Lecture 1 Introduction

Introduction to Database

Outline

- Understand the reasons of using a database
- Know the topics covered in this class
- Know the preliminary of database
- Know the data model and data schema

Purpose of database

CS211 – Introduction to Database

To store data using file system

- University info
  - Instructors, students, departments, course offerings
- File Processing System: store data in files
- What to store and how to store?
- Single file
  - Store all the information in a single list
  - Structure of a node: Instructor ID, instructor Name, studentID, studentName, departmentID for instructor, departmentID for student …
- Multiple files
  - Instructor file, student file, department file, course file, course enrollment file
  - Instructor file: Instructor ID, instructor Name, department ID …

Structure of data

To store data using file system

- How to interact with the data?
  - Read, write, update
- Applications are needed
  - Query: return the average scores of courses
  - Add new students, instructors and courses
  - Course enrollment for students
  - Generate class rosters
  - Generate transcripts
  - …

Potential problems with file system

- Considering the application code for
  - Adding a new student
    - Input the student info and write it in the file or
    - Checking duplicate first to avoid repeat writing
  - Update the course name “DB” as “database”
    - Just update “DB” in course file or
    - Search all files containing course name DB and update it

Student file

0001	Amy Choo	20	CS
0002	John Yeo	21	CS
0002	John Yeo	21	CS

Course file

cs211	~~DB~~ database
cs385	Machine learning

Course enrolment file

0001	cs211	~~DB~~ database
0002	cs385	Machine learning
0001	cs385	Machine learning

Potential problems with file system

- Considering the application code for
  - Deleting a student
    - Just delete the student from student file or
    - Search all files containing the student and delete its related information

Student file

0001	Amy Choo	20	CS
~~0002~~	~~John Yeo~~	~~21~~	~~CS~~

Course enrolment file

0001	cs211	DB
~~0002~~	~~cs385~~	~~Machine learning~~
0001	cs385	Machine learning

Potential problems with file system

- Considering the application code for
  - Displaying students and enrolled class
    - Display all the data from enrollment file, but
    - Student name is missing

Course enrolment file

0001	cs211	DB
0002	cs385	Machine learning
0001	cs385	Machine learning

Difficulty in accessing data

- Solution 1, coding for searching in student file
- E.g. get 0001 from course enrollment file,
- search the name of student with ID 0001
- return the name Amy Choo

Student file

0001	Amy Choo	20	CS
0002	John Yeo	21	CS

Potential problems with file system

- Considering the application code for
  - Displaying students and enrolled class
    - Display all the data from enrollment file, but
    - Student name is missing
    - Solution 2, add student name into the course enrollment file

Course enrollment file

0001 cs211 DB
0002 cs385 Machine learning
0001 cs385 Machine learning

Course enrollment file

0001 Amy Choo cs211 DB
0002 John Yeo cs385 Machine learning
0001 Amy Choo cs385 Machine learning

Redundancy,
Keeping Integrity

Student file

0001 Amy Choo 20 CS
0002 John Yeo 21 CS

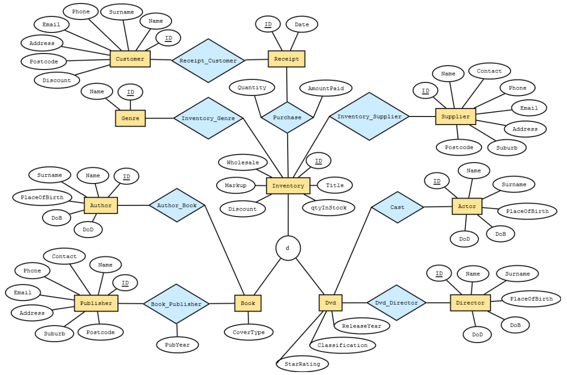
Purpose of database

- To store data
- To provide organizational structure for data
- To provide a mechanism for querying, creating, modifying and deleting data
- To provide functionalities such as keeping consistency, easily accessing data, keeping integrity, supporting concurrent-accessing, security etc.…

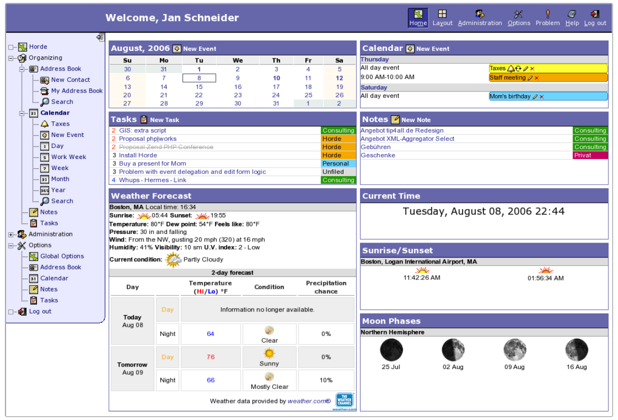
Topics covered in this class

CS211 – Introduction to Database

Data Modeling



Database Design



Welcome, Jan Schneider

August, 2006 New Event

Su	Mo	Tu	We	Th	Fr	Sa
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2

Tasks New Task

2	GIS: extra script	Consulting
2	Proposal php|works	Horde
2	~~Proposal Zend PHP Conference~~	Horde
3	Install Horde	Horde
3	Buy a present for Mom	Personal
3	Problem with event delegation and edit form logic	Unfiled
4	Whups - Hermes - Link	Consulting

Weather Forecast

Boston, MA Local time: 16:34
Sunrise: 05:44 **Sunset:** 19:55
Temperature: 80°F **Dew point:** 54°F **Feels like:** 80°F
Pressure: 30 in and falling
Wind: From the NW, gusting 20 mph (320) at 16 mph
Humidity: 41% **Visibility:** 10 sm **U.V. index:** 2 - Low
Current condition: Partly Cloudy

2-day forecast

Day		Temperature (Hi/Lo) °F	Condition	Precipitation chance
Today Aug 08	Day	Information no longer available.		
	Night	64	Clear	0%
Tomorrow Aug 09	Day	76	Sunny	0%
	Night	66	Mostly Clear	10%

Weather data provided by weather.com®

Calendar New Event

Thursday	
All day event	Taxes
9:00 AM-10:00 AM	Staff meeting
Saturday	
All day event	Mom's birthday

Notes New Note

Angebot tip4all.de Redesign	Consulting
Angebot XML-Aggregator Select	Consulting
Gebühren	Consulting
Geschenke	Privat

Current Time

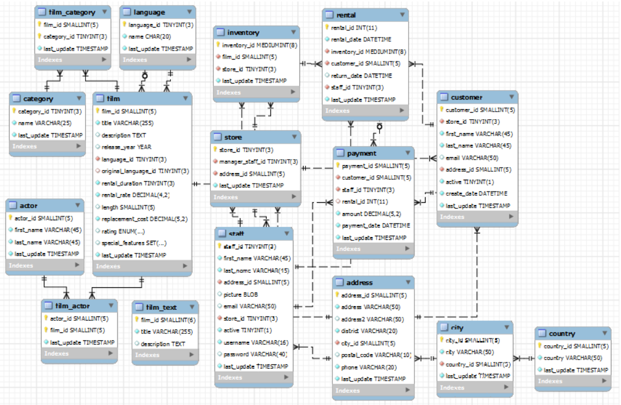
Tuesday, August 08, 2006 22:44

Sunrise/Sunset

Boston, Logan International Airport, MA
11:42:26 AM 01:56:34 AM

Moon Phases

Northern Hemisphere
25 Jul 02 Aug 09 Aug 16 Aug



film_category
film_id SMALLINT(5)
category_id TINYINT(3)
last_update TIMESTAMP
Indexes

language
language_id TINYINT(3)
name CHAR(20)
last_update TIMESTAMP
Indexes

inventory
inventory_id MEDIUMINT(8)
film_id SMALLINT(5)
store_id TINYINT(3)
last_update TIMESTAMP
Indexes

rental
rental_id INT(11)
rental_date DATETIME
inventory_id MEDIUMINT(8)
customer_id SMALLINT(5)
return_date DATETIME
staff_id TINYINT(3)
last_update TIMESTAMP
Indexes

category
category_id TINYINT(3)
name VARCHAR(25)
last_update TIMESTAMP
Indexes

film
film_id SMALLINT(5)
title VARCHAR(255)
description TEXT
release_year YEAR
language_id TINYINT(3)
original_language_id TINYINT(3)
rental_duration TINYINT(3)
rental_rate DECIMAL(4,2)
length SMALLINT(5)
replacement_cost DECIMAL(5,2)
rating ENUM(...)
special_features SET(...)
last_update TIMESTAMP
Indexes

