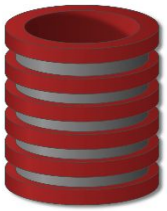
Data Modelling and Databases

Jooyoung Lee

http://www.dainfos.com

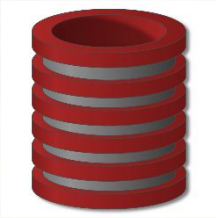
Slides are adopted from Jennifer Widom @ Stanford University



Intro to Databases

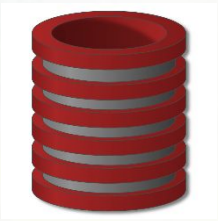
Database Management System (DBMS) provides….

… efficient, reliable, convenient, and safe multi-user storage of and access to massive amounts of persistent data.



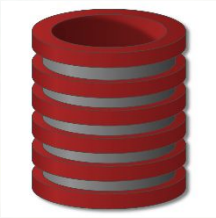
Intro to Databases

- Massive
- Persistent
- Safe
- Multi-user
- Convenient
- Efficient
- Reliable



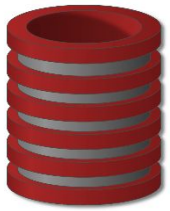
Intro to Databases

- Database applications may be programmed via “frameworks”
- DBMS may run in conjunction with “middleware”
- Data-intensive applications may not use DBMS at all



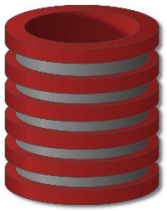
Key concepts

- Data model
- Schema versus data
- Data definition language (DDL)
- Data manipulation or query language (DML)



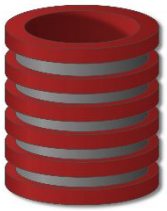
Key people

- DBMS implementer
- Database designer
- Database application developer
- Database administrator



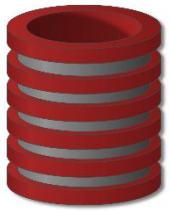
The Relational Model

- Used by all major commercial database systems
- Very simple model
- Query with high-level languages: simple yet expressive
- Efficient implementations



The Relational Model

Schema = structural description of relations in database
Instance = actual contents at given point in time



The Relational Model

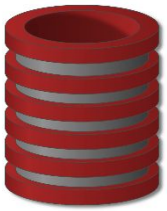
Database = set of named **relations** (or **tables**)
Each relation has a set of named **attributes** (or **columns**)
Each **tuple** (or **row**) has a value for each attribute
Each attribute has a **type** (or **domain**)

Student

ID	name	GPA	photo
123	Emil	3.4	😐
142	Artur	3	:+)
521	Damir	NULL	☹

Dorm

name	unit	CAP
dorm1	205	4
dorm2	205	5
dorm1	403	4



The Relational Model

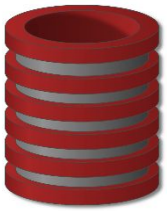
Schema – structural description of relations in database
Instance – actual contents at given point in time

Student

ID	name	GPA	photo
123	Emil	3.4	😐
142	Artur	3	:+)
521	Damir	NULL	☹

Dorm

name	unit	CAP
dorm1	205	4
dorm2	205	5
dorm1	403	4



The Relational Model

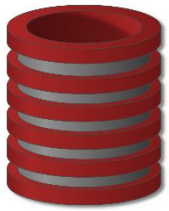
NULL – special value for “unknown” or “undefined”

Student

ID	name	GPA	photo
123	Emil	3.4	😐
142	Artur	3	:+)
521	Damir	NULL	☹

Dorm

name	unit	CAP
dorm1	205	4
dorm2	205	5
dorm1	403	4



The Relational Model

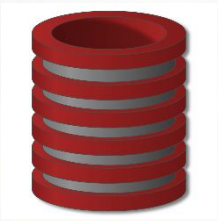
Key – attribute whose value is unique in each tuple
Or set of attributes whose combined values are unique

Student

ID	name	GPA	photo
123	Emil	3.4	😐
142	Artur	3	:+)
521	Damir	NULL	☹

Dorm

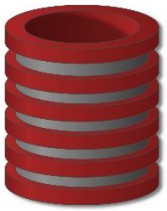
name	unit	CAP
dorm1	205	4
dorm2	205	5
dorm1	403	4



Creating relations (tables) in SQL

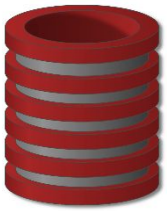
```
Create Table Student(ID, name, GPA, photo)

Create Table Dorm
  (name string, unit char(3), CAP integer)
```



