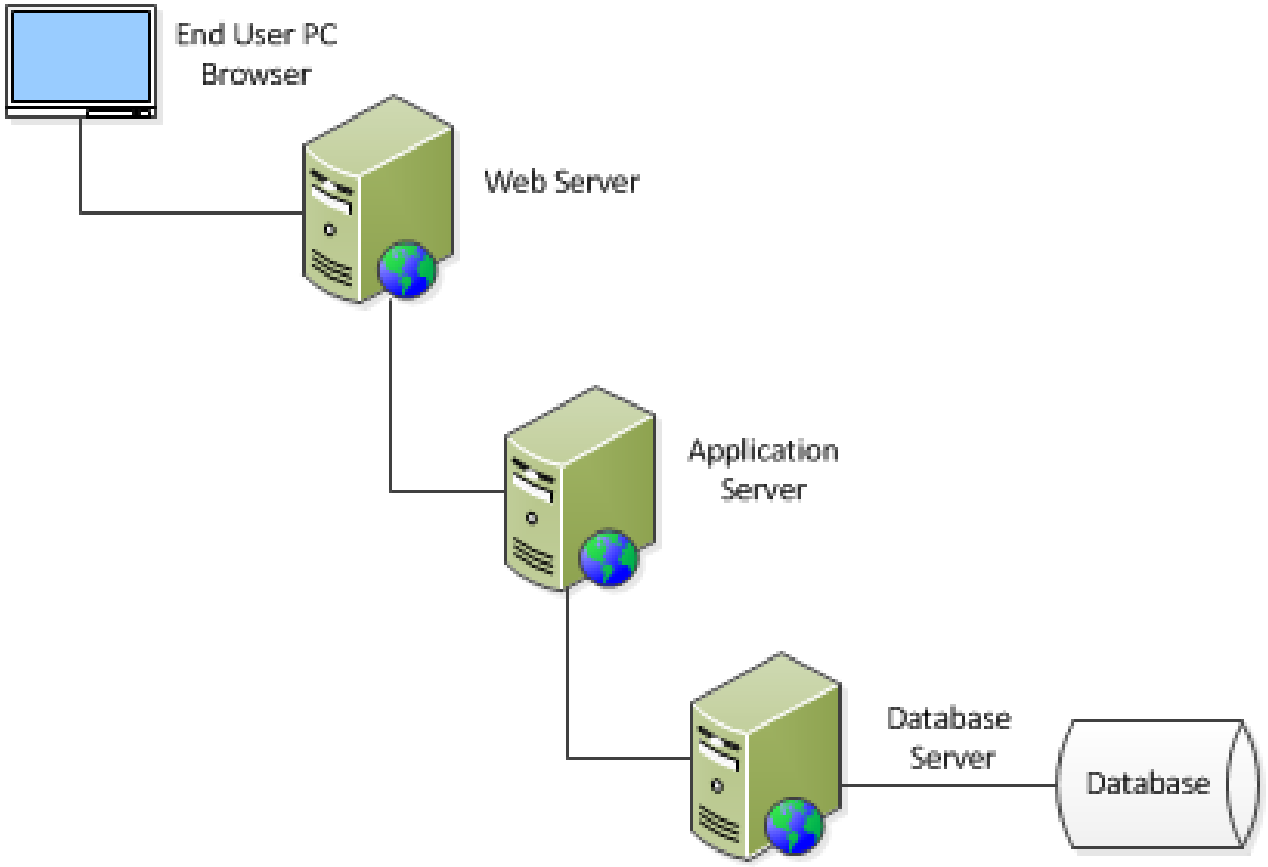
Application Architecture

- Let's say you work in Human Resources at a small company
- You use the browser on your PC to access the Payroll System