store
store_id TINYINT(3)
manager_staff_id TINYINT(3)
address_id SMALLINT(5)
last_update TIMESTAMP
Indexes

payment
payment_id SMALLINT(5)
customer_id SMALLINT(5)
staff_id TINYINT(3)
rental_id INT(11)
amount DECIMAL(5,2)
payment_date DATETIME
last_update TIMESTAMP
Indexes

customer
customer_id SMALLINT(5)
store_id TINYINT(3)
first_name VARCHAR(45)
last_name VARCHAR(45)
email VARCHAR(50)
address_id SMALLINT(5)
active TINYINT(1)
create_date DATETIME
last_update TIMESTAMP
Indexes

actor
actor_id SMALLINT(5)
first_name VARCHAR(45)
last_name VARCHAR(45)
last_update TIMESTAMP
Indexes

staff
staff_id TINYINT(3)
first_name VARCHAR(45)
last_nome VARCHAR(15)
address_id SMALLINT(5)
picture BLOB
email VARCHAR(50)
store_id TINYINT(3)
active TINYINT(1)
username VARCHAR(16)
password VARCHAR(40)
last_update TIMESTAMP
Indexes

address
address_id SMALLINT(5)
address VARCHAR(50)
address2 VARCHAR(50)
district VARCHAR(20)
city_id SMALLINT(5)
postal_code VARCHAR(10)
phone VARCHAR(20)
last_update TIMESTAMP
Indexes

film_actor
actor_id SMALLINT(5)
film_id SMALLINT(5)
last_update TIMESTAMP
Indexes

film_text
film_id SMALLINT(6)
title VARCHAR(255)
description TEXT
Indexes

city
city_id SMALLINT(5)
city VARCHAR(50)
country_id SMALLINT(5)
last_update TIMESTAMP
Indexes

country
country_id SMALLINT(5)
country VARCHAR(50)
last_update TIMESTAMP
Indexes

SQL

SYBASE
iAnywhere
MySQL
Tibero RDBMS
ORACLE
Informix SOFTWARE
The database company
Microsoft SQL Server

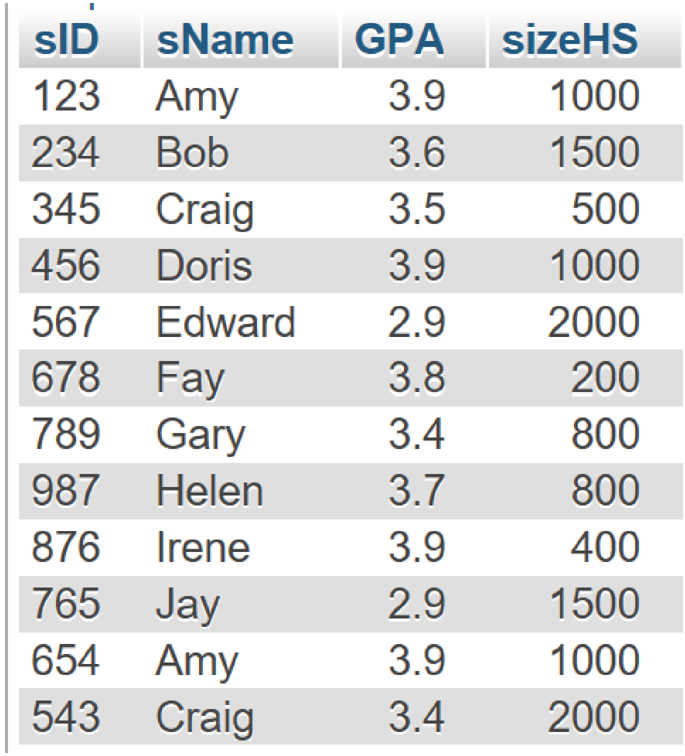
Transaction
Indexing
Concurrency
etc.

No-SQL Database

- What is No-SQL?
  - Not only SQL or Non-SQL
  - NoSQL databases are databases that store data in a format other than relational tables
- Why No-SQL?
  - Huge amounts of data came in all shapes and sizes.
  - It makes data Management using Relational DB become nearly impossible.
  - NoSQL databases allow developers to store huge amounts of unstructured data, giving them a lot of flexibility.

SQL vs No-SQL

- Student in relational databases



sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

- Student in key-value databases

sID:”123”, sName:”Amy”,GAP:3.9 etc.

Topics

- SQL
  - Relational algebra, relational model
  - Basic SQL (create, insert, select, update, delete)
  - Advanced SQL(aggregation, create view)
  - Integrity (create trigger)
  - Embedded SQL

- Data modelling & design
  - ER model
  - Functional dependencies
  - Normal form
- Transaction
  - Transaction
  - Isolation level
  - Indexing
- NoSQL
  - Data Model
  - MongoDB Document
  - Data manipulation and design

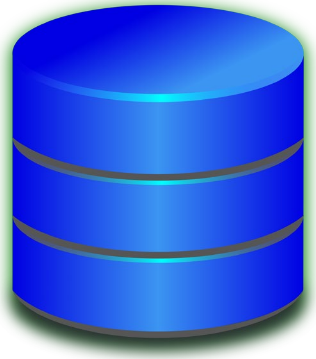
Database – A Preliminary Understanding

CS211 – Introduction to Database

Database

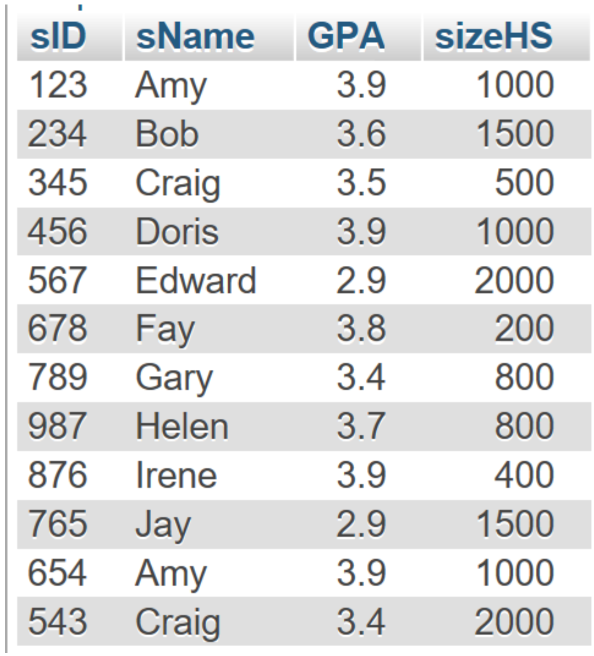
- Everything over DB - Support Query and Analysis

- Database
  - Collection of related data
  - Storage place for data
  - Get information from



Database, table

Students



sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

- Table
  - an arrangement of words, numbers, or signs, or combination of them, as in parallel columns, to exhibits a set of facts or relations in a definite, compact, and comprehensive form

–

Webster's Dictionary of the English Language

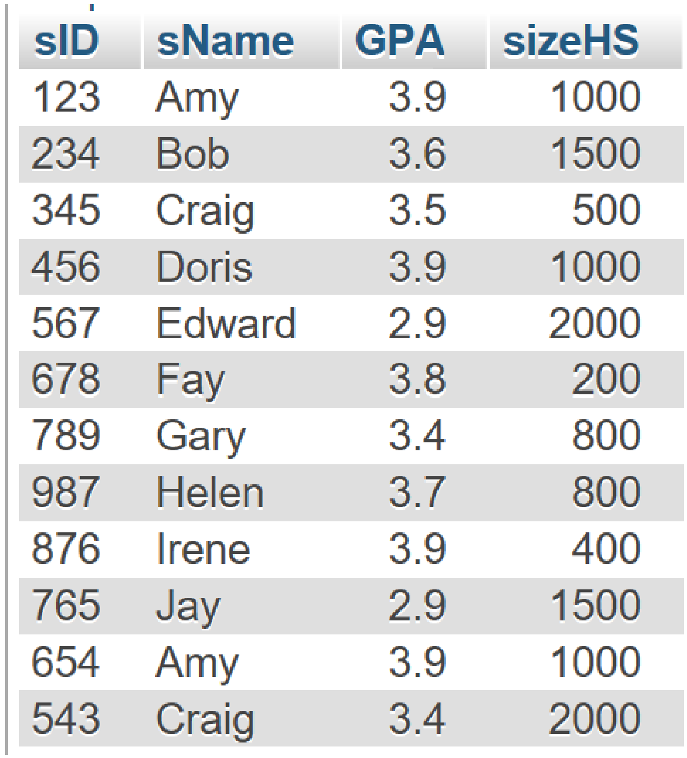
Table and relation

- E.F.Codd, basing on the understanding of table,
  - Proposed: relation, relational model
  - They are the theoretical basis for relational databases and relational database management systems

What’s inside a table?

