Motivate

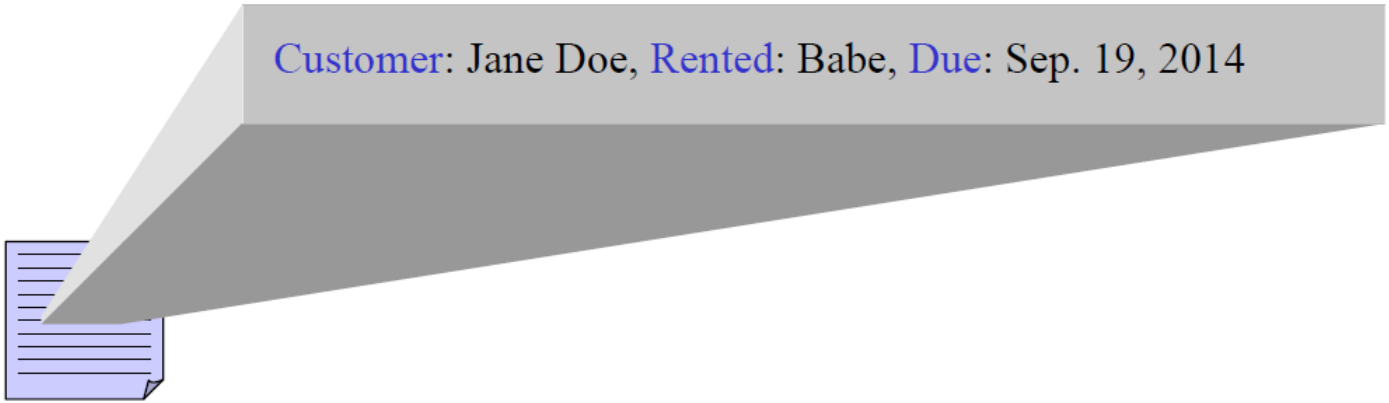
Data Management Example

- Scenario
  - You are a startup Netflix: Aggieflix
  - Customers rent DVD copies of movies.
  - Several copies of each movie.

- Needs
  - Which DVDs has a customer rented?
  - Are any DVDs overdue?
  - When will a DVD become available?

Solution: A File-based System

- Edit `rented.txt` file



- Advantages
  - Text editors are easy to use
  - Simple to insert/delete/update records

Complication: Queries

- Does not address needs

  Query: What DVDs has Joe Jenkins rented?

  Execute (not quite right): Search for ‘Joe Jenkins’.

  Execute: Search for ‘^\s+Customer:\s*Joe\s+Jenkins\s*,\s+Rented:’.

  Query: Are any DVDs overdue?

  Execute: ???

- Requirements
  - Robust, sophisticated query language
  - Clear separation between data organization (schema) and data

Database Concepts
Schema
DML
SQL

Complication: Integrity

- Lacks data *integrity*, *consistency*
  - Clerk misspells value/field

    Customer: Jane Doek, Rented: Bourne Identity, Deu: Sep. 19, 2014
  - Inputs improper value, same value differently

    Customer: Jane Doe, Rented: The Bourne Identity, Due: Feb. 29, 2014
  - Forgets/adds/reorders field

    Terms: weekly special Due: Sep. 19, 2014, Rented: Bourne Identity

- Requirements
  - Enforce *constraints* to permit only *valid* information to be input.

Database Concepts
Integrity constraints
Types

Complication: Update

- Add/delete/update fields in every record
  - Record store location.

    Customer: Jane Doe, Rented: Babe, Due: Sep. 19, 2014, Store: Providence
  - Modify customer to first and surname.

    First: Jane, Surname: Doe, Rented: Babe, Due: Sep. 19, 2014
- Add/delete/update new information collections
  - `customer.txt` file to record information

    Customer: Jane Doe, Phone: 555 3344

- Requirements
  - Ability to manipulate the way data is organized.

Complication: Multiple Users

- Two clerks edit `rented.txt` file at the same time.
  1) Ben starts to edit `rented.txt`, reads it into memory.
  2) Sarah starts to edit `rented.txt`.
  3) Ben adds a record.
  4) Ben saves `rented.txt` to disk.
  5) Sarah saves `rented.txt` to disk.

  Ben's added record disappears!

- Requirements
  - Must support multiple readers and writers.
  - Updates to data must (appear to) occur in *serial* order.

Database Concepts
Serializability
Concurrency control

Complication: Crashes

- Crash during update may lead to inconsistent state.
  - Ben makes 250 of 500 edits to change Jane Doe to her preferred name Jan Doe. Does periodic saves.
  - In middle, Windows crashes!

- Requirements
  - Must update on all or none basis.
  - Implemented by *commit* or *rollback* if necessary.

