Database Overview

Yufei Tao

Department of Computer Science and Engineering
Chinese University of Hong Kong
A very high-level definition

**A database** is a system that *manages* data, where “management” means

1. data storage
2. support of operations on data.
A database example

Microsoft Excel

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>age</th>
<th>city</th>
<th>state</th>
<th>salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adam</td>
<td>38</td>
<td>New York</td>
<td>NY</td>
<td>4500</td>
</tr>
<tr>
<td>2</td>
<td>Bob</td>
<td>55</td>
<td>Pittsburg</td>
<td>PA</td>
<td>3000</td>
</tr>
<tr>
<td>3</td>
<td>Charlie</td>
<td>25</td>
<td>New York</td>
<td>NY</td>
<td>2800</td>
</tr>
<tr>
<td>4</td>
<td>Dorothy</td>
<td>42</td>
<td>San Fransisco</td>
<td>CA</td>
<td>5300</td>
</tr>
<tr>
<td>5</td>
<td>Emily</td>
<td>32</td>
<td>Pittsburgh</td>
<td>PA</td>
<td>2100</td>
</tr>
<tr>
<td>6</td>
<td>Frank</td>
<td>37</td>
<td>Los Angeles</td>
<td>CA</td>
<td>4800</td>
</tr>
</tbody>
</table>

Supported operations:

- Find the record of Dorothy.
- Find the average salary of Pittsburgh.
- For each state, report its average salary.
- ...
Limitations of Excel:

1. Has a low limit on the data volume.
2. Offers weak error detection capability.
3. Slow data retrieval.
4. Supports only relatively simple operations.
5. Data redundancy (can you find redundancy in the table of the previous slide?).
6. No multi-level access control.
7. Cannot run as a server (that reacts to an incoming query any time).
8. ...

Database Overview
A professional database should:

1. Be able to handle colossal datasets.
2. Enforce strong integrity constraints.
3. Have very efficient mechanisms for query processing.
4. Support a huge variety of queries.
5. Minimize data redundancy.
6. Provide fine-grained multi-level access control.
7. Be a powerful server that can even run long running programs simultaneously.
8. ...