What is a Database?

Data versus Information

Data is collected but not processed – unprocessed data
Information is data that has been processed - processed data

Data

- Consists of raw facts
- Stored in the tables

Information

- Interpretation of these facts
- Manipulated and presented as information

Database = a collection of information managed by DBMS software

Typically, the information has structure, and the database reflects this structure

What is a Database?

Introduction

- Database Analysis and Design is concerned with the creation of well structured databases.
  - Analysis - The collection and understanding of user requirements.
  - Design - The creation and design of mappings of the requirements to a database.

What is a Database?

Database Management System

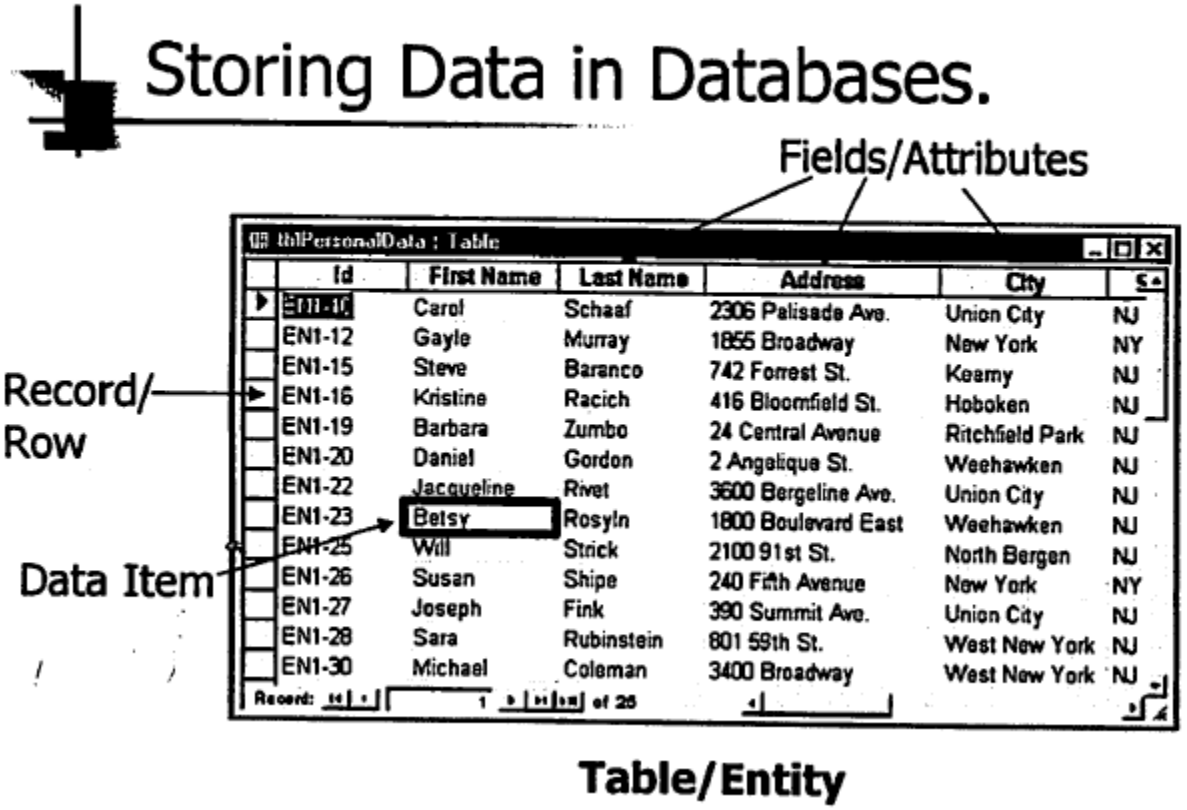
- A Database Management System (DBMS) is a software system for manipulating databases.
- A DBMS supports:
  - Logical views (known as schema).
  - Physical view (access methods)
  - Data Definition Language
  - Data Manipulation Language
  - Utilities
    - (transaction management, concurrency control, data integrity,security, disaster recovery)

What is a Database?

Database Tables

- A single database will usually have many tables.
- The ONLY place data is stored in a database is in it's tables.
- Each table will hold information about only ONE "thing".
- Each field in the table will further describe this one "thing".
- Each record in the table will define one *instance* of the "thing".

What is a Database?



