Distributed DBMS: Conceptual Data Model Issues

- Due to heterogeneity and autonomy,
  - Multiple types of schemas may exist to describe the same dataset

- Motivation: Query data across all these databases
  - Leads to Schema integration problem
  - Integrate multiple schemas into a single schema

- Considered a very HARD (e.g. undecidable?) computational problem
  - Challenges come from synonyms, homonyms,
  - Diversity of meanings for a word in natural language
  - Consensus building (e.g. XML, semantic Web) is needed

- Best explained using an Exercise
Conceptual Data Model: Schema Integration

**Exercise:** Consider Employee, Engineering and Project databases. Develop an integrated schema to capture information across all three databases.

**HRD Human Resource DB:** CODASYL/XML

- Department (dept-name, budget, manager)
- Employee (e#, name, address, title, salary)
- Department Employs Employee (1:N relationship)

**Engineering– Relational DB**

- E(eno, ename, title)
- J(jno, jname, budget, loc, cname)
- G(eno, jno, resp, dur)
- S(title, sal)

**Projects Database:** Entity Relationship Model

**Entities:**

- Engineer (Engineer No, name, title, salary)
- Project (PNo, project name, budget, location)
- Client (Client Name, Address)

**Relationships:**

- Engineer Works_In Project: (Responsibility, Duration)
- Project Contract_By Client: (Contract Date)

**Hint:** First, find the common entities and relationships between schemas. Then, draw a conceptual diagram for the common entities and relationships.
Conceptual Data Management: Schema Integration

- It may be seen that there are multiple solutions to this problem
- No single correct solution to this exercise

A candidate solution:
- HRD.Employee, Eng.E, Eng.S and Project.Engineer may be overlapping entities, which may be merged as entity-types salary and Employee (with a sub-Entity-type Engineer).
- Eng.J and Project.Project may be similar and merged into an entity-type Project.
- Eng.G and Project.Works_In may be similar relationships and maybe merged
- HRD.department, Project.Client may stay as entity-types
- HRD.Employs and Project.Contracted_by may stay as relationships
Case Study: Delta's Full Integration of Northwest Gets Final Clearance (WSJ, Jan. 2010). (online.wsj.com/article/SB10001424052748704789404574636351284060282.html)

...In the coming weeks, pilots from Delta and Northwest are scheduled to share cockpits for the first time. By the end of March, Delta plans to replace separate ticketing systems with a single one as Northwest's flight codes are retired and visitors to Northwest's Web site get redirected to Delta.

But plenty still can go wrong. When US Airways tried to combine its predecessors' reservation systems in early 2007, more than a year after the initial merger, some check-in kiosks crashed. It also struggled with flight delays.

"We continue to take a 'show me' approach" to Delta and Northwest's merger benefits, William Greene, a Morgan Stanley airline analyst, wrote in a research note last month.

...

Delta and Northwest announced their merger in April 2008. They immediately began planning for what turned out to be an 18-month sprint to integrate 1,200 systems across the two airlines — everything from customer loyalty programs to aircraft operations, all without interrupting service. Managers built this master guide to break down when these systems would need to start working together. Each note represents a project that could involve thousands of tasks. (www.nytimes.com/2011/05/19/business/19air.html?gwh=3CEBE68095B9F170E0A3AB81582B83F0)
**Organization Merger: Family Tree of Delta** (deltamuseum.org/M_Education_DeltaHistory_Facts_Family_Tree.htm).

Imagine database schema integration efforts along the mergers of various organizations the family tree!