Database Concepts
Transactions
Commit
Rollback
Recovery

Complication: Data Physically Separate

- Wants
  - Want to advise Batman fans about new Batman movie.
- Method
  - `customer.txt` contains addresses of customers.
  - Must merge with `rented.txt` to create mailing list.
- Problems
  - Text editors incapable of such a merge (write a program)
  - Several Joe Jenkins
  - No information on some customers!?
- Requirements
  - Uniquely identify each customer.
  - Make sure we have information on customers that rent DVDs.

Database Concepts
Joins
Keys
Foreign keys
Referential integrity

Complication: Security

- Customers want to know how many times a movie has been rented.
  - Provide access to *rented.txt*, but not to customer field, how to I do that in an editor?
- Underage clerks should not see history of R-rated rentals.
  - Keep two lists of rentals?
- Requirements
  - Ability to control who has access to what information.

Database Concepts
Security
Views

Complication: Efficiency

- Merge with other Aggieflix stores in a state
  - `rented.txt` file gets huge (gigabytes of data).
  - Slow to edit.
  - Slow to query for customer information.

- Requirements
  - New data structures to improve query performance.
  - System automatically modifies queries to improve speed.
  - Ability of system to scale to handle huge datasets.

Database Concepts
Indexes
Query optimization
Database tuning

Complication: New Needs

- Combined data collection leads to new kinds of queries.
  - What pairs of DVDs are often rented together?
    Calculate probability of movie combinations.
  - Do we need more copies of the X-men movie anywhere?
    Plot rental history of X-men by store area.

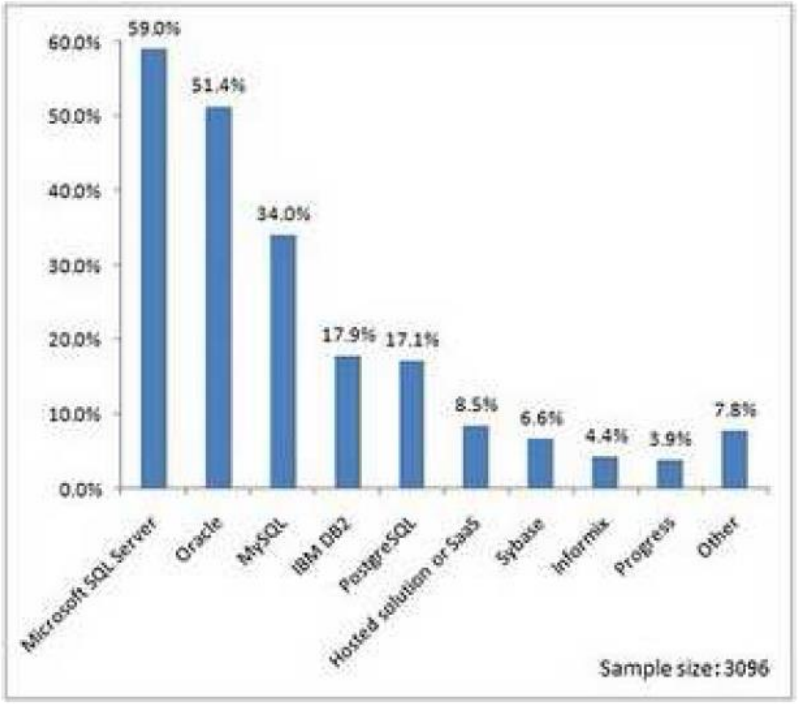
- Requirements
  - Collect and analyze summary data.
  - Use computer to *mine* for interesting trends.
  - Support access to data by sophisticated programs.

Database Concepts
Data warehouses
Data mining
Database API

Limitations of File-based Systems

- Program must implement
  - Security
  - Concurrency control
  - Support for schema reorganization
  - Performance enhancing data structures, e.g., indexes
- Observation
  - Many applications need these services.
- Solution
  - Build and sell a software system to provide services!

DBMS popularity



source: http://giannisgiataganas.blogspot.com/2010/06/business-process-management-in-free-and.html

