Database Systems

Gary KL Tam

Department of Computer Science
Swansea University

Why Study Databases?

- Databases are important for computing
  - Many computing applications deal with ***large amounts*** of information
  - Database systems give a set of ***tools*** for storing, searching and managing this information

- Databases are a '***core topic***' in computer science and IT
- ***Basic concepts and skills*** with database systems are part of the skill set you will be assumed to have as a CS and IT graduate

Why Study Databases?

Really important in Computer Science degree

Databases are (virtually) everywhere!

- Library catalogues
- Medical records
- Bank accounts
- Stock market data
- Personnel systems
- Product catalogues
- Telephone directories
- Train timetables
- Airline bookings
- Credit card details
- Student records
- Customer histories
- Stock market prices
- Discussion boards

and many more...

Director of Student Recruitment

Institution:	Kennesaw State University
Location:	Kennesaw, GA
Category:	Admin - Admissions and Enrollment
Posted:	06/29/2013
Application Due:	08/28/2013
Type:	Full Time

Not just for I.T. people!
Useful for many management and administrative jobs

Kennesaw State University is seeking an experienced, highly collaborative and goal-oriented enrollment management professional to lead a university-wide undergraduate and international student recruitment program. Reporting to the Associate Dean for Enrollment Services, the Director works closely with the Student Recruitment and Marketing team to develop data enabled recruitment strategies or initiatives.

...related experience, management/supervisory skills are required.

Preferred qualifications:

A master's degree and seven to ten years of progressive leadership experience in the management of a college/university-wide student recruitment program is preferred. The ideal candidate will have a proven track record of managing multiple teams and recruitment related initiatives to exceed enrollment goals. Knowledge and use of student search databases (SSS), student information systems, and new technologies including social media and management of higher education CRM systems are preferred.

Demonstrated management and leadership ability and experience in motivating, supervising and evaluating the performance of staff. Decision making and problem solving skills. Knowledge of relational databases and computer operations. Knowledge and use of student search databases (SSS), student information systems, and new technologies including social media and management of higher education CRM systems. Ability to

Goal

- To teach students how to ***model***, ***communicate*** and ***implement*** business functions and data in a device and software independent way
- The underlying principles are 50+ years old.
- Every UK universities teach database (esp. SQL).
- Students should be able to ***adapt*** to new software and technology
- Students will ***not learn specific packages*** which may cease to exist or evolve out of recognition within a few years (e.g. RethinkDB)

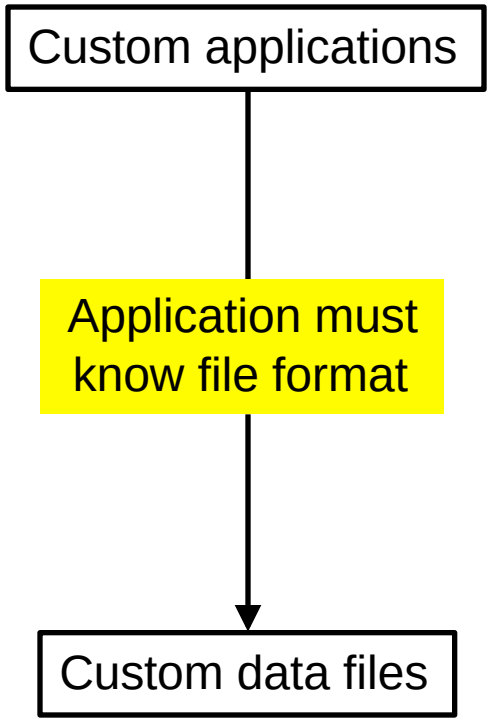
What is a Database?

- "A set of information held in a computer"
  - Oxford English Dictionary
- "One or more large structured sets of persistent data, usually associated with software to update and query the data"
  - Free On-Line Dictionary of Computing
- "A collection of data arranged for ease and speed of search and retrieval by a computer. "
  - American Heritage Science Dictionary

