Database Systems I

Final Words

Instructor: Ouldooz Baghban Karimi

CMPT 354 - Summer 2019
Database Systems

1970
Navigational DBs
Relational DBs
SQL

2000
Data Warehouses
OLAP & OLTP
No SQL

2019
Big Data
New SQL
Database Systems I

• Challenges of data-intensive systems
  • Reliability
  • Concurrency
  • Performance
  • Accessibility
  • Security

• Relational Databases
  • Basics (Relational Model of Data, Relational Algebra, SQL)
  • Design (ER Model, Design Theory, Normalization, Constraint & Triggers)
  • Operate (Views, Indexing, Query Processing, Transaction Processing, Database Applications)
What to learn next?

• More on Relational DBMSs
  • Concurrency Control
  • Optimizer Implementation
  • Logging and Recovery Control
  • Join Algorithms
  • …

• High-level Trends & Architectures
  • Data Warehouses
  • Parallel Databases
  • Big Data Processing
  • Cloud Databases
  • …
Data Warehouses

• Collecting and organizing historical data from multiple sources

• Semi-structured & unstructured data

• Data Lake & Data Warehouse
Parallel Data Processing

• Multi-cores
  • Most processors have multiple cores
  • This trend will likely increase in the future

• Big data
  • Too large to fit in main memory
  • Distributed query processing on 100x-1000x servers
  • Widely available now using cloud services

• Parallel DBMSs
  • How to evaluate a parallel DBMS? Speedup (more nodes, same data: higher speed), Scaleup (more nodes, more data: same speed)
  • How to architect a parallel DBMS? Shared Memory, Shared Nothing, Shared Disk
  • How to partition data in a parallel DBMS?
    • Round Robin, Range Partitioning, Hash Partitioning (Load balancing vs. query performance)
Big Data Processing

• **Volume**: Data Size

• **Velocity**: Rate of data generation

• **Variety**: Data sources, formats, and workloads

• **Veracity**: Trustworthiness
Cloud Databases

- Internet-based computing
  - Amazon
    - Aurora (relational Database Engine)
    - Redshift (petabyte-scale data warehouse)
  - Microsoft
    - Azure SQL Server
    - Azure SQL Data Warehouse
  - Google
    - Spanner (relational database service)
    - BigQuery

\[
\text{Application} + \text{Cloud} = \text{SaaS} \text{ (Software as a service)}
\]
\[
\text{Platform} + \text{Cloud} = \text{PaaS} \text{ (Platform as a service)}
\]
\[
\text{Infrastructure} + \text{Cloud} = \text{IaaS} \text{ (Infrastructure as a service)}
\]

Picture from: https://azure.microsoft.com/en-ca/overview/what-is-paas/
Acknowledgements

I have used materials from the following resources in preparation of this course:

- **Database Systems: The Complete Book**
- **Database Systems (Kifer, Bernstein, Lewis)**
- **Database System Concepts**: [https://www.db-book.com](https://www.db-book.com)
- **Course offerings**
  - **CMPT 354 (Jiannan Wang - SFU)**: [https://sfu-db.github.io/cmpt354/](https://sfu-db.github.io/cmpt354/)
  - **W 4111 (Eugene Wu - Columbia)**: [https://w4111.github.io/](https://w4111.github.io/)
  - **CS 186 (Joe Hellerstein - Berkeley)**: [https://sites.google.com/site/cs186fall17/](https://sites.google.com/site/cs186fall17/)
  - **CSE 344 (Dan Suciu - Washington)**: [https://courses.cs.washington.edu/courses/cse344/17au/](https://courses.cs.washington.edu/courses/cse344/17au/)