

# 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

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

...

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 ...

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 ...