Id	First Name	Last Name	Address	City	S
EN1-10	Carol	Schaaf	2306 Palisade Ave.	Union City	NJ
EN1-12	Gayle	Murray	1855 Broadway	New York	NY
EN1-15	Steve	Baranco	742 Forrest St.	Kearny	NJ
EN1-16	Kristine	Racich	416 Bloomfield St.	Hoboken	NJ
EN1-19	Barbara	Zumbo	24 Central Avenue	Ritchfield Park	NJ
EN1-20	Daniel	Gordon	2 Angelique St.	Weehawken	NJ
EN1-22	Jacqueline	Rivet	3600 Bergeline Ave.	Union City	NJ
EN1-23	Betsy	Rosyln	1800 Boulevard East	Weehawken	NJ
EN1-25	Will	Strick	2100 91st St.	North Bergen	NJ
EN1-26	Susan	Shipe	240 Fifth Avenue	New York	NY
EN1-27	Joseph	Fink	390 Summit Ave.	Union City	NJ
EN1-28	Sara	Rubinstein	801 59th St.	West New York	NJ
EN1-30	Michael	Coleman	3400 Broadway	West New York	NJ

Record: 1 of 26

A Little History

- DBMS software allows me to query the information in the database
- DBMS software supports huge data sets (terabytes, petabytes) plus history of information over time
- DBMS provides durability – don't let me lose anything
- DBMS helps me control access

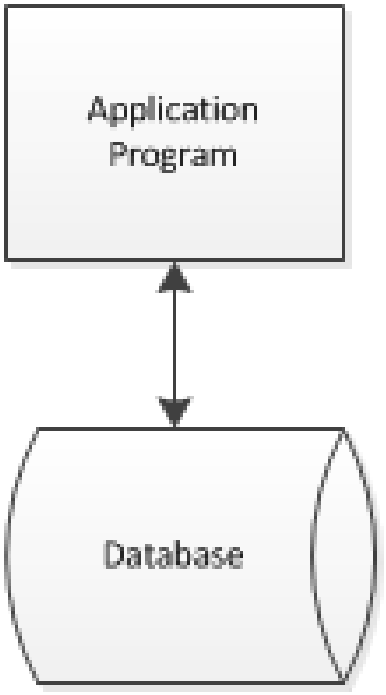
At first, there were files.

Early databases allowed me to link records in one file to records in another file (via "pointers" at the record level)

The database allowed me to select a starting point and read through chains of linked records

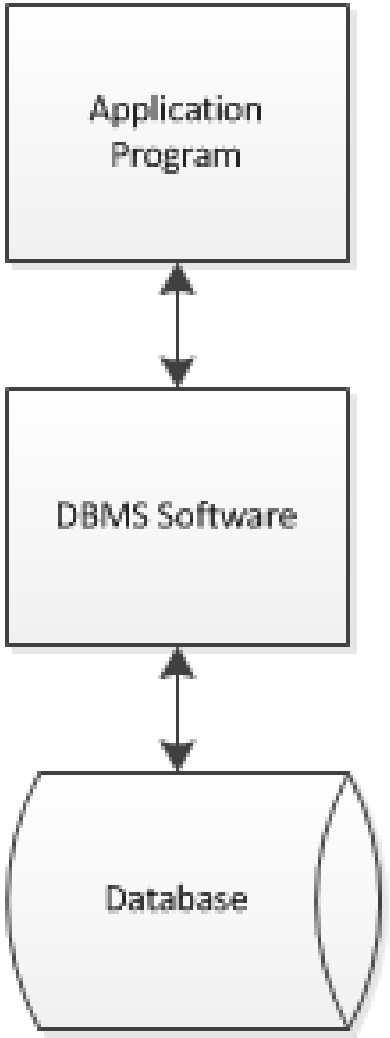
Application programs were required to programmatically manage all files, records, and pointers.

Early Databases



Every program required code to be written to directly map file & record layouts, starting points, and chains.

The Relational Model



Adds a layer of separation between the application code and the files, records, fields, pointer, chains.