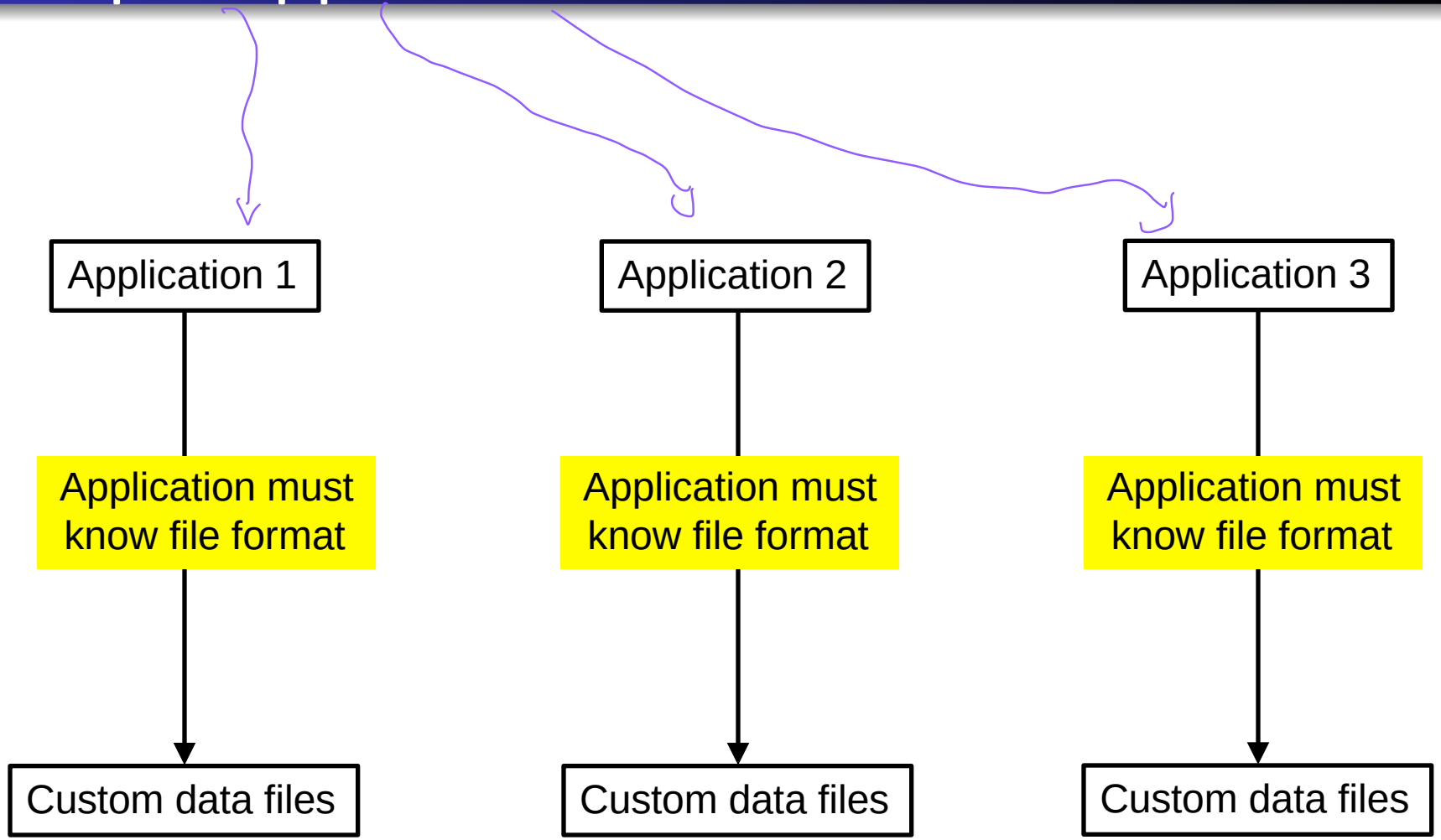
The VERY early days…

All application was ***tailor-made***

- Applications store (& persist) their data in files
- Each file has its own format
- Program has to know format
- Any other program using file has to know format