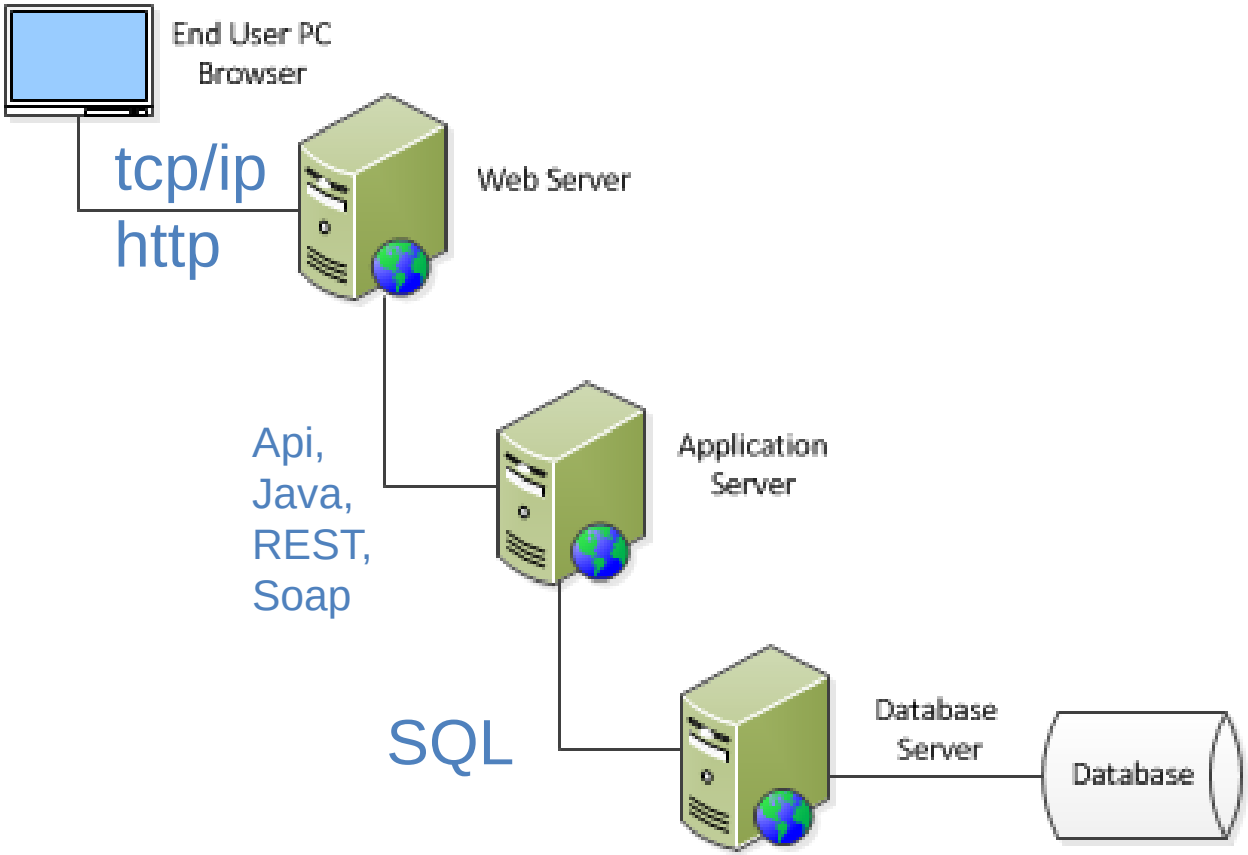
Application Architecture

- Let's say you work in Human Resources at a small company
- You use the browser on your PC to access the Payroll System