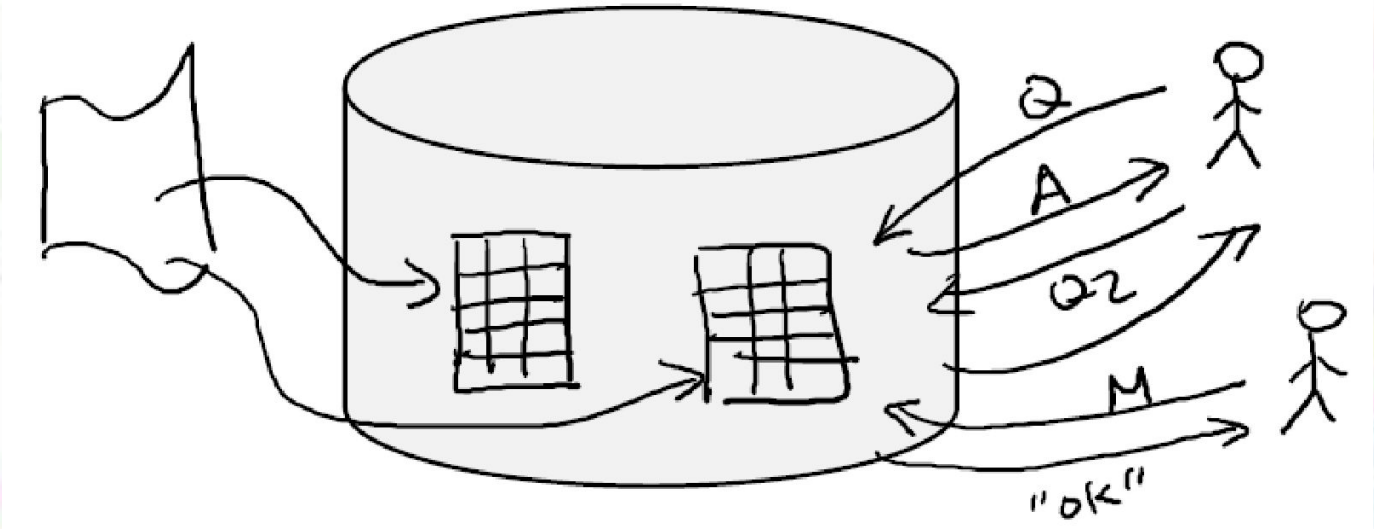
The Relational Model

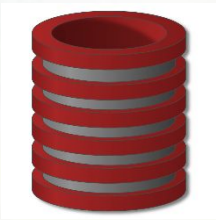
- Used by all major commercial database systems
- Very simple model
- Query with high-level languages: simple yet expressive
- Efficient implementations



Steps in creating and using a (relational) database

1. Design schema; create using DDL
2. “Bulk load” initial data
3. Repeat: execute queries and modifications



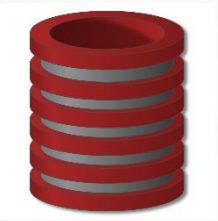


Querying Relational Databases

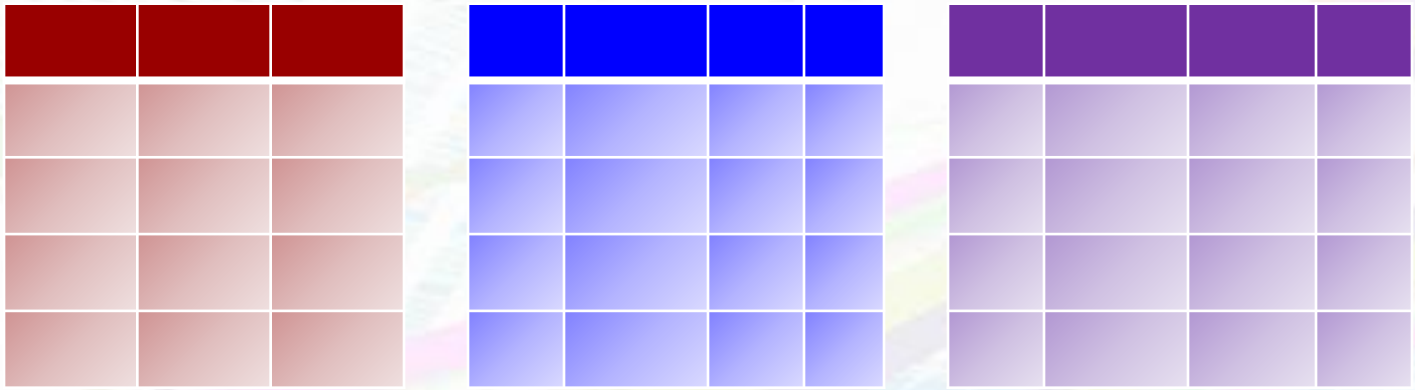
Ad-hoc queries in high-level language

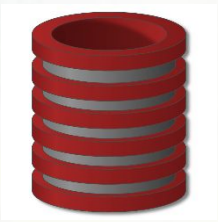
- *All students with GPA > 3.7 applying to Stanford and MIT only*
- *All engineering departments in CA with < 500 applicants*
- *College with highest average accept rate over last 5 years*

- Some easy to pose; some a bit harder
- Some easy for DBMS to execute efficiently; some harder
- "Query language" also used to modify data



Queries return relations (“compositional”, “closed”)





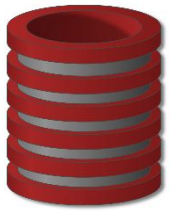
Query Languages

- Relational Algebra

- SQL

```
Select Student.ID
From Student, Apply
Where Student.ID=Apply.ID
And GPA>3.7 and college='Stanford'
```

IDs of students with GPA > 3.7 applying to Stanford



Assignment 1

- Write one page essay in latex [sharelatex.com] that includes the followings:
- Your name and email.
- Your short bio.
- Categorize databases based on your opinion by using any search engine.
- Cite all the sources you use.
- No copy-paste.

Whether you know it or not,
you're using a database every day