Multiple Applications

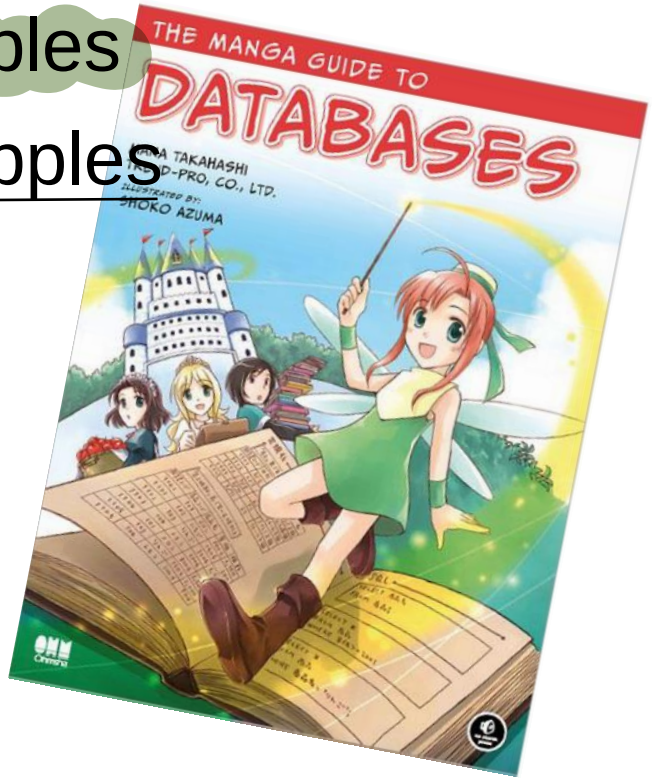


An Example

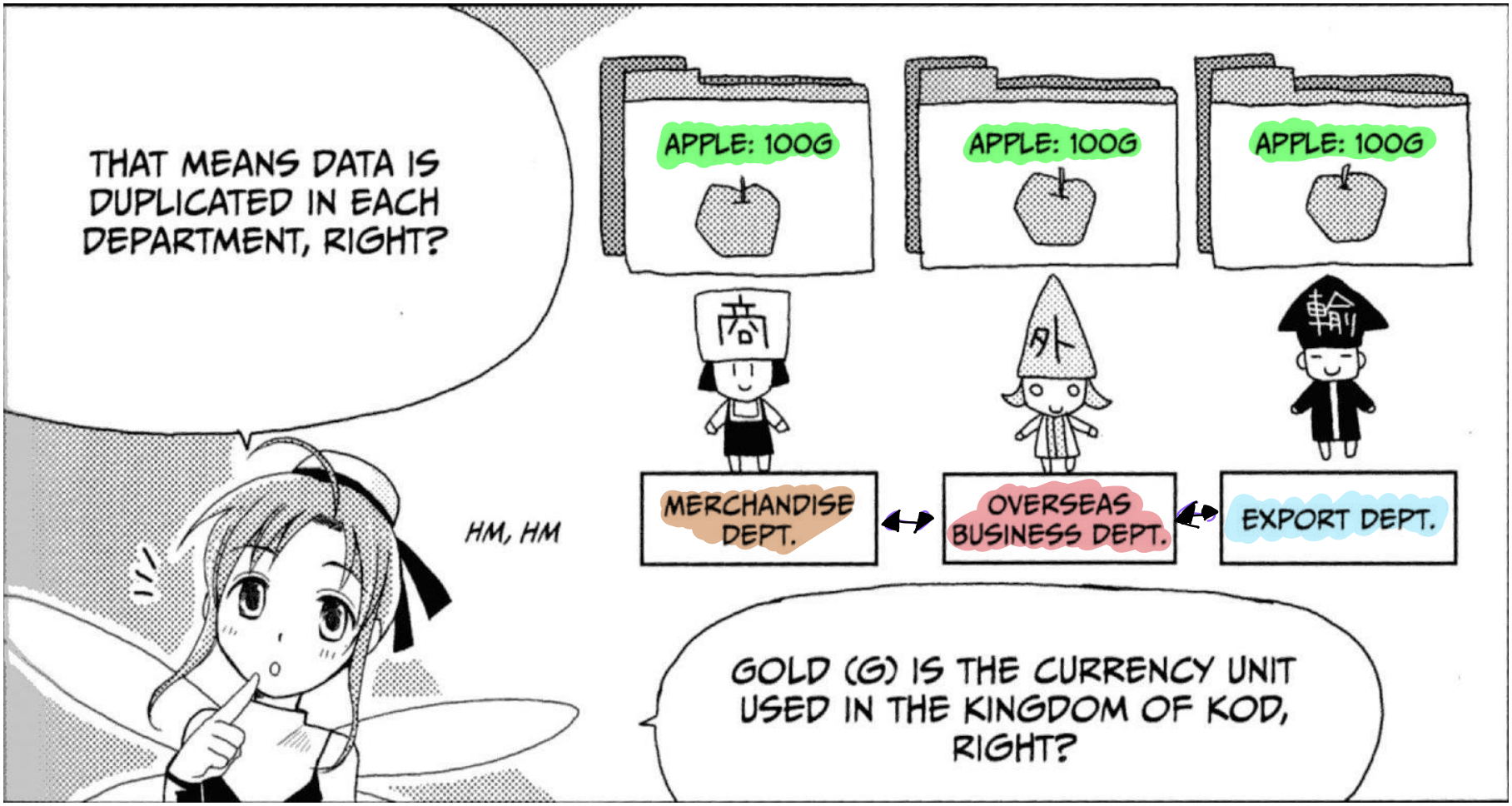
- From the Manga Guide to Databases

- The Kingdom of Kod exports apples
- Price is 100G per container of apples
- Three separate departments:
  - Merchandise
  - Overseas Business
  - Export