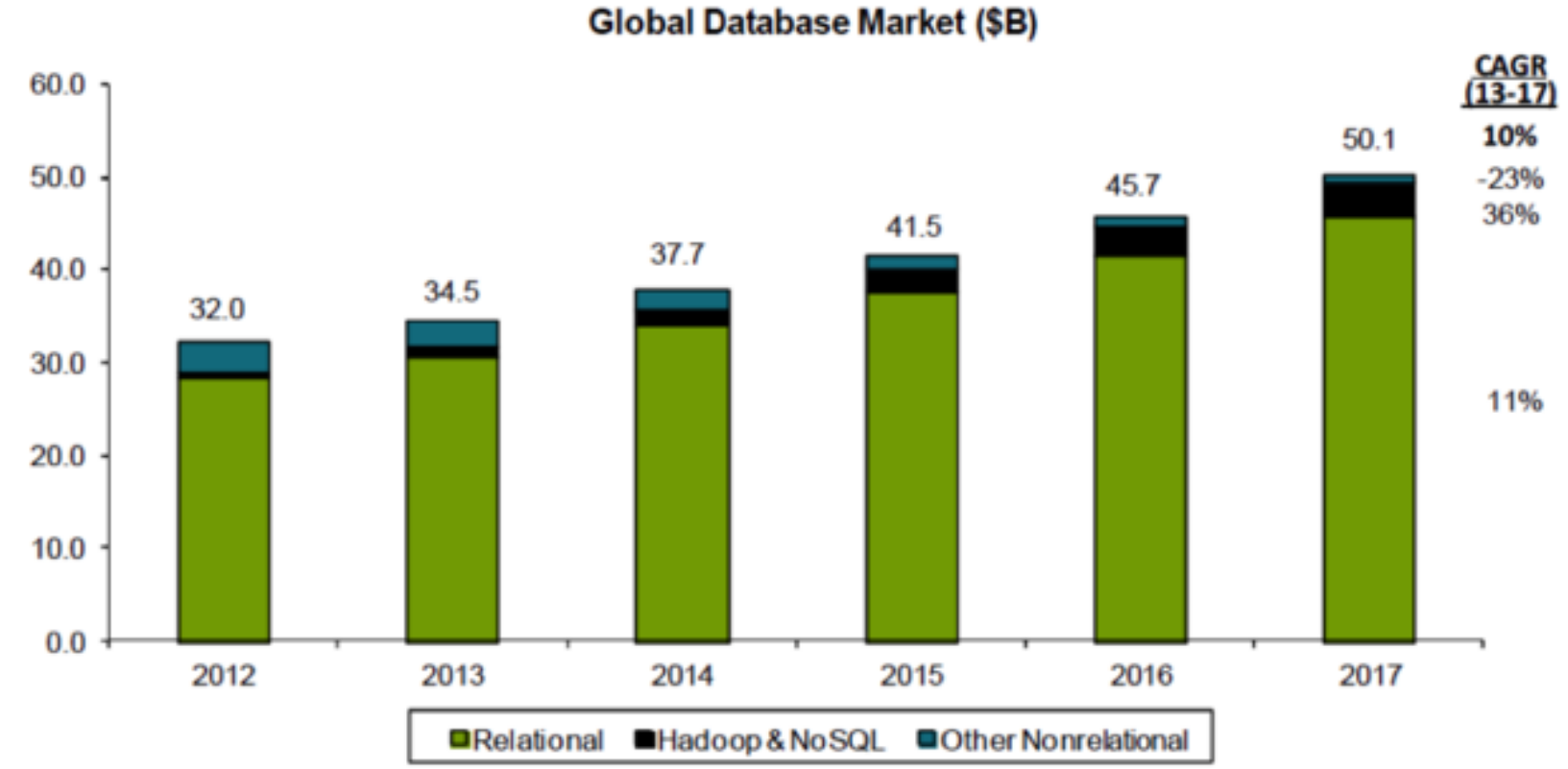


Application Architecture

- Let's say you work in Human Resources at a small company
- You use the browser on your PC to access the Payroll System



DBMS Software Use



Source: IDC, Bernstein analysis

DBMS Software Use

315 systems in ranking, September 2016

Rank			DBMS	Database Model	Score		
Sep 2016	Aug 2016	Sep 2015			Sep 2016	Aug 2016	Sep 2015
1.	1.	1.	Oracle	Relational DBMS	1425.56	-2.16	-37.81
2.	2.	2.	MySQL	Relational DBMS	1354.03	-3.01	+76.28
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1211.55	+6.51	+113.72
4.	↑ 5.	↑ 5.	PostgreSQL	Relational DBMS	316.35	+1.10	+30.18
5.	↓ 4.	↓ 4.	MongoDB	Document store	316.00	-2.49	+15.43
6.	6.	6.	DB2	Relational DBMS	181.19	-4.70	-27.95
7.	7.	↑ 8.	Cassandra	Wide column store	130.49	+0.26	+2.89
8.	8.	↓ 7.	Microsoft Access	Relational DBMS	123.31	-0.74	-22.68
9.	9.	9.	SQLite	Relational DBMS	108.62	-1.24	+0.97
10.	10.	10.	Redis	Key-value store	107.79	+0.47	+7.14