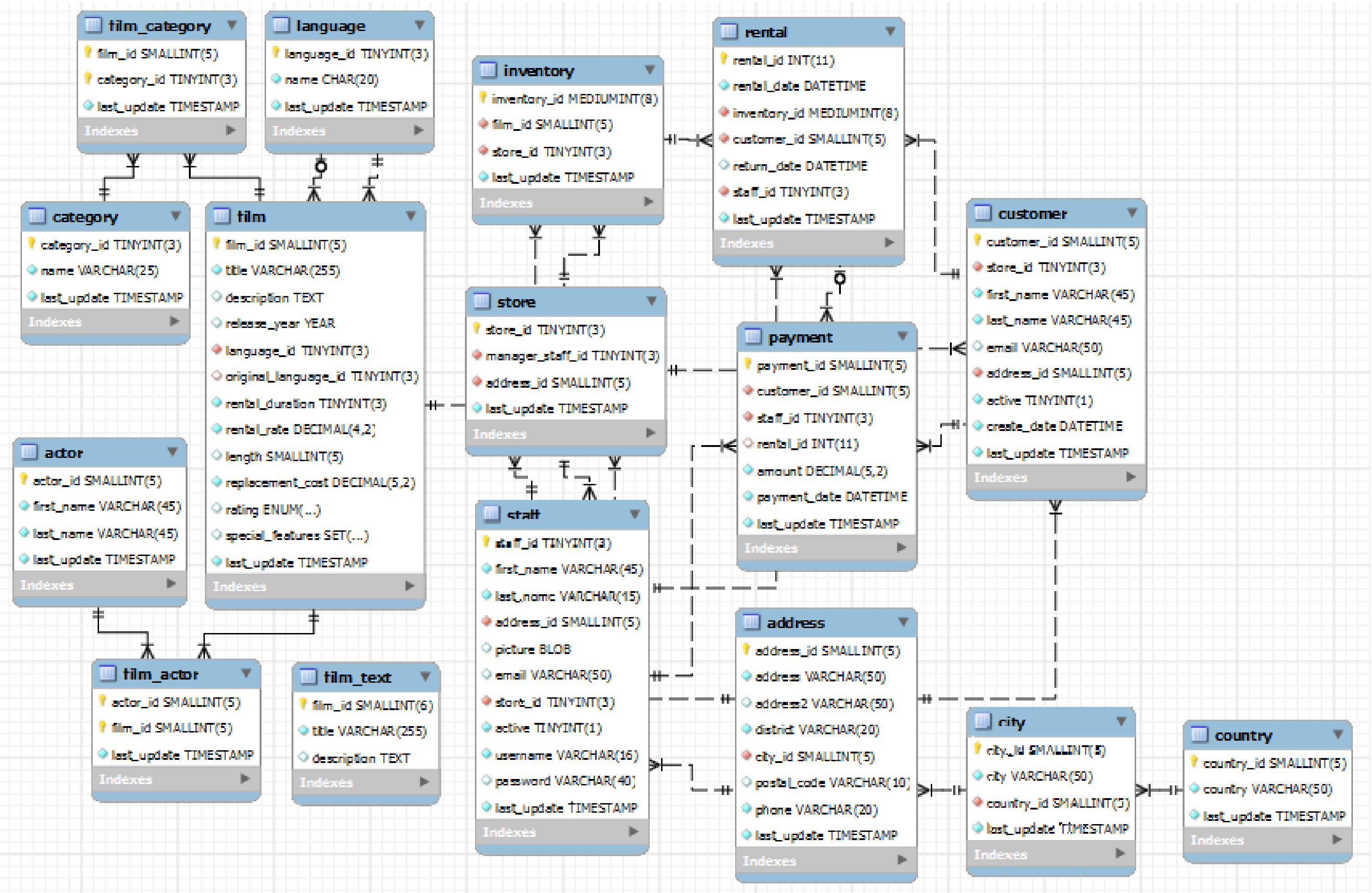
- Table / relation
  - Relational schema
    - Table name
    - Table format
  - Table content
    - row / tuple / record
    - column / field / attribute
      - Column name
      - value

Student

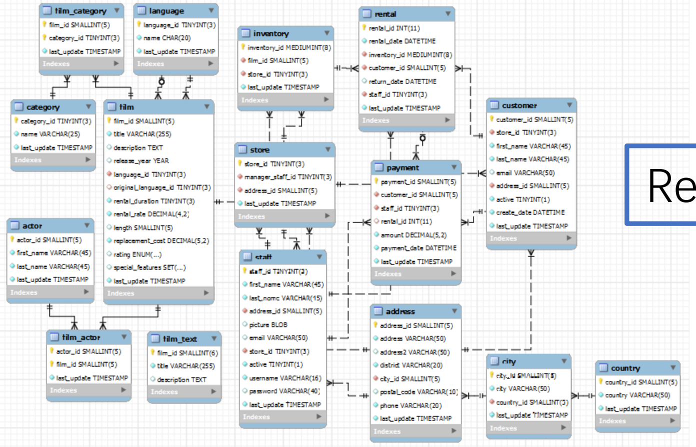


sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

Database: collection of related tables

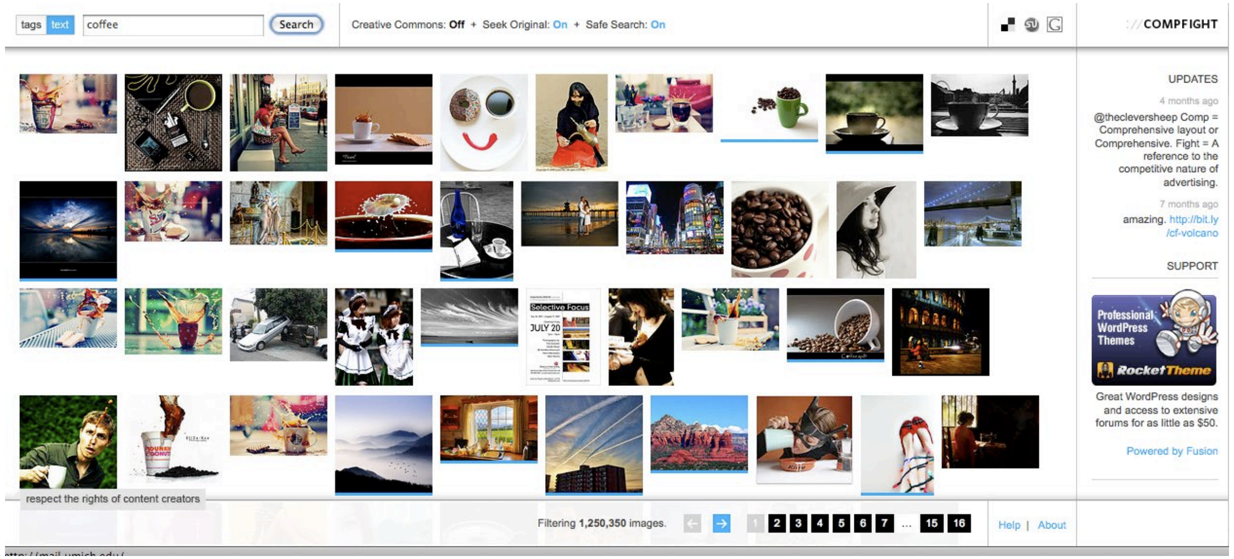


Database: different types



Relational

Image database



tags text coffee Search

Creative Commons: Off • Seek Original: On • Safe Search: On

COMPFIGHT

UPDATES

4 months ago

@thecleversheep Comp = Comprehensive layout or Comprehensive. Fight = A reference to the competitive nature of advertising.

7 months ago

amazing. http://bit.ly/cf-volcano

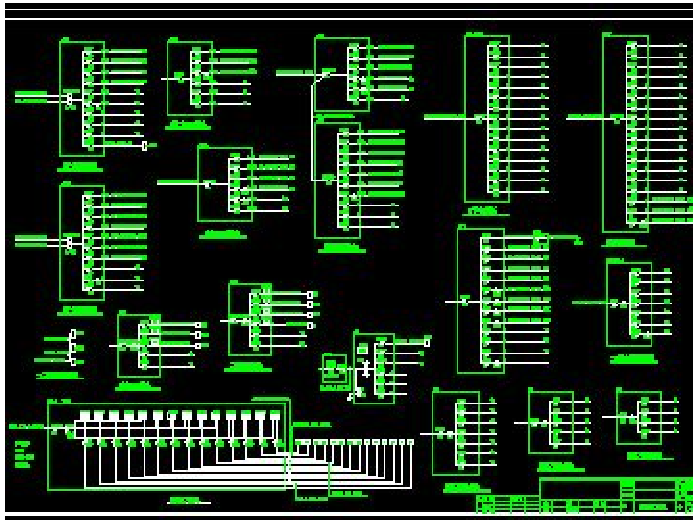
SUPPORT

Great WordPress designs and access to extensive forums for as little as $50.