Revolutionized database software

The DBMS

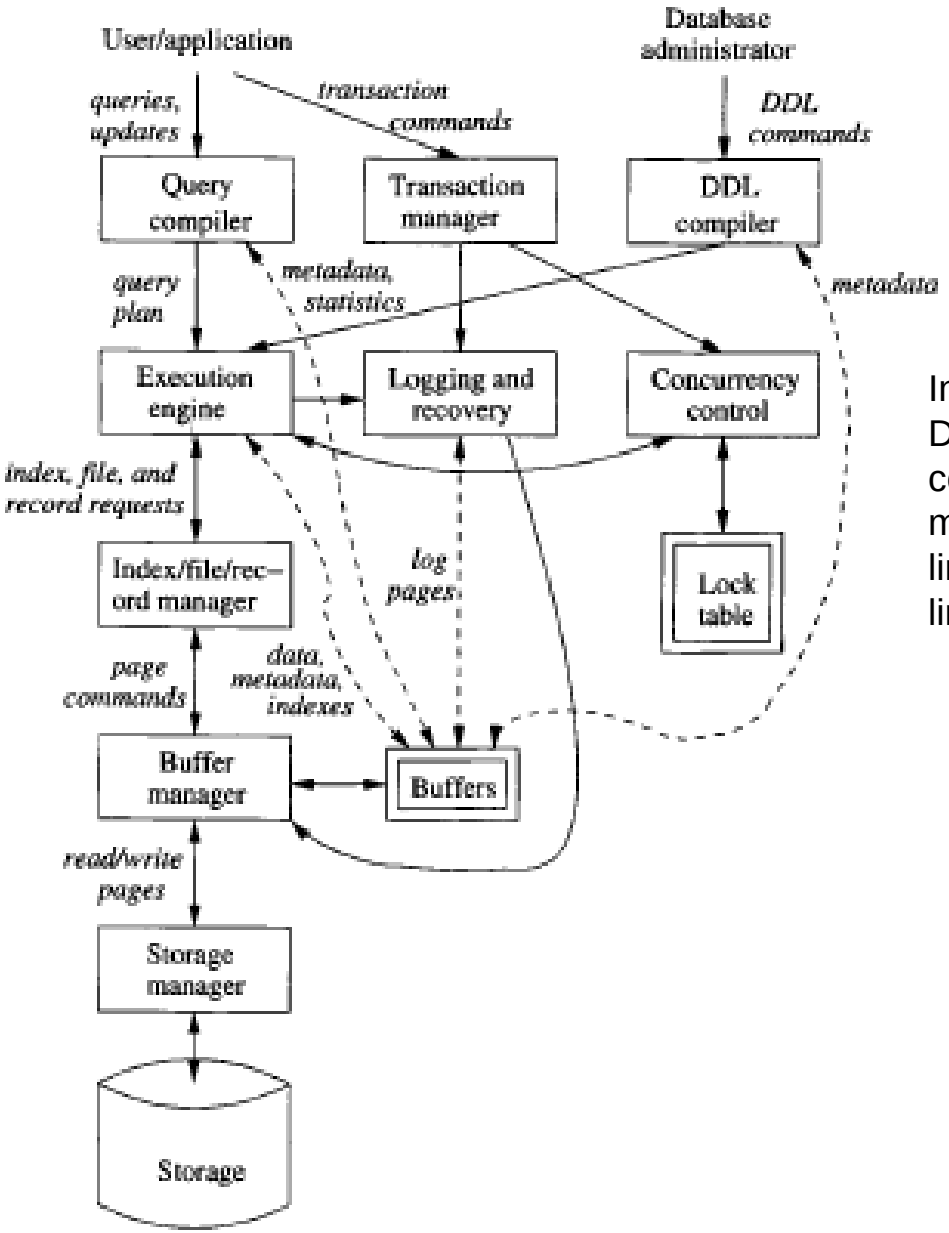


Figure 1.1: Database management system components

Overview of a Database Management System

In Fig. 1.1 we see an outline of a complete DBMS. Single boxes represent system components, while double boxes represent in-memory data structures. The solid lines indicate control and data flow, while dashed lines indicate data flow only.

The Relational Model

Dr. Edgar F. Codd
wrote a famous article* describing how a relational model could eliminate the lock between the application and the data

- Data is stored in TABLES (“relations”)
- Tables are related to each other by keys
- Data is queried using a language (SQL) based on relational algebra.

* “A relational model for large shared data banks”, 1970, in the ACM journal.

The Relational Model

The relational database is “self-describing”

It contains within itself all information about itself

(referred to as “metadata”)

For example, the DBMS keeps a “Catalog”

The catalog consists of tables.

Catalog tables contain information about

tables, columns, rows, indexes, users, etc.

Metadata

USER_TABLES Table

TableName	NumberColumns	PrimaryKey
STUDENT	3	StudentNumber
CLASS	4	ClassNumber
GRADE	3	(StudentNumber, ClassNumber)

USER_COLUMNS Table

ColumnName	TableName	DataType	Length (bytes)
StudentNumber	STUDENT	Integer	4
LastName	STUDENT	Text	25
FirstName	STUDENT	Text	25
EmailAddress	STUDENT	Text	100
ClassNumber	CLASS	Integer	4
Name	CLASS	Text	25
Term	CLASS	Text	12
Section	CLASS	SmallInteger	2
StudentNumber	GRADE	Integer	4
ClassNumber	GRADE	Integer	4
Grade	GRADE	Decimal	(2, 1)

ACID

The DBMS software must maintain ACID compliance

A = Atomicity. A transaction is all or nothing

C = Consistency. A transaction must maintain data consistency.

That is, all databases have consistency constraints (e.g., account balances may not be negative after a transaction finishes).

I = Isolation. Multiple users/transactions cannot step on each other.

D = Durability. Once a transaction is finished, the results cannot be lost.

ACID

Payroll Example:

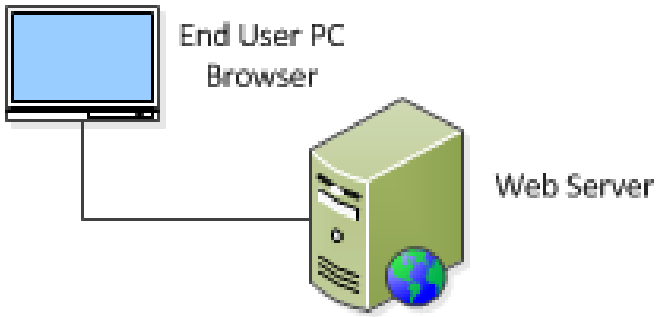
Paying an employee requires updates to
- The Payroll Register table
- The General Ledger Journal Entries table

What if a transaction dies midstream?