Top 10 out of 285 by popularity (links, google searches, job offers, twitter)
Source: http://db-engines.com/en/ranking

DBMS Licensing Cost

- Commercial Products
  - MS Access is FREE
  - MySQL is FREE (open source GPL license), but we buy support
  - Microsoft SQL Server, Enterprise Edition
    - List price is about $15,000 per core
  - Oracle, Enterprise Edition
    - List price is about $47,500 per core

So, if I am running a 4 CPU HP server
And each CPU is “quad core”
My server is using 16 cores
To license MS SQL Server:
16 * $15,000 = $240,000
To license Oracle Enterprise DBMS:
16 * $47,500 = $760,000

MySQL

- “Open Source” relational database management software
  Written by Monty Widenus (Finland)
- Monty formed MySQL AG to govern it
- MySQL AG was purchased by Sun Microsystems
- Sun Microsystems was purchased by Oracle
- Monty branched out and created MariaDB
- MySQL Big Three:
  - Oracle
  - MariaDB
  - Percona

Characteristics of Relational Databases

- The purpose of a database is to help people track things of interest to them
- Data is stored in tables, which have rows and columns like a spreadsheet
- A database stores BOTH data (in tables) AND relationships (between tables)

Characteristics of Relational Databases

- A database may have multiple tables, where each table stores data about a different thing
  - Example: a STUDENT table, a CUSTOMER table
- Each row in a table stores data about one occurrence of the thing of interest
  - Example: one student's data, one customer's data
- A database stores metadata – data about itself stored within itself
  - Example: in the DBMS catalog, there is a table containing one row for every table in the database

Data in Tables

ProductID	ProductName	SupplierID	CategoryID	QuantityPerUnit	UnitPrice	UnitsInStock	UnitsOnOrder
1	Chai	1	1	10 boxes x 20 bags	18.00	39	0
2	Chang	1	1	24 - 12 oz bottles	19.00	17	40
3	Aniseed Syrup	1	2	12 - 550 ml bottles	10.00	13	70
4	Chef Antons Cajun Seasoning	2	2	48 - 6 oz jars	22.00	53	0
5	Chef Antons Gumbo Mix	2	2	36 boxes	21.35	0	0
6	Grandmas Boysenberry Spread	3	2	12 - 8 oz jars	25.00	120	0
7	Uncle Bobs Organic Dried Pears	3	7	12 - 1 lb pkgs.	30.00	15	0
8	Northwoods Cranberry Sauce	3	2	12 - 12 oz jars	40.00	6	0
9	Mishi Kobe Niku	4	6	18 - 500 g pkgs.	97.00	29	0
10	Ikura	4	8	12 - 200 ml jars	31.00	31	0
11	Queso Cabrales	5	4	1 kg pkg.	21.00	22	30
12	Queso Manchego La Pastora	5	4	10 - 500 g pkgs.	38.00	86	0
13	Konbu	6	8	2 kg box	6.00	24	0
14	Tofu	6	7	40 - 100 g pkgs.	23.25	35	0
15	Genen Shouyu	6	2	24 - 250 ml bottles	15.50	39	0
16	Pavlova	7	3	32 - 500 g boxes	17.45	29	0
17	Alice Mutton	7	6	20 - 1 kg tins	39.00	0	0