Largest Software Companies

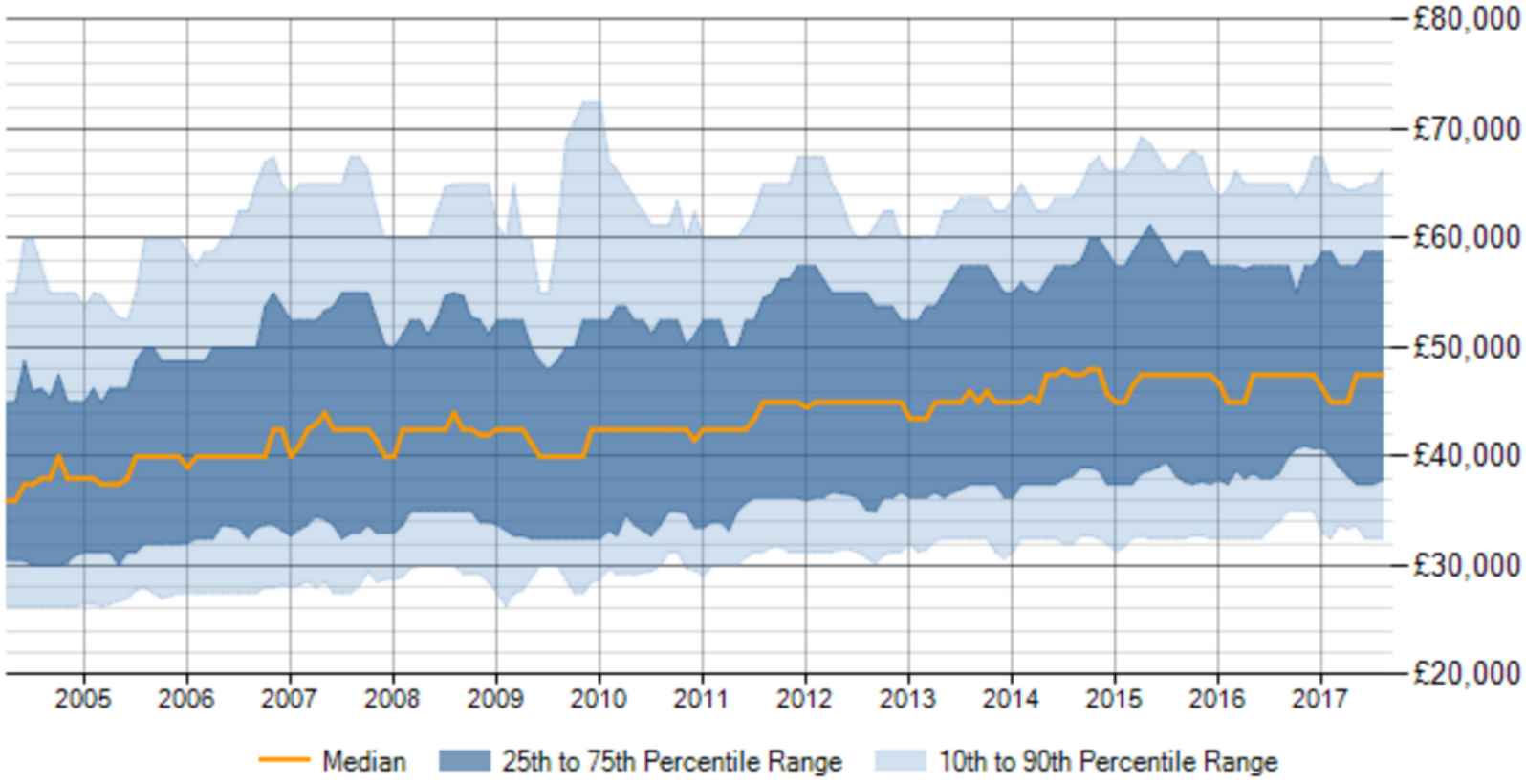
Table 1. Top 10 Worldwide Software Vendors, Worldwide, 2012-2013 (Billions of Dollars)

Rank 2013	Rank 2012	Vendor	2013 Revenue	2012 Revenue	2012-2013 Growth Rate (%)
1	1	Microsoft	65.7	62.0	6.0
2	3	Oracle	29.6	28.7	3.4
3	2	IBM	29.1	28.7	1.4
4	4	SAP	18.5	16.9	9.5
5	5	Symantec	6.4	6.4	-0.8
6	6	EMC	5.6	5.4	4.9
7	7	HP	4.9	5.0	-2.7
8	9	VMware	4.8	4.2	14.1
9	8	CA Technologies	4.2	4.3	-2.6
10	12	Salesforce.com	3.8	2.9	33.3
		Others	**234.6**	**224.0**	**4.7**
		Total	**407.3**	**388.5**	**4.8**

Source: Gartner (March 2014)

DBA
Salary Trend

This chart provides the 3-month moving average for salaries quoted in permanent IT jobs citing DBA.



Acknowledgements

- These slides were written by Richard T. Snodgrass (University of Arizona, SIGMOD chair), Christian S. Jensen (Aalborg University), and Curtis Dyreson (Utah State University).
- Michael Soo (amazon.com) provided some of the query processing slides.
- Kristian Torp (Aalborg University) converted the slides from Island Presents to Powerpoint.
- The motivational example was given by Gary Lindstrom and Wei Tao (University of Utah).