Commit and Rollback

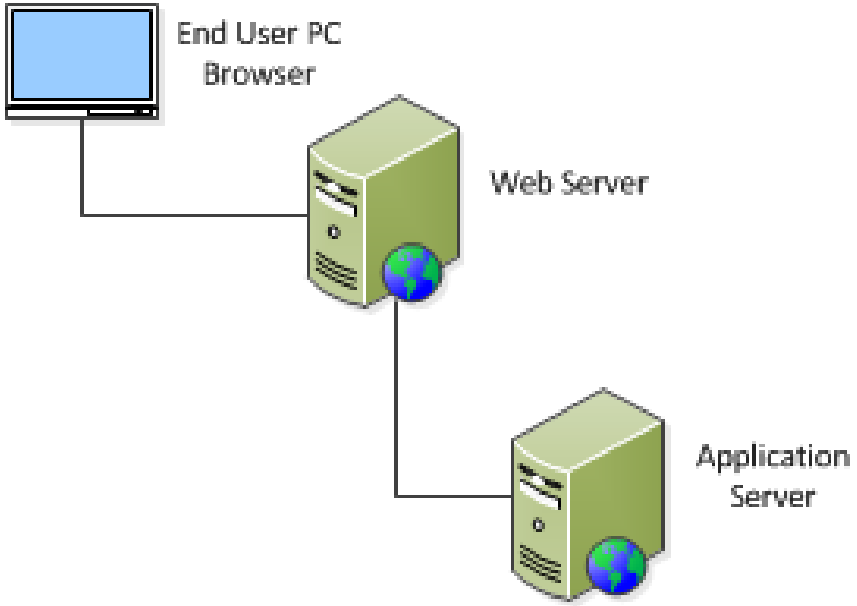
Application Architecture

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



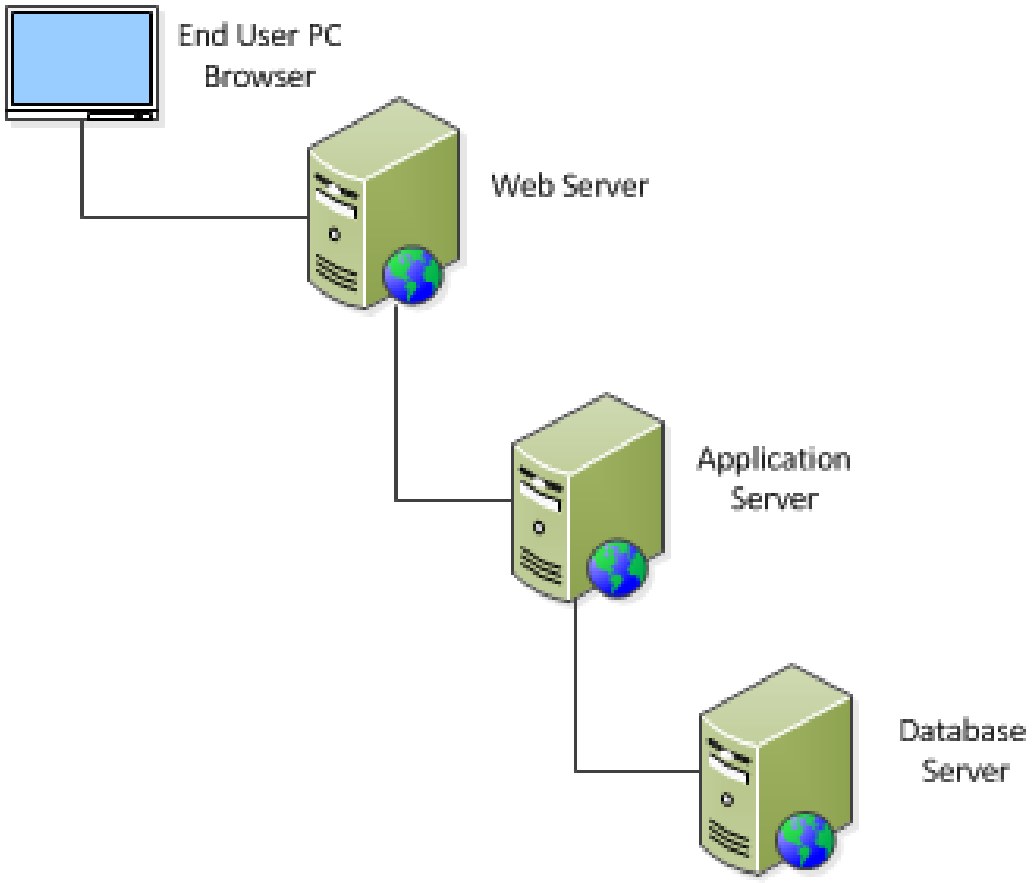
Application Architecture

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