SupplierID	CompanyName	ContactName	ContactTitle	Address	City
1	Exotic Liquids	Charlotte Cooper	Purchasing Manager	49 Gilbert St.	London
2	New Orleans Cajun Delights	Shelley Burke	Order Administrator	P.O. Box 78934	New Orleans
3	Grandma Kelly's Homestead	Regina Murphy	Sales Representative	707 Oxford Rd.	Ann Arbor
4	Tokyo Traders	Yoshi Nagase	Marketing Manager	9-8 Sekimai Musashino-shi	Tokyo
5	Cooperativa de Quesos 'Las Cabras'	Antonio del Valle Saavedra	Export Administrator	Calle del Rosal 4	Oviedo
6	Mayumi's	Mayumi Ohno	Marketing Representative	92 Setsuko Chuo-ku	Osaka
7	Pavlova Ltd.	Ian Devling	Marketing Manager	74 Rose St. Moonie Ponds	Melbourne
8	Specialty Biscuits Ltd.	Peter Wilson	Sales Representative	29 King's Way	Manchester
9	PB Knackebrod AB	Lars Peterson	Sales Agent	Kaloadagatan 13	Goteborg
10	Refrescos Americanas LTDA	Carlos Diaz	Marketing Manager	Av. das Americanas 12.890	Sao Paulo
11	Heli Susswaren GmbH & Co. KG	Petra Winkler	Sales Manager	Tiergartenstrasse 5	Berlin
12	Plutzer Lebensmittelgrossmarkte AG	Martin Bein	International Marketing Mgr.	Bogenallee 51	Frankfurt
13	Nord-Ost-Fisch Handelsgesellschaft mbH	Sven Petersen	Coordinator Foreign Markets	Frahmredder 112a	Cuxhaven
14	Formaggi Fortini s.r.l.	Elio Rossi	Sales Representative	Viale Dante 75	Ravenna
15	Norske Meierier	Beate Vileid	Marketing Manager	Hatlevegen 5	Sandvika
16	Bigfoot Breweries	Cheryl Saylor	Regional Account Rep.	3400 - 8th Avenue Suite 210	Bend
17	Svensk Sjofoda AB	Michael Bjorn	Sales Representative	Brovallavagen 231	Stockholm
18	Aux joyeux ecclesiastiques	Guylene Nodier	Sales Manager	203 Rue des Francs-Bourgeois	Paris
19	New England Seafood Cannery	Robb Merchant	Wholesale Account Agent	Order Processing Dept. 2100 Paul Revere Blvd.	Boston
20	Leka Trading	Chandra Leka	Owner	471 Serangoon Loop Suite #402	Singapore
21	Lyngbysild	Niels Petersen	Sales Manager	Lyngbysild Fiskebakken 10	Lyngby
22	Zaanse Snoepfabriek	Dirk Luchte	Accounting Manager	Verkoop Rijnweg 22	Zaandam

The Relational Model

- Introduced in 1970
- Created by E.F. Codd
  - An IBM systems engineer and mathematician
  - Codd's model is based on the branch of mathematics known as "relational algebra"
- Now the standard model for most commercial relational DBMS products.

Important Relational Model Terms

- Entity
- Attribute
- Relation
- Composite key
- Primary key
- Surrogate key
- Foreign key
- Referential integrity constraint
- Normalization

Entity

- An entity is some identifiable person, place, thing or event that users want to keep track of (that is, store data about)
  - Customers
  - Students
  - Computers
  - Sales

Attribute

- An attribute is a FACT or CHARACTERISTIC describing the occurrences of an entity
  - Customer first name, last name, middle initial
  - Student number, address, city, state zip
  - Computer manufacturer, model, processor
  - Sale item, quantity, date, dollar amount

Relation

- Relational DBMS products store data about entities in relations, which are a special type of table.
- A relation is a two-dimensional table that has the following characteristics: (Codd's relational rules)
  - Rows contain data about an entity.
  - Columns contain data about attributes of the entity.
  - All entries in a column are of the same kind.
  - Each column has a unique name.
  - One cell of the table holds a single value.
  - The order of the columns is unimportant.
  - The order of the rows is unimportant.
  - No two rows may be identical.
  - Every row has a column that uniquely identifies the row

A Relation

EmployeeNumber	FirstName	LastName	Department	Email	Phone
100	Jerry	Johnson	Accounting	JJ@somewhere.com	834-1101
200	Mary	Abernathy	Finance	MA@somewhere.com	834-2101
300	Liz	Smathers	Finance	LS@somewhere.com	834-2102
400	Tom	Caruthers	Accounting	TC@somewhere.com	834-1102
500	Tom	Jackson	Production	TJ@somewhere.com	834-4101
600	Eleanore	Caldera	Legal	EC@somewhere.com	834-3101
700	Richard	Bandalone	Legal	RB@somewhere.com	834-3102