multi application

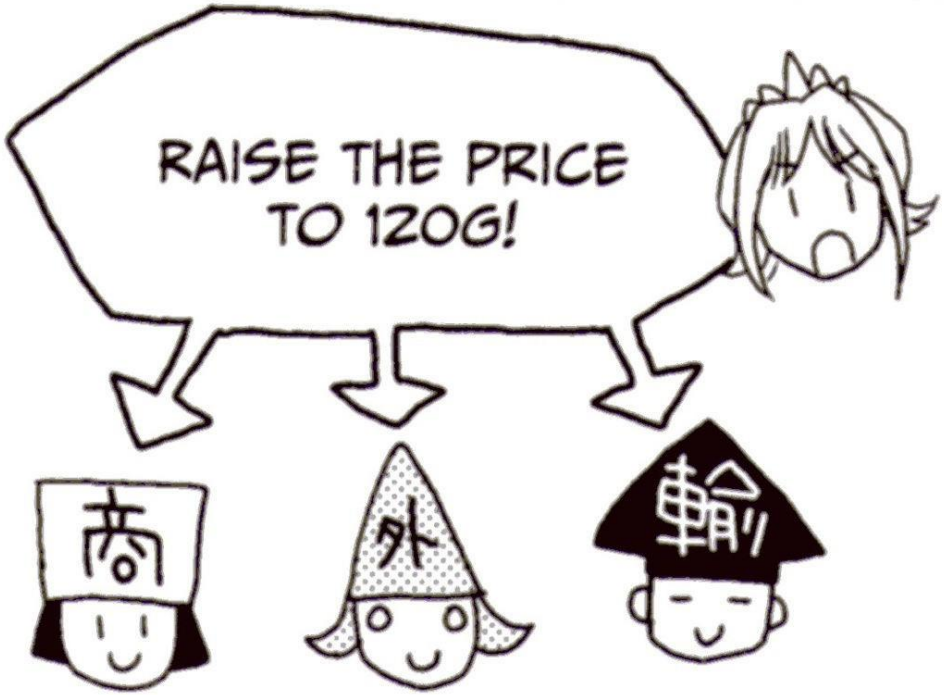


Multiple Copies of Same Data

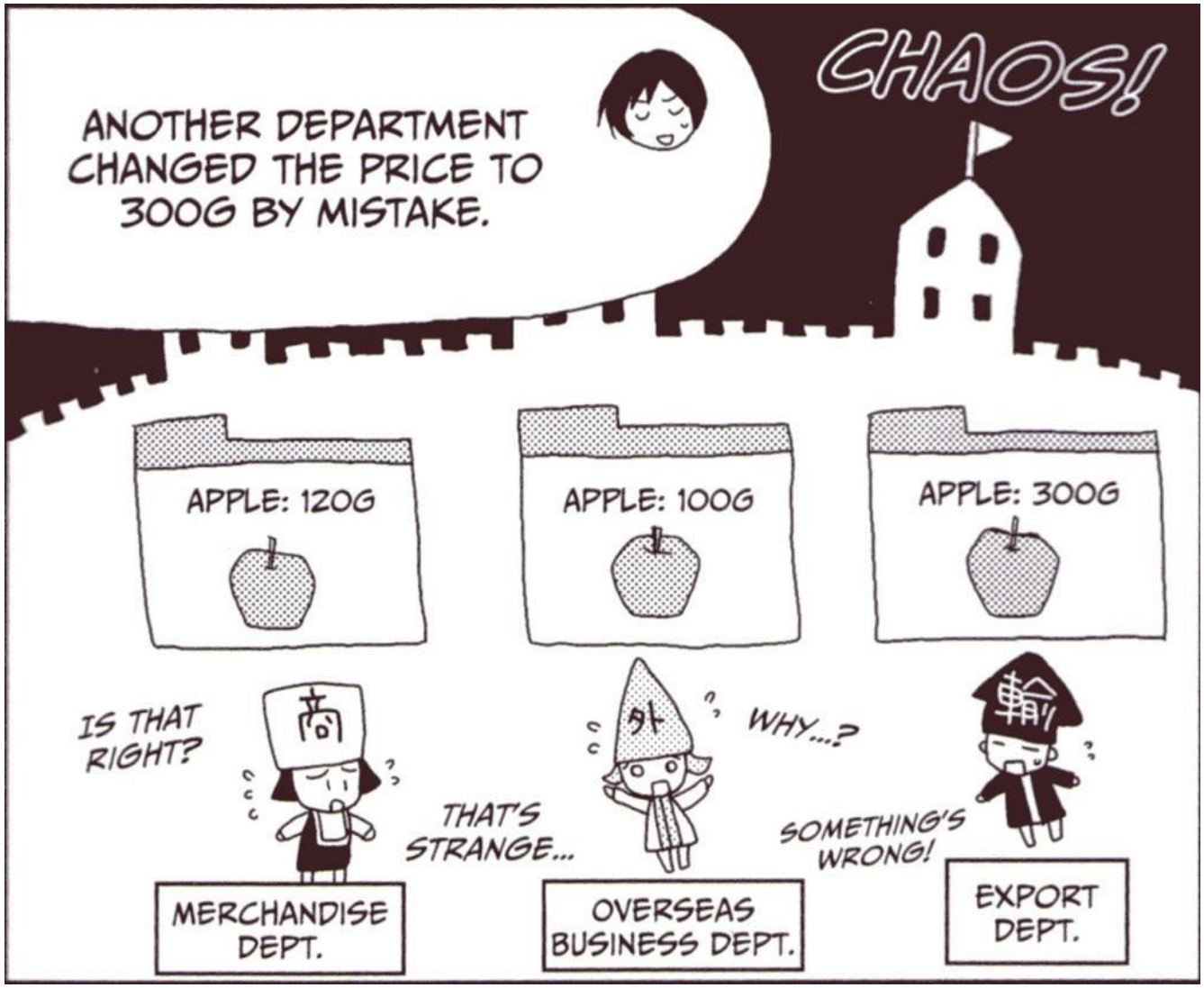


Can lead to errors…

I SENT A MESSAGE TO EACH DEPARTMENT TO CHANGE THE PRICE TO 120G, BUT...