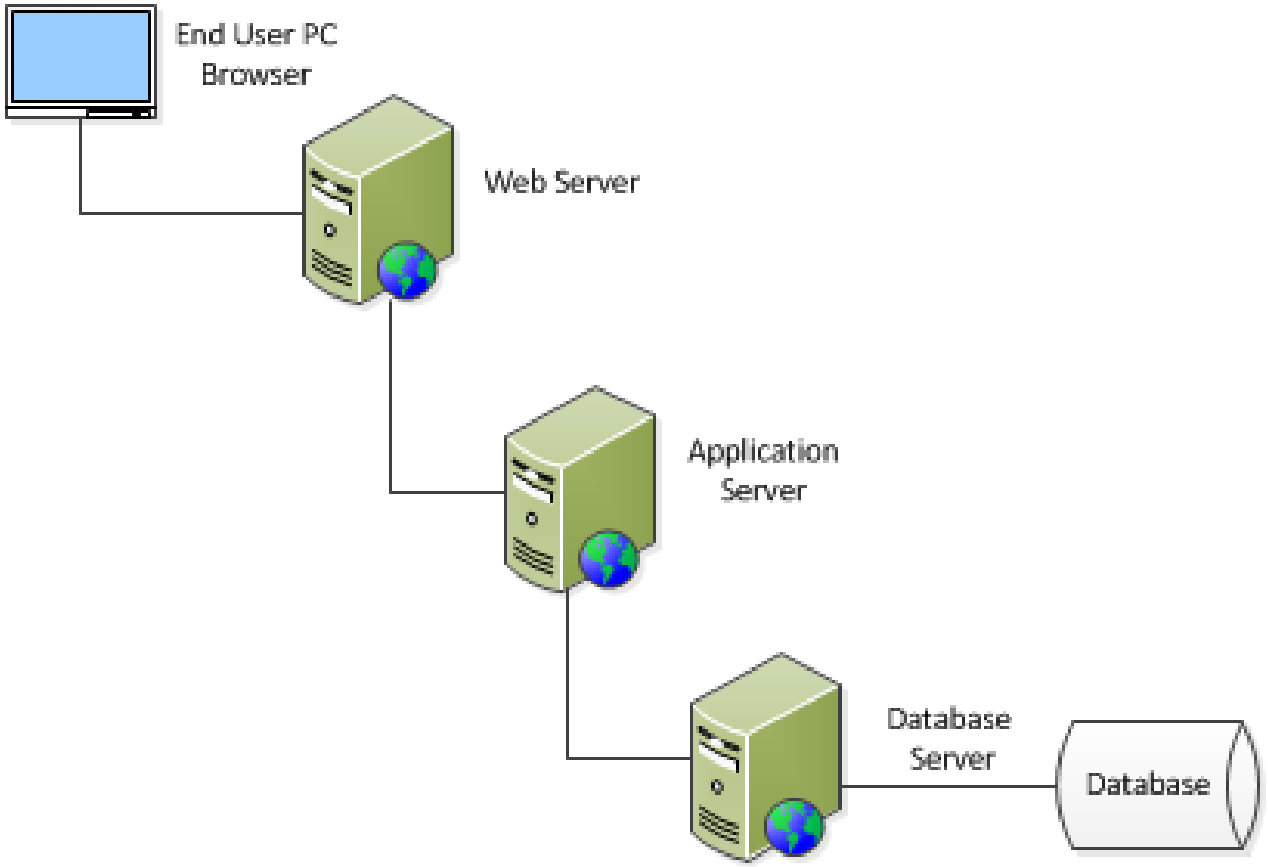
Application Architecture

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



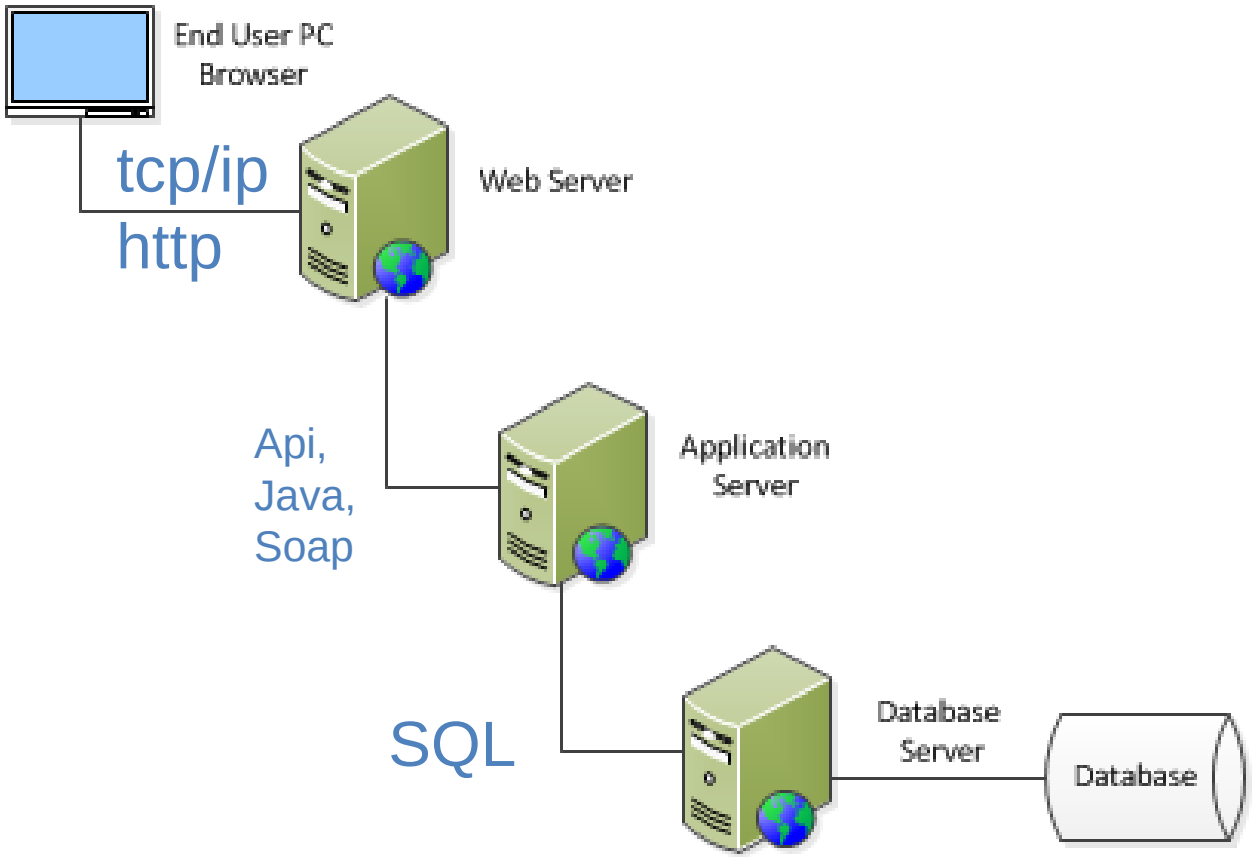
Application Architecture

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