Powered by Fusion

respect the rights of content creators

Filtering 1,250,350 images. 1 2 3 4 5 6 7 ... 15 16 Help | About

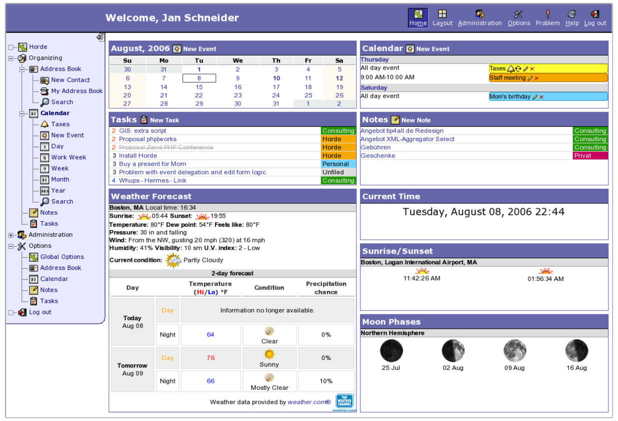


Project database

Database System

- DB: DataBase
- DBMS: DataBase Management System
- DBAP: DataBase APplication
- DBA: DataBase Administrator
- Computer System: computer, network, os, I/O etc.

HADOOP
MySQL
MS SQL
ORACLE
POSTGRE



Welcome, Jan Schneider

August, 2006 New Event

Su	Mo	Tu	We	Th	Fr	Sa
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2

Calendar New Event

Thursday	
All day event	Taxes
9:00 AM-10:00 AM	Staff meeting
Saturday	
All day event	Mom's birthday

Tasks New Task

2	GIS: extra script	Consulting
2	Proposal php|works	Horde
2	~~Proposal Zend PHP Conference~~	Horde
3	Install Horde	Horde
3	Buy a present for Mom	Personal
3	Problem with event delegation and edit form logic	Unfiled
4	Whups - Hermes - Link	Consulting

Notes New Note

Angebot tip4all.de Redesign	Consulting
Angebot XML-Aggregator Select	Consulting
Gebühren	Consulting
Geschenke	Privat

Weather Forecast

Boston, MA Local time: 16:34
Sunrise: 05:44 **Sunset:** 19:55
Temperature: 80°F **Dew point:** 54°F **Feels like:** 80°F
Pressure: 30 in and falling
Wind: From the NW, gusting 20 mph (320) at 16 mph
Humidity: 41% **Visibility:** 10 sm **U.V. index:** 2 - Low
Current condition: Partly Cloudy

2-day forecast

Day		Temperature (Hi/Lo) °F	Condition	Precipitation chance
Today Aug 08	Day	Information no longer available.		
	Night	64	Clear	0%
Tomorrow Aug 09	Day	76	Sunny	0%
	Night	66	Mostly Clear	10%

Weather data provided by weather.com®

Current Time

Tuesday, August 08, 2006 22:44

Sunrise/Sunset

Boston, Logan International Airport, MA

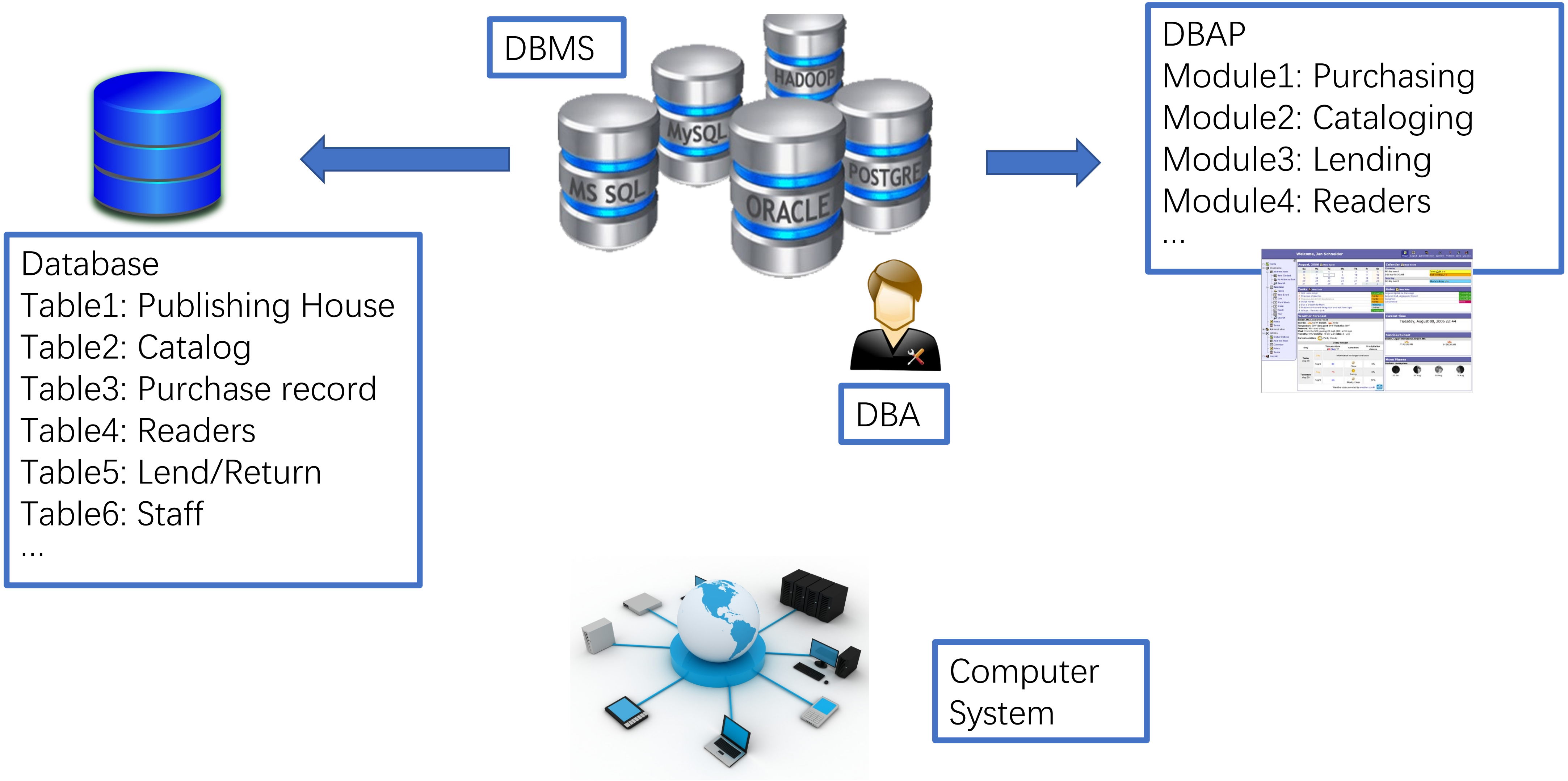
11:42:26 AM 01:56:34 AM

Moon Phases

Northern Hemisphere

25 Jul 02 Aug 09 Aug 16 Aug

Books Management System



DBMS

Database
Table1: Publishing House
Table2: Catalog
Table3: Purchase record
Table4: Readers
Table5: Lend/Return
Table6: Staff
…

