Distributed Database Systems

Course Goal

1. Review the historic development of data management technologies;
2. Enhance the previous knowledge of database systems by deepening the understanding of the theoretical and practical aspects of the database technologies;
3. Understand the need for distributed database technology to tackle deficiencies of centralized database systems;
4. Introduce basic principles and implementation techniques of distributed database systems;
5. Expose active and emerging issues in distributed database systems and application development;
6. Apply theory to practice by building a data center in a distributed context.

Course Content – Theory

❖ Historroic development of data managment technologies

❖ Distributed DataBase Systems (DDBS)

- ◆ Architecture
- ◆ Design (fragmentation and allocation)
- ◆ Query processing and optimization
- ◆ Transaction management and concurrecy control
- ◆ Failure recovery and reliability

❖ State-of-Art big data management

- ◆ SQL→ NoSQL → NewSQL
- ◆ parallel and streaming data management
- ◆ data warehousing and OLAP

Site 1

Site 2

Site 3

3
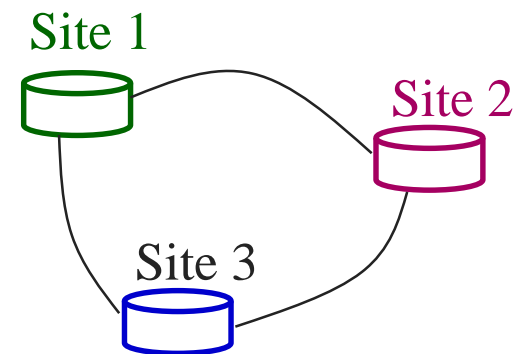
Course Topics

❖ Chapter 0: Overview

❖ Chapter 1: Introduction

❖ Chapter 2: Distributed DBMS Architecture

- ◆ Data independence (logical/physical)

- ◆ Transparency (distribution/fragmentation/replication)
  (the major goals of DDBMS)

- ◆ ANSI/SPARC 3-level architecture (internal/conceptual/external views)

- ◆ Components of DDBMS

- ◆ User Processor at local site, plus Data Processor at remote site

- ◆ Global directory

Course Topics (*cont.*)

❖ Chapter 3: Distributed DB Design

  ◆ DDB design = Data fragmentation and allocation

  ◆ Why and How

  ◆ Correctness of fragmentation (completeness, reconstruction, disjointness)

❖ Chapter 6/7: Overview of Query Processing

  ◆ Problem

  ◆ Objective

  ◆ Complexity, characterization

  ◆ Layers of query processing

Course Topics (*cont.*)

❖ Chapter 8: Optimization of Distributed Queries

  ◆ Cost model

  ◆ Centralized query optimization

    – INGRES

    – System R

  ◆ Distributed query optimization

    – Distributed INGRES

    – System R*

Course Topics (*cont.*)

❖ Chapter 10: Introduction to Transaction Management

- ◆ Properties of transactions: ACID

- ◆ Formalization – partial order, or DAG

- ◆ Termination of transactions

Course Topics (*cont.*)

❖ Chapter 11: Distributed Concurrency Control

 ◆ Serializability theory

 ◆ Locking-based algorithms

   – Basic

   – 2PL

   – Strict 2PL

 ◆ Timestamp-based algorithms

   – Basic

   – Conservative

   – Extremely conservative

 ◆ Optimistic versus pessimistic

Course Topics (*cont.*)

❖ Chapter 12: Distributed DBMS Reliability

  ◆ Reliability and types of failures

  ◆ Local recovery protocols

    – Architecture and log file

    – Execution of LRM commands

    – Checkpoint

  ◆ Distributed reliability protocols

    – 2PC protocol

    – Termination protocols

    – Recovery protocols

    – 3PC – an non-blocking protocols

Course Topics (*cont.*)

❖ Chapter 13: Parallel Database Systems

- ◆ Parallel architectures
- ◆ Parallel DBMS techniques

❖ Chapter 18: Streaming Data Management

- ◆ Challenges
- ◆ Architecture
- ◆ Query Processing

❖ Data Warehouse and OLAP

Question & Answer