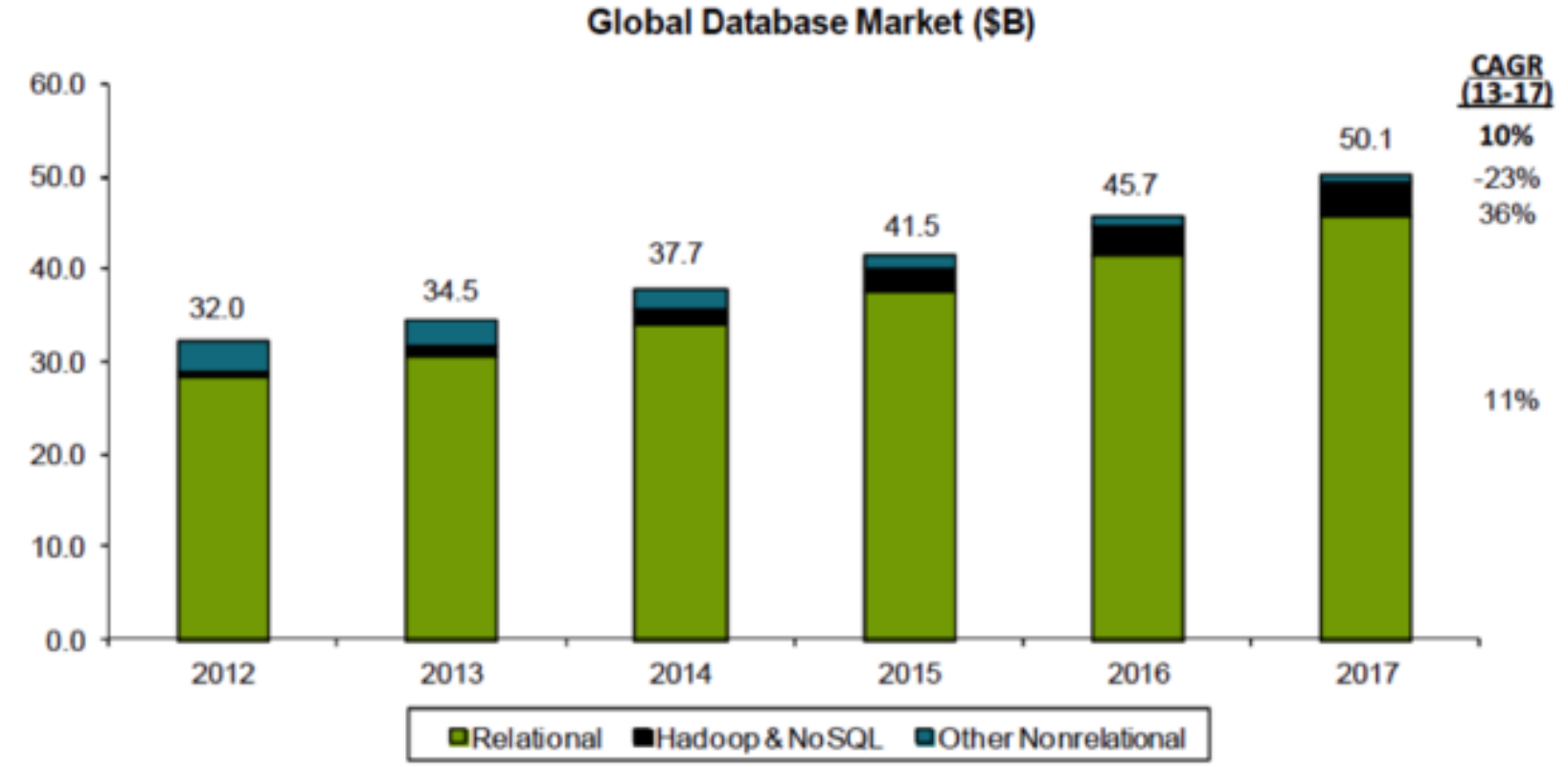


Application Architecture

- 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

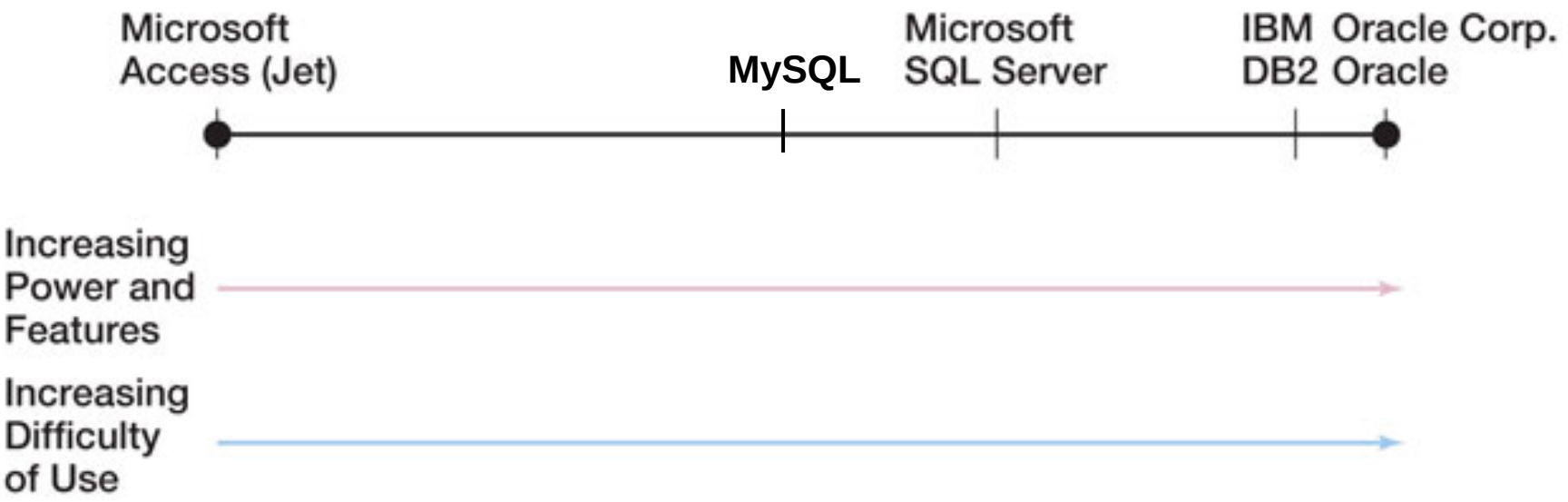
DBMS Software Use

315 systems in ranking, September 2016

Rank Sep 2016	Rank Aug 2016	Rank Sep 2015	DBMS	Database Model	Score Sep 2016	Score Aug 2016	Score 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

Relational DBMS Power vs. Ease of Use



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)