DBAP
Module1: Purchasing
Module2: Cataloging
Module3: Lending
Module4: Readers
…

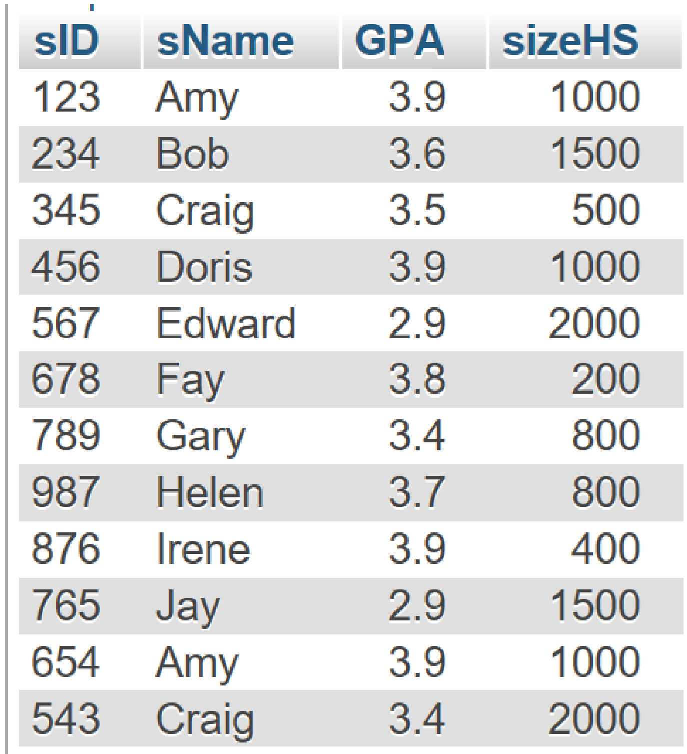
DBA

Computer System

DataBase Management System

- From user's point of view
  - Database definition: schema definition for tables
    - DDL: Data Definition Language
  - Database manipulation: query/add/delete/update data
    - DML: Data Manipulation Language
  - Database access control: Authorization
    - DCL: Data Control Language
    - control access to data stored in a database
  - Database maintenance
    - App: backup/restore/performance monitoring

Student



sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

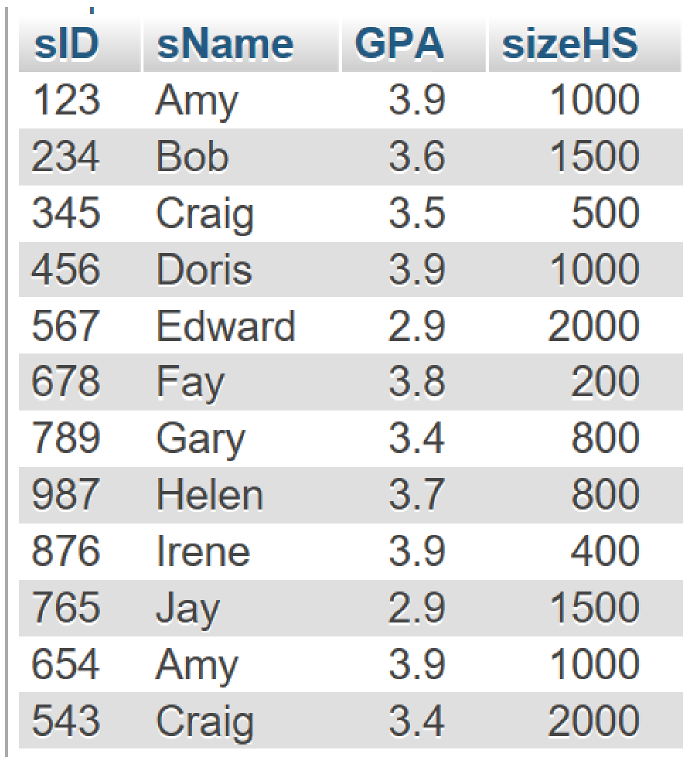
SQL

- SQL vs. High-level programming language

```
select sName from student where GPA>3.4
```

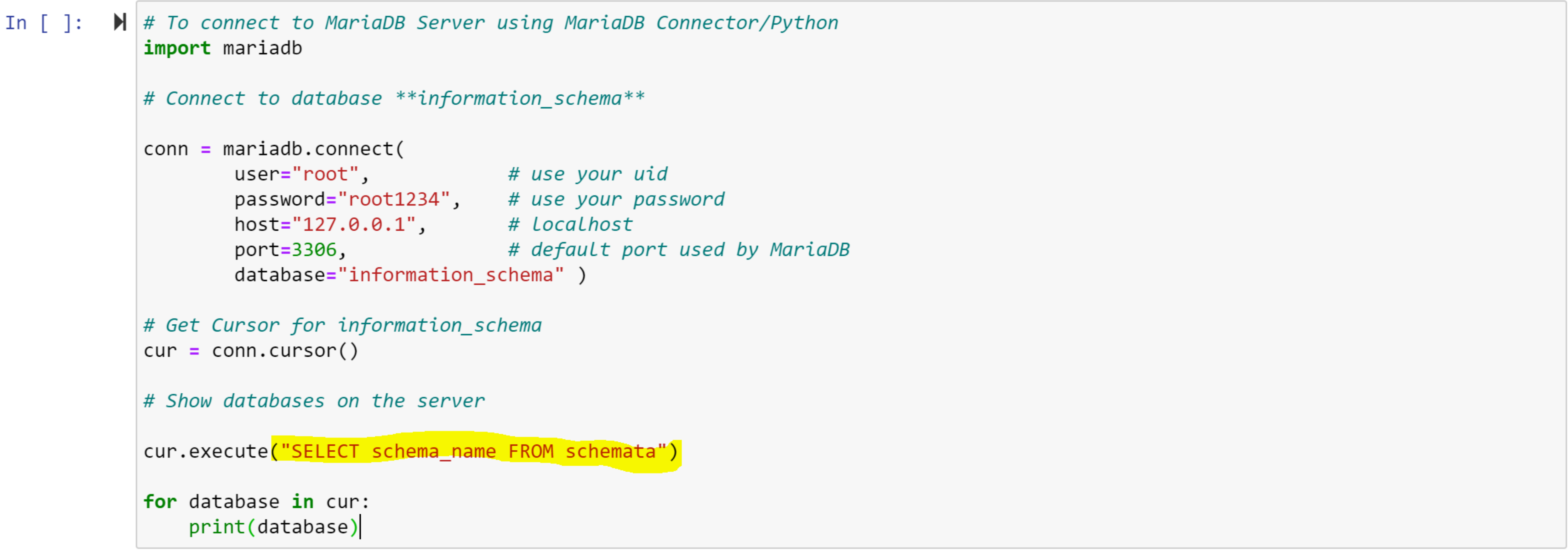
```
for each tuple
  if tuple.GPA > 3.4 then
     printout(tuple.sName)
```