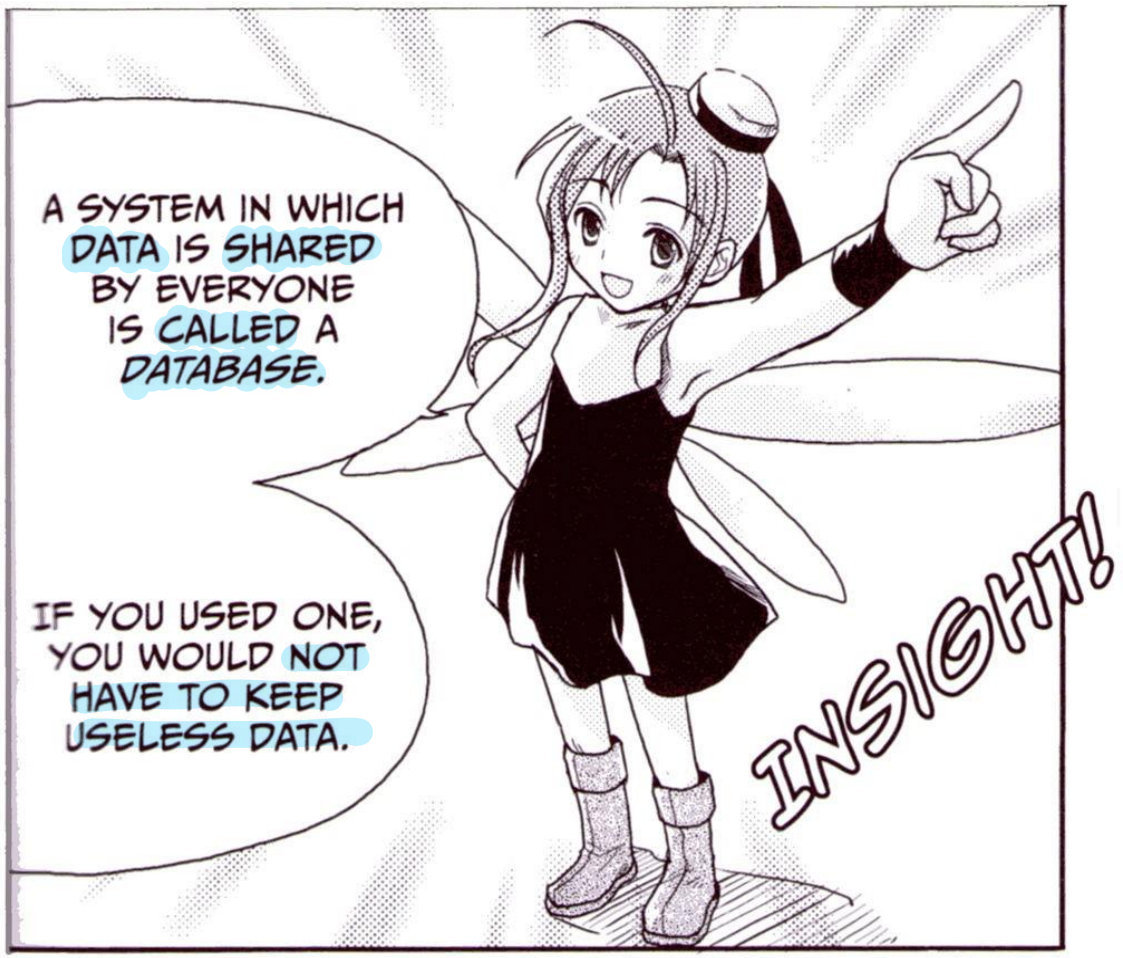
Mistakes happen



Overseas business department didn't get the message and ...

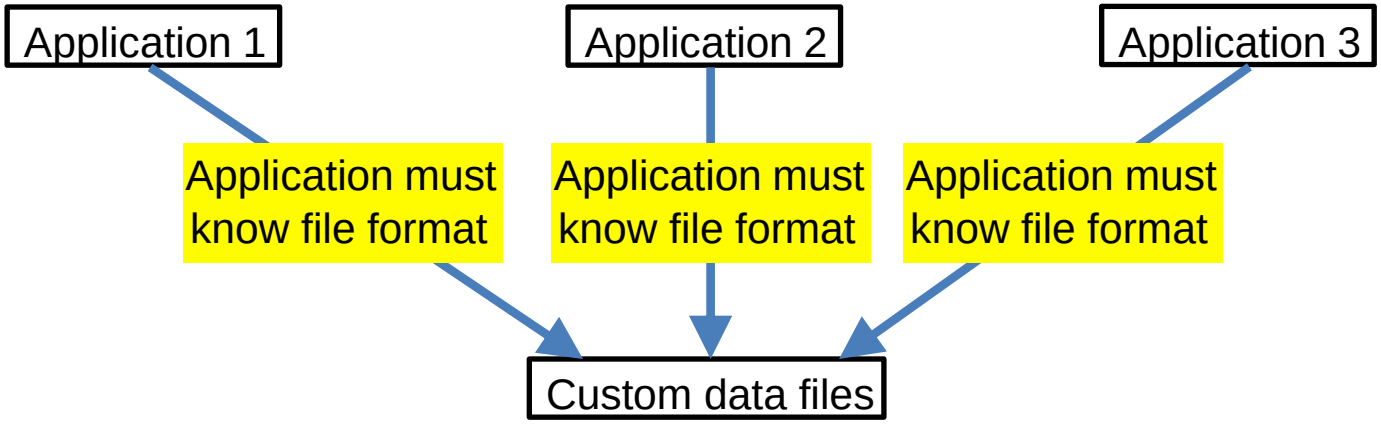
Redundant Data

Storing the same data several times in different places (redundancy) is error-prone!