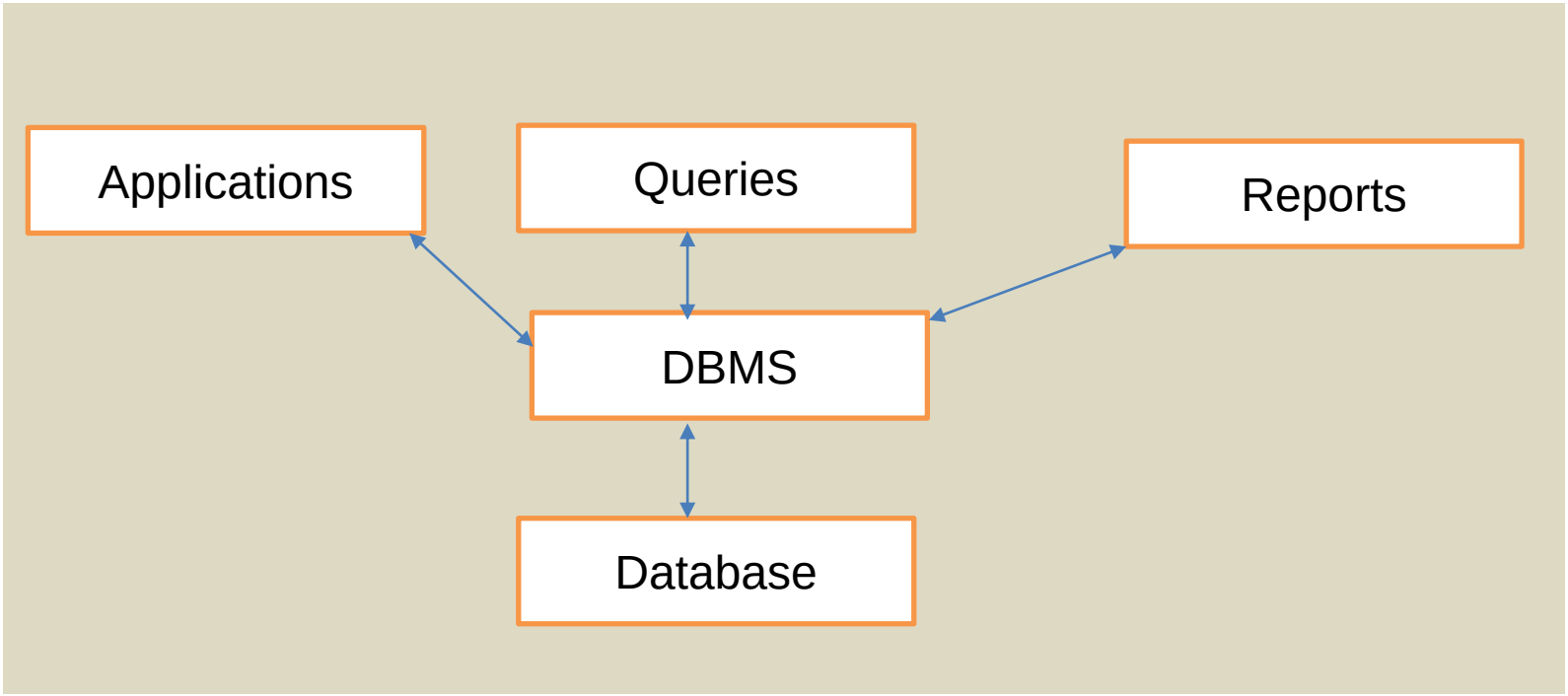
Characteristics of Relational Databases

- 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

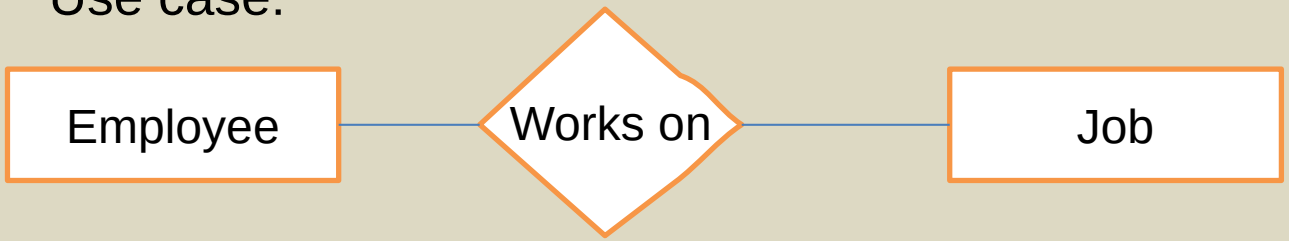
Data in Tables

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
18	Carnarvon Tigers	7	8	16 kg pkg.	62.50	42	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


Applications
Queries
Reports
DBMS
Database

- Use case:



- Employee = (emp#, fname, lname, dept)
- Works on (emp#, job#)
- Job = (job#, job_desc)
- There is 2 entities: Employee and Job
- There is one association: works on

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)

- Database is a collection of interrelated stored data: think of a electronic filing system.
- DBMS is a software system / application for manipulating databases.
- DBMS supports:
  - Logical views; Physical views; DDL; DML
- A single database normally contains many tables:
  - each table holds information about only one “thing”.
  - Each field in the table further describe this one “thing”.
  - Each record in the table define one instance of the “thing”.