Student



sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

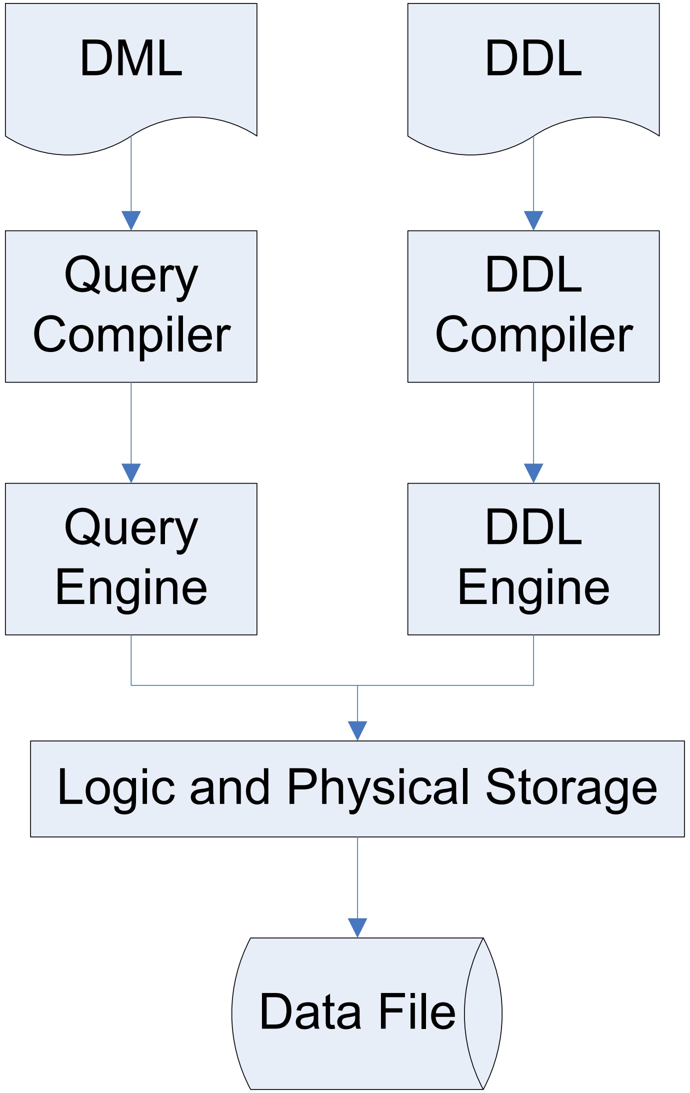
Embedded SQL



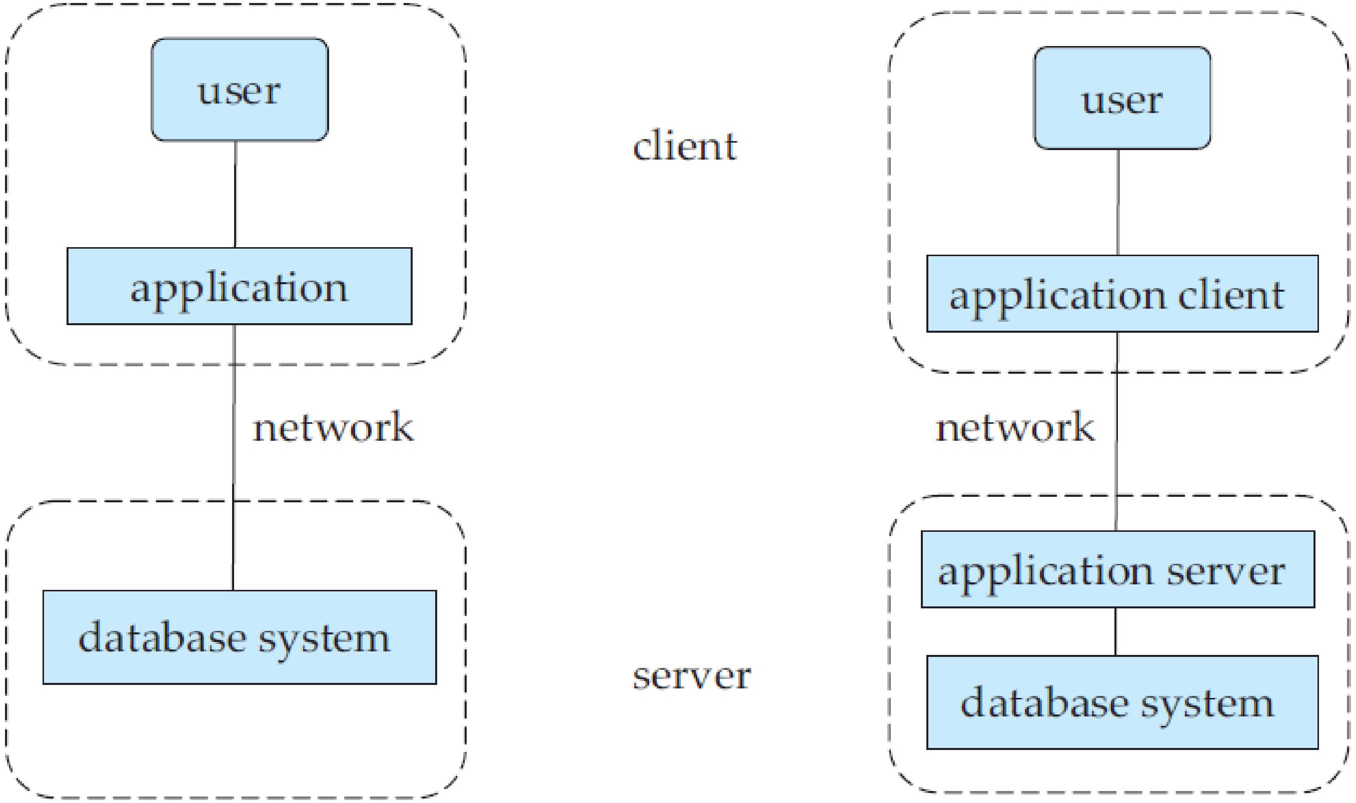
```
# To connect to MariaDB Server using MariaDB Connector/Python
import mariadb

# Connect to database **information_schema**

conn = mariadb.connect(
        user="root",            # use your uid
        password="root1234",    # use your password
        host="127.0.0.1",       # localhost
        port=3306,              # default port used by MariaDB
        database="information_schema" )

# Get Cursor for information_schema
cur = conn.cursor()

# Show databases on the server

cur.execute("SELECT schema_name FROM schemata")

for database in cur:
    print(database)
```

DataBase Management System

- From system’s point of view
  - Compiler
  - Execution Engine: optimization, query algorithm···
  - Data access and indexing
  - Interaction with the OS File System
  - Concurrency control
  - Transaction Management
  - Recovery: backup, log ···
  - Security control
  - Integrity control
  - Data dictionary



Database Architecture



(a) Two-tier architecture

(b) Three-tier architecture

Roles

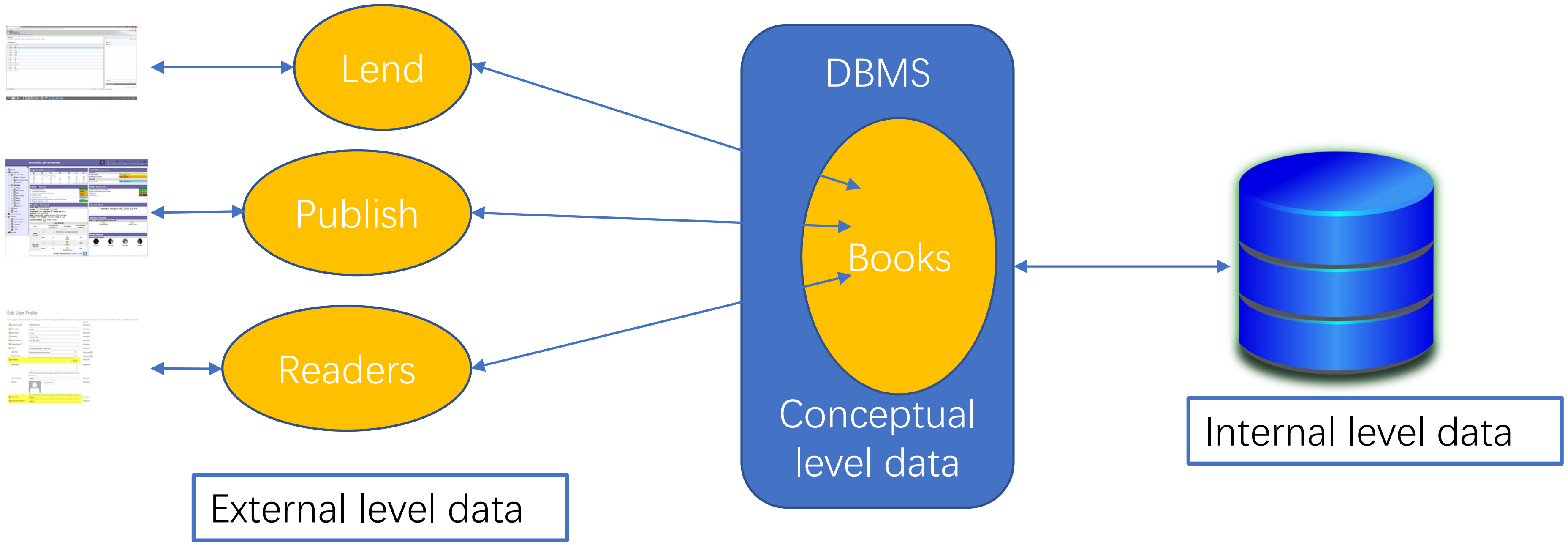
- End-user, Naïve users – Application interface
- Sophisticated users – query language, interactive SQL
- Application programmers: Embedded SQL, Concept about transaction
- Database analyzer and designer: data modeling, Concept about normalization
- Database administrator: Database maintenance, Definition for: Security, Integrity, Recovery, Schema, Storage structure and access-method
- Database management system designer and implementor: implementation technique of above
- Specialized users - write specialized database applications
- Techniques for special and new database management system – e.g. noSQL, XML DB, Multimedia DB

Data Model and Data Schema

CS211 – Introduction to Database

3-Schema Architecture

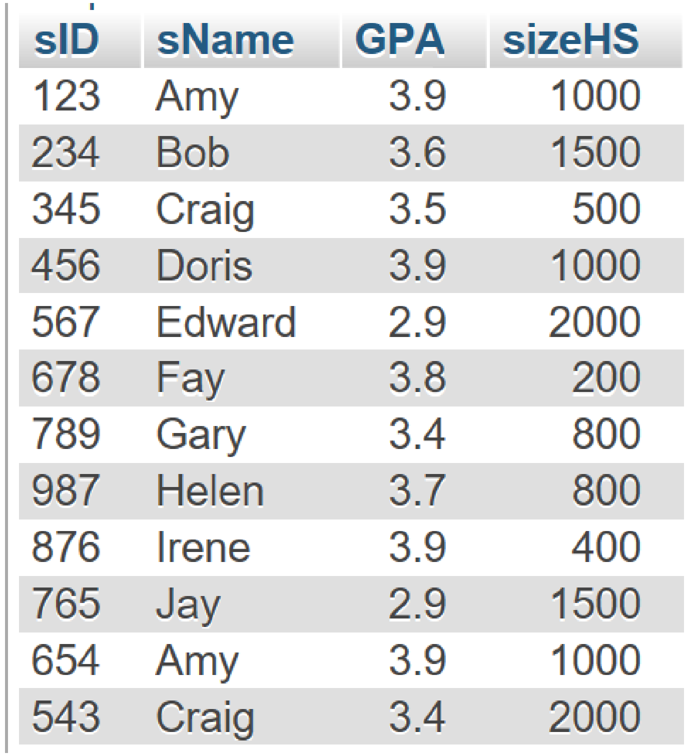
- External level = user level (Local)
- Conceptual level = logic level (Global)
- Internal level = physical level (Index…)