Tables That Are Not Relations: Multiple Entries per Cell

EmployeeNumber	FirstName	LastName	Department	Email	Phone
100	Jerry	Johnson	Accounting	JJ@somewhere.com	834-1101
200	Mary	Abernathy	Finance	MA@somewhere.com	834-2101
300	Liz	Smathers	Finance	LS@somewhere.com	834-2102
400	Tom	Caruthers	Accounting	TC@somewhere.com	834-1102, 834-1191, 834-1192
500	Tom	Jackson	Production	TJ@somewhere.com	834-4101
600	Eleanore	Caldera	Legal	EC@somewhere.com	834-3101
700	Richard	Bandalone	Legal	RB@somewhere.com	834-3102, 834-3191

Tables That Are Not Relations: Required Row Order

EmployeeNumber	FirstName	LastName	Department	Email	Phone
100	Jerry	Johnson	Accounting	JJ@somewhere.com	834-1101
200	Mary	Abernathy	Finance	MA@somewhere.com	834-2101
300	Liz	Smathers	Finance	LS@somewhere.com	834-2102
400	Tom	Caruthers	Accounting	TC@somewhere.com	834-1102
				Fax:	834-9911
				Home:	723-8795
500	Tom	Jackson	Production	TJ@somewhere.com	834-4101
600	Eleanore	Caldera	Legal	EC@somewhere.com	834-3101
				Fax:	834-9912
				Home:	723-7654
700	Richard	Bandalone	Legal	RB@somewhere.com	834-3102

A Relation with Values of Varying Length

EmployeeNumber	FirstName	LastName	Department	Email	Phone	Comment
100	Jerry	Johnson	Accounting	JJ@somewhere.com	834-1101	Joined the Accounting Department in March after completing his MBA. Will take the CPA exam this fall.
200	Mary	Abernathy	Finance	MA@somewhere.com	834-2101	
300	Liz	Smathers	Finance	LS@somewhere.com	834-2102	
400	Tom	Caruthers	Accounting	TC@somewhere.com	834-1102	
500	Tom	Jackson	Production	TJ@somewhere.com	834-4101	
600	Eleanore	Caldera	Legal	EC@somewhere.com	834-3101	
700	Richard	Bandalone	Legal	RB@somewhere.com	834-3102	Is a full time consultant to Legal on a retainer basis.

Keys

- A key is a combination of one or more columns that is used to identify rows in a relation.
- A composite key is a key that consists of two or more columns (also referred to as a concatenated key)
- A primary key is a candidate key selected as the primary means of identifying rows in a relation.
  - There is only one primary key per relation.
  - The primary key may be a composite key.
  - The ideal primary key is short, numeric, and never changes.

Surrogate Keys

- A surrogate key is an artificial column added to a relation to serve as a primary key.
  - DBMS supplied (a sequence number + 1)
  - Short, numeric, and never changes—an ideal primary key
  - Has artificial values that are meaningless to users

Surrogate Keys

- NOTE: The primary key of the relation is underlined below:

- RENTAL_PROPERTY without surrogate key:

  RENTAL_PROPERTY (Street, City, State/Province, Zip/PostalCode, Country, Rental_Rate)

-

- RENTAL_PROPERTY with surrogate key:

  RENTAL_PROPERTY (PropertyID, Street, City, State/Province, Zip/PostalCode, Country, Rental_Rate)

Foreign Keys

- A foreign key is the primary key of one relation that is placed in another relation to form a link between the relations.
  - A foreign key can be a single column or a composite key.
  - The term refers to the fact that key values are not primary to the relation in which they appear as foreign key values.

The Referential Integrity Constraint

- A referential integrity constraint is a rule that limits the values of the foreign key to those already existing as primary key values in the corresponding relation.
- In other words, the constraint keeps me from adding a row to a table if the value in a foreign key column is “not on file”

- Example: (consider slide # 21)

  – I cannot insert a new PRODUCT into the Product table if it has an invalid (not on file) value in in the SupplierID column referencing the SUPPLIER table

Designing a Database

- Data must be normalized
- Construct a Data Model
  - Define all keys
  - Define all attributes

Normalization

- A structured, defined, detailed process
- Prepares the database design to make sure it complies with Codd's rules for the RELATIONAL MODEL

