Data Management Technologies

Dr. Seng-cho Chou
Professor
Department of Information Management
National Taiwan University

Feb. 2003
Course Objective

- Give an overview of the development of major database technologies from relational database to the most current challenges
- Focus on what they are, where they can be applied, and what their impacts are to the business world
- Cover concepts and principles, platforms and tools, and application modeling ideas

Reading:
- Papers and class handouts
## Schedule

<table>
<thead>
<tr>
<th></th>
<th>Topic</th>
<th>Reading/ pres.</th>
<th>Assignment</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DB overview</td>
<td>E&amp;N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>XML/DB (and WS)</td>
<td>Some WS/XML ex.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Ontology</td>
<td></td>
<td>1st report (ppt)</td>
<td>Sg to share</td>
</tr>
<tr>
<td>4</td>
<td>Wf, KM, PM, and Adv appl modeling and data mgmt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DW, DM</td>
<td></td>
<td>2nd report (doc)</td>
<td>Sg to study</td>
</tr>
<tr>
<td>6</td>
<td>NN, GA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Bioinformatics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8/9</td>
<td>Presentation and/or exam</td>
<td></td>
<td>Group project</td>
<td>(ppt &amp; code)</td>
</tr>
</tbody>
</table>
International Conference on Service Oriented Computing
http://icsoc03.dit.unitn.it/

1st International Workshop on XML Schema and Data Management
www.umr.edu/~madrias/XSDM-workshop'03.html
Historic Development of Data/Database Management Technology

pre-relational (fms, network, hierarchical; programmatic & pragmatic)
Codd’s vision, System R, INGRES, Chen’s ER model

relational db (theoretic but performance)
Oracle, Sybase (C/S), DB2, Informix
Dbase, Clipper, Foxpro, Access, SQL server

post-relational (semantic data model, extended relational, rule-based/active db)
object-oriented db
temporal & spatial, multimedia
distributed db, mdb
Web database management
workflow/doc mgmt/knowledge management
data warehouse & data mining
AI & financial engineering applications
e-commerce applications & mobile data management
XML, semantic Web, Web services

1980 database tech for mainly bus data processing
1990 data mgmt tech for bus apps
2000
Relational Database and its Evolution

- file system (missing interrelationship among files; missing data independence; missing query language)
- hierarchical and network model (navigational)
- relational (3-tier architecture for data independence; simple and sound data model; powerful declarative query language; ad hoc queries with a 4GL; efficient transaction processing)
RDB data model & query language

- Sound mathematical foundation – “relation”
- Based on “tables” with rows and columns
- Closure property of relational operators – manipulate with tables
- A set of simple operators – select, project, join, union, difference
- SQL – a powerful declarative query language
  - Example: find all high pay employees (>1.5M)
  - Select empl_id, name from employee where salary > 1500000
### Customer Table

<table>
<thead>
<tr>
<th>cust_ID</th>
<th>name</th>
<th>address</th>
<th>age</th>
<th>income</th>
<th>credit_info</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Smith, Sandy</td>
<td>5463 E Hastings, Burnaby, BC V5A 4S9, Canada</td>
<td>21</td>
<td>$27000</td>
<td>1</td>
</tr>
</tbody>
</table>

### Item Table

<table>
<thead>
<tr>
<th>item_ID</th>
<th>name</th>
<th>brand</th>
<th>category</th>
<th>type</th>
<th>price</th>
<th>place_made</th>
<th>supplier</th>
<th>cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>high-res-TV</td>
<td>Toshiba</td>
<td>high resolution</td>
<td>TV</td>
<td>$988.00</td>
<td>Japan</td>
<td>NikoX</td>
<td>$600.00</td>
</tr>
<tr>
<td>18</td>
<td>multidisc-CDplay</td>
<td>Sanyo</td>
<td>multidisc</td>
<td>CD player</td>
<td>$369.00</td>
<td>Japan</td>
<td>MusicFront</td>
<td>$120.00</td>
</tr>
</tbody>
</table>

### Employee Table

<table>
<thead>
<tr>
<th>empl_ID</th>
<th>name</th>
<th>category</th>
<th>group</th>
<th>salary</th>
<th>commission</th>
</tr>
</thead>
<tbody>
<tr>
<td>E55</td>
<td>Jones, Jane</td>
<td>home entertainment</td>
<td>manager</td>
<td>$18,000</td>
<td>2%</td>
</tr>
</tbody>
</table>

### Branch Table

<table>
<thead>
<tr>
<th>branch_ID</th>
<th>name</th>
<th>address</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>City Square</td>
<td>369 Cambie St., Vancouver, BC V5L 3A2, Canada</td>
</tr>
</tbody>
</table>

### Purchases Table

<table>
<thead>
<tr>
<th>trans_ID</th>
<th>cust_ID</th>
<th>empl_ID</th>
<th>date</th>
<th>time</th>
<th>method_paid</th>
<th>amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>T100</td>
<td>C1</td>
<td>E55</td>
<td>09/21/98</td>
<td>15:45</td>
<td>Visa</td>
<td>$1357.00</td>
</tr>
</tbody>
</table>

### Items Sold Table

<table>
<thead>
<tr>
<th>trans_ID</th>
<th>item_ID</th>
<th>qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>T100</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>T100</td>
<td>18</td>
<td>2</td>
</tr>
</tbody>
</table>

### Works At Table

<table>
<thead>
<tr>
<th>empl_ID</th>
<th>branch_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>E55</td>
<td>B1</td>
</tr>
</tbody>
</table>
Transaction management

- Transaction and its properties – no need to worry about partial/incomplete transactions
  - Atomic
  - Consistent
  - Isolation
  - Durable
- Concurrency control – handle multiple simultaneous transactions
- Recovery
- Efficient OLTP – OnLine Transaction Processing
Key players

- Codd’s vision, System R, INGRES
- Chen’s ER model (cf. Chen’s ppt)
- Oracle, Sybase (C/S), DB2, Informix
- Dbase, Clipper, Foxpro, Access
- SQL server
RDB in a nutshell

- all good for commercial data processing (in particular after performance issues were under control)
- the main technology behind today's business information systems
- full of precious data for today's business (DM or treasure hunt)
- post-relational (the need for support for more complex applications)
Post-relational DB applications

- semantic data model, extended relational, rule-based/active db – supports advanced database features, including integrity constraint enforcement, derived data maintenance, triggers, alerts, protection, version control, and others
- object-oriented db – IC design
- temporal & spatial, multimedia – GIS
- distributed db, multi-db – heterogeneous databases
- Web database management – web as a database
- workflow/doc mgmt/knowledge management
- data warehouse & data mining
- AI & financial engineering applications – time-series appl
- e-commerce applications & mobile data management
- XML, semantic Web, Web services