Schema – organization of data

- The **database schema** of a database system is its structure

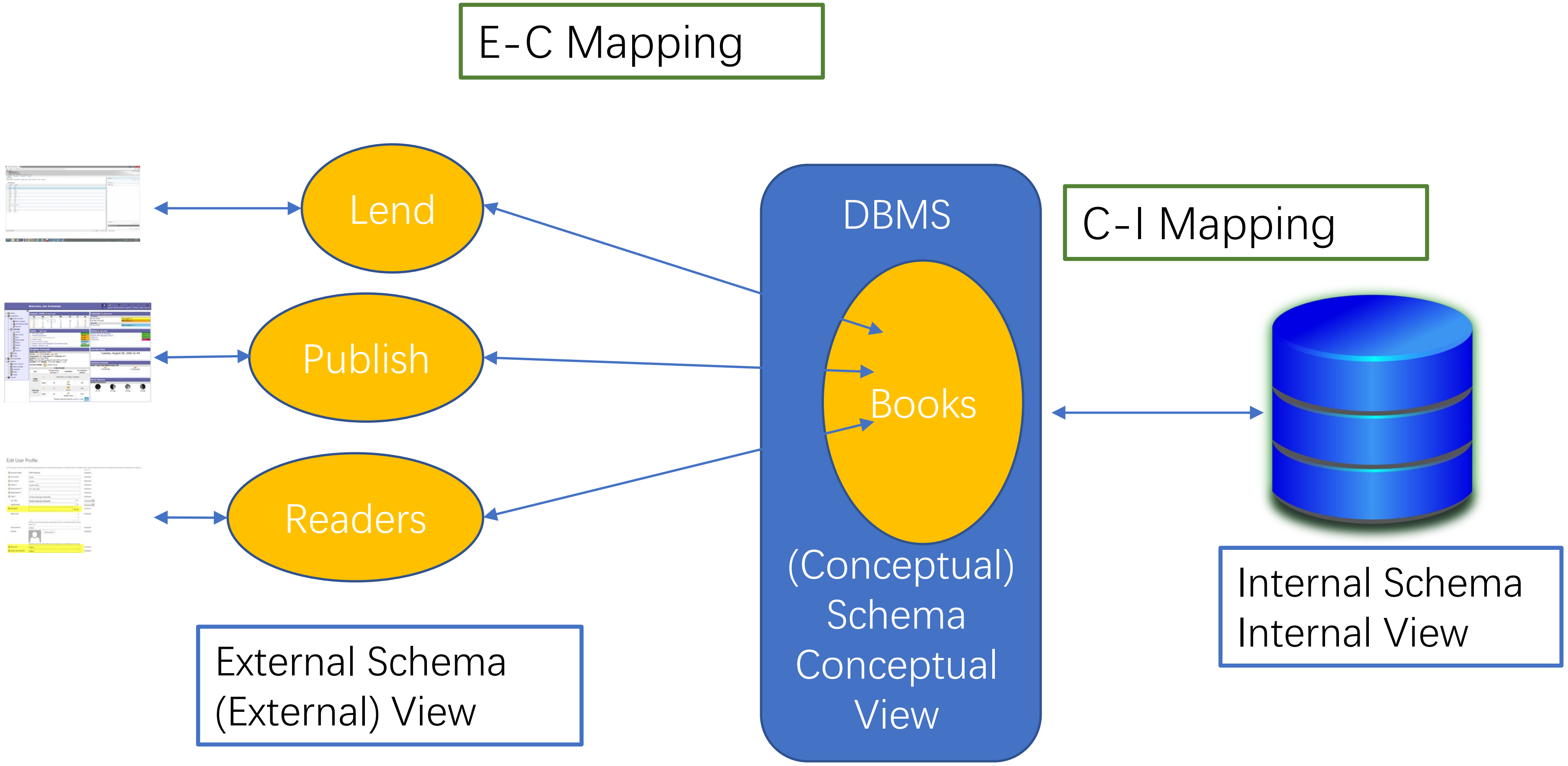
- Schema
  - Describe the structure
  - sID char(8)
  - sName varchar(20)
  - GPA real
  - sizeHS int



sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000

View/data

3-schema/view, 2 mappings, 2 independences



E-C Mapping

Lend

Publish

Readers

External Schema
(External) View

DBMS

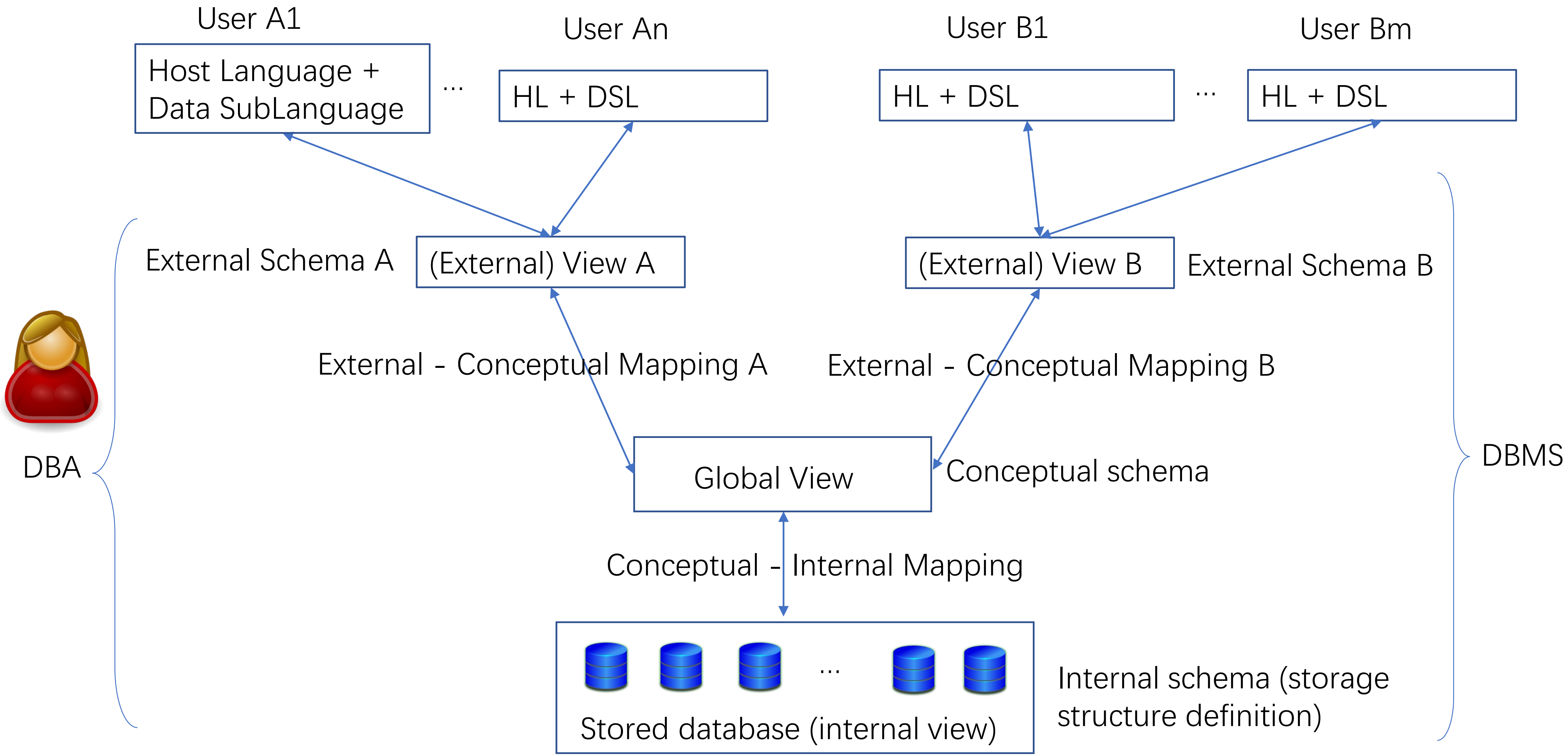
Books

(Conceptual)
Schema
Conceptual
View

C-I Mapping

Internal Schema
Internal View

- Logic data independence
- Physical data independence


User A1
Host Language +
Data SubLanguage
…
User An
HL + DSL
User B1
HL + DSL
…
User Bm
HL + DSL
External Schema A
(External) View A
(External) View B
External Schema B
External－Conceptual Mapping A
External－Conceptual Mapping B
DBA
Global View
Conceptual schema
DBMS
Conceptual－Internal Mapping
…
Stored database (internal view)
Internal schema (storage
structure definition)