Normalization

Objective of Normalization

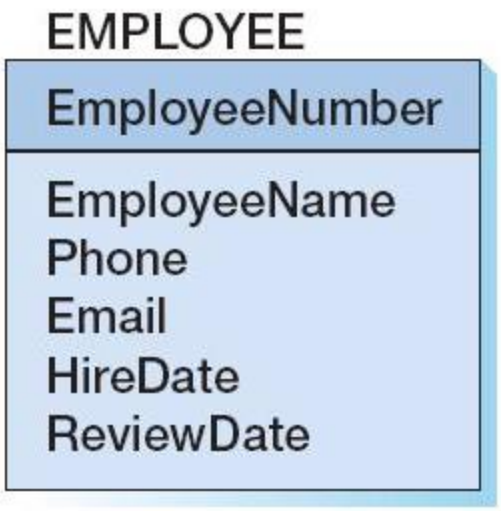
- To Arrange the data into a series of clearly defined, well-organized RELATIONS
  - Each with a primary key
  - All attributes are fully dependent on the primary key
  - Each relation may have foreign keys referencing other relations

Normalization

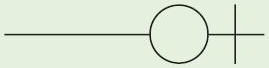
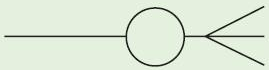
- Normalization Step-by-Step
  - First Normal Form
    - Remove any multi-valued cells and/or any rows requiring a specific sequence
  - Second Normal Form
    - For entities with composite keys, make sure that all attributes are dependent on the full key
  - Third Normal Form
    - Make sure that no attributes are dependent on any other non-key attributes

- Symbols you will use in data modeling

Entity (rectangle) with an entity name, Primary Key, and Attributes listed



Cardinality & Optionality Symbols

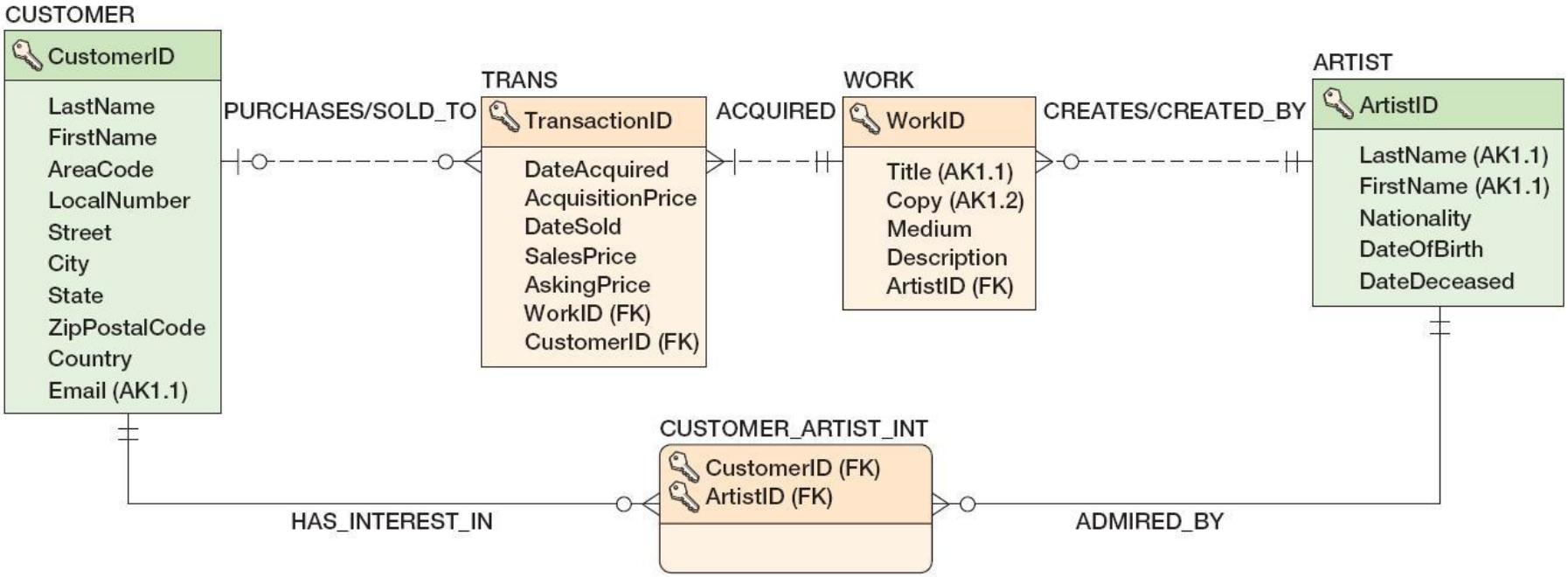
Symbol	Meaning
	One—Mandatory
	Many—Mandatory
	One—Optional
	Many—Optional

