Foundation of Relational Databases

Yanlei Diao
A **database** is a large, integrated collection of data

A **database management system** (DBMS) is a software system designed to store and manage a large amount of data

- *Declarative interface* to define, add/update, and query data
- *Efficient querying*
- *Concurrent users*
- *Crash recovery*
- *Access control*...
What Type of Data is stored?

- All critical business data!
  - Banking
  - Ticketing
  - Retail
  - Electronic commerce
  - Insurance
  - Healthcare
  - Enterprise HR
  - Government
  - Telecommunications
  - Social networks

ORACLE CUSTOMERS INCLUDE
- 10 of the 10 top aerospace and defense companies
- 20 of the 20 top airlines
- 20 of the 20 top automotive companies
- 20 of the 20 top banks
- 9 of the 10 top consumer goods companies
- 9 of the 10 top engineering and construction companies
- 20 of the 20 top governments
- 20 of the 20 top high tech companies
- 20 of the 20 top insurers
- 20 of the 20 top manufacturers
- 20 of the 20 top oil and gas companies
- 20 of the 20 top pharmaceutical companies
- 20 of the 20 top retailers
- 10 of the 10 top SaaS providers
- 20 of the 20 top supply chains
- 20 of the 20 top telecommunications companies
- 20 of the 20 top universities
- 10 of the 10 top utilities
Early DBMS’s

- Many small data items, many queries and updates
  - Banking, airline reservations

- 1960s Navigational DBMS
  - Tree / graph-based data model
  - Manual navigation to find what you want
  - Support for “search” = “programming”

- 1973 Turing Award Winner
  - Charles William Bachman
  - “The Programmer as Navigator”
  - The network data model
Relational DBMS’s

- **Relational model** (1970)
  - Data independence: hides details of physical storage from users
  - Declarative query language: say *what* you want, not *how* to compute it
  - Mathematical foundation: what queries mean, possible implementations

- **1981 Turing Award Winner**
  - Edgar F. (“Ted”) Codd
  - Mathematically-inclined researcher
  - Legitimized DBMS’s as a theoretically respectable research field in CS
Relational DBMS

Query optimization (1970’s till now)
- Earliest: System R at IBM, INGRES at UC Berkeley
- Queries can be **efficiently executed** despite data independence and declarative queries!

2014 Turing Award Winner
- Michael Stonebraker (Berkeley / MIT)
- “For fundamental contributions to modern database systems”
Evolution of DBMS’s

INGRES
UC Berkeley, Stonebraker et al

System R
IBM San Jose, Gray, Selinger et al

Informix
Postgres
Sybase
MS SQL Server

IBM DB2
Oracle
MySQL
The Picture Today (Gartner 2015)