Data Model

- Data model: structure of schema
- Relational model
  - Table/Relation
  - Relation Calculation
  - Constrains about table

data

schema

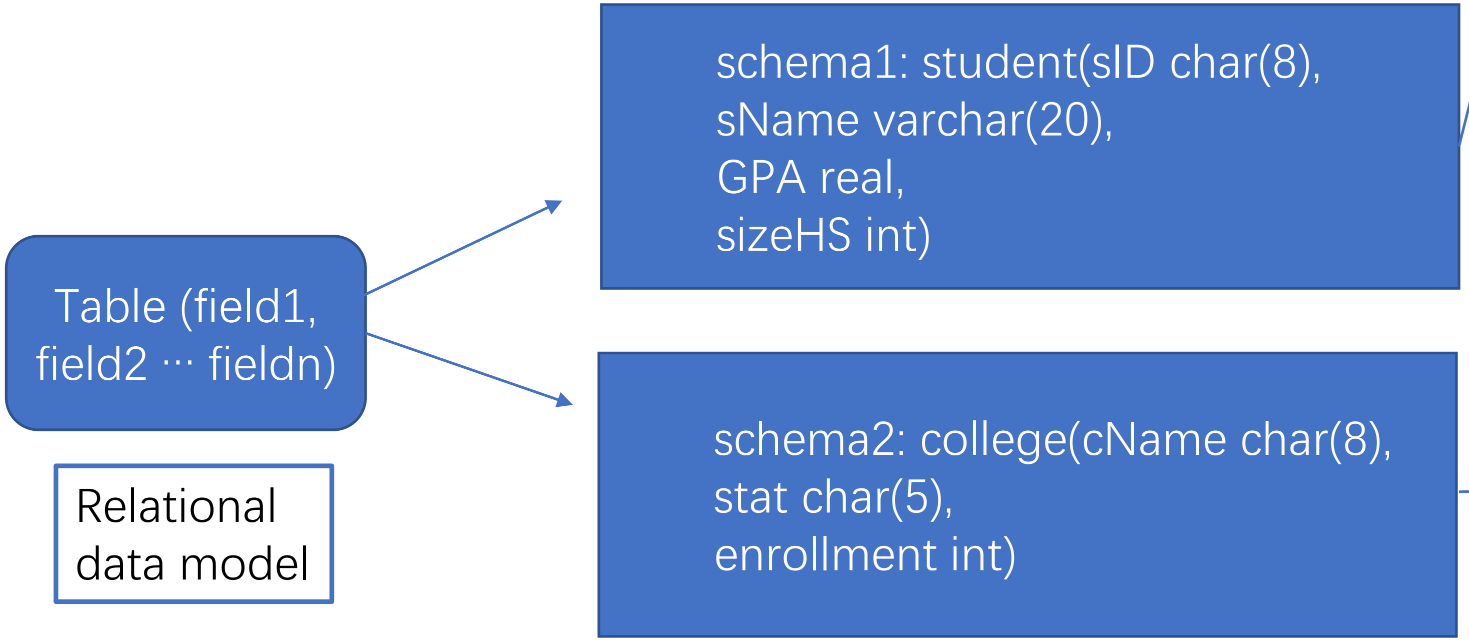
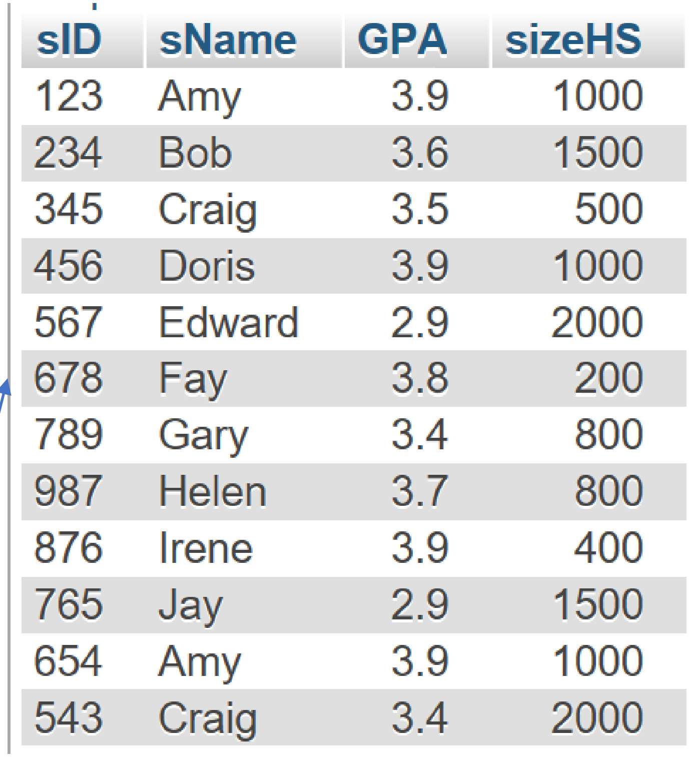


Table (field1, field2 ··· fieldn)

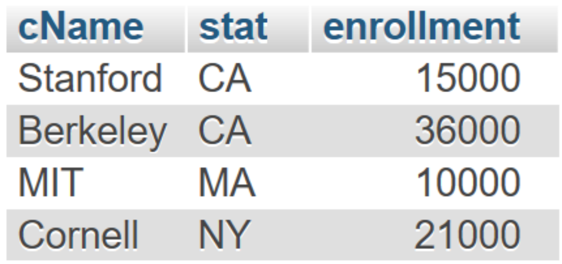
Relational data model

schema1: student(sID char(8),
sName varchar(20),
GPA real,
sizeHS int)

schema2: college(cName char(8),
stat char(5),
enrollment int)



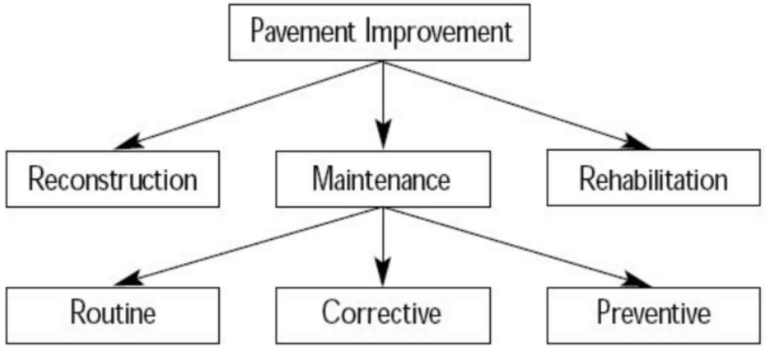
sID	sName	GPA	sizeHS
123	Amy	3.9	1000
234	Bob	3.6	1500
345	Craig	3.5	500
456	Doris	3.9	1000
567	Edward	2.9	2000
678	Fay	3.8	200
789	Gary	3.4	800
987	Helen	3.7	800
876	Irene	3.9	400
765	Jay	2.9	1500
654	Amy	3.9	1000
543	Craig	3.4	2000



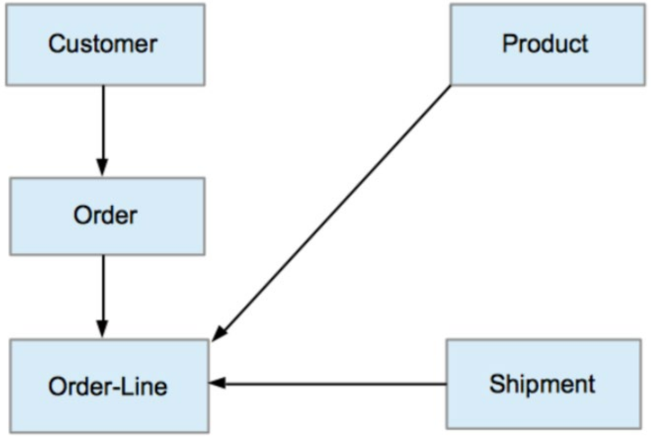
cName	stat	enrollment
Stanford	CA	15000
Berkeley	CA	36000
MIT	MA	10000
Cornell	NY	21000

3 types of data model

- Relational: table
- Hierarchical: tree
- Network: graph



Hierarchical



Graph