Data Modeling

- Cardinality
  - How many of THESE are related to how many of THESE
  - Typically: zero, one, or many
  - On both ends of the relationship
- Optionality
  - Is the relationship mandatory (one or more) or optional (zero)

Data Modeling

- An example:

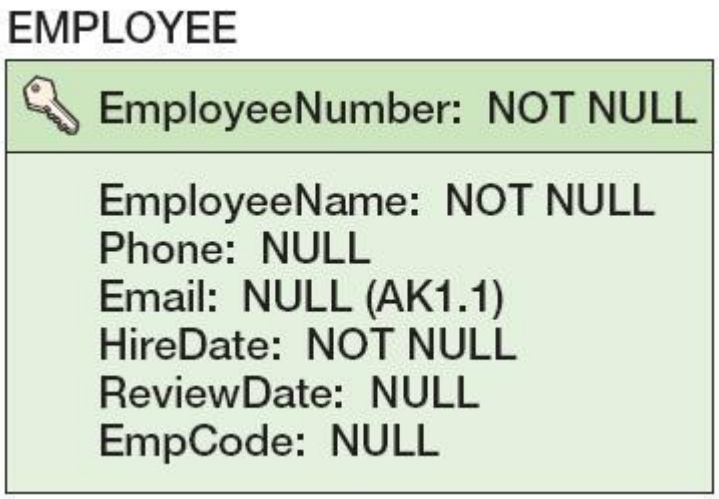


Look at this data model, then consider the questions on the next slide.

Data Modeling

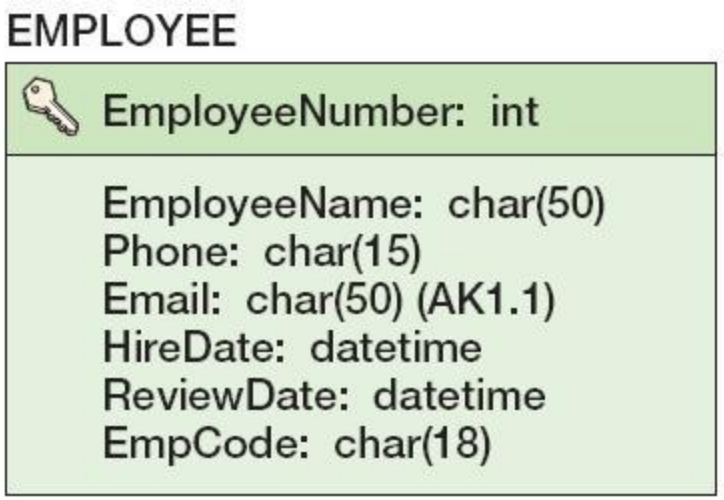
- What are the names of the five entities?
- What is the primary key of each entity?
- Why are some of the relationship lines dashed, and some are solid?
- Why do 4 of the entities have square corners and one has rounded corners?
- Which entity has a composite (or “concatenated”) key?
- Relationship descriptions are read clockwise: a customer purchases a work; a work is sold to a customer. Which entity represents the fact that a customer purchased a work?

- **Null status** indicates whether or not the value of the column can be NULL.



Specify Column Properties: Data Type

- Generic data types:
  - CHAR(n)
  - VARCHAR(n)
  - DATE
  - TIME
  - MONEY
  - INTEGER
  - DECIMAL



In Summary

- Databases are Ubiquitous
- DBMS software is extremely powerful and complex
- Oracle and MS SQL Server are expensive, yet widely used
- MySQL is free, and widely used
- Relational DBMS software holds 90% of market share, but
- NoSQL and Hadoop are booming so folks can handle "Big Data"
- Database design requires specialized skills

- Careers using Databases:
  - Database Administrator ("DBA")
  - Application Developer (uses SQL)
  - Data Architect (designs databases)
  - Data Scientist (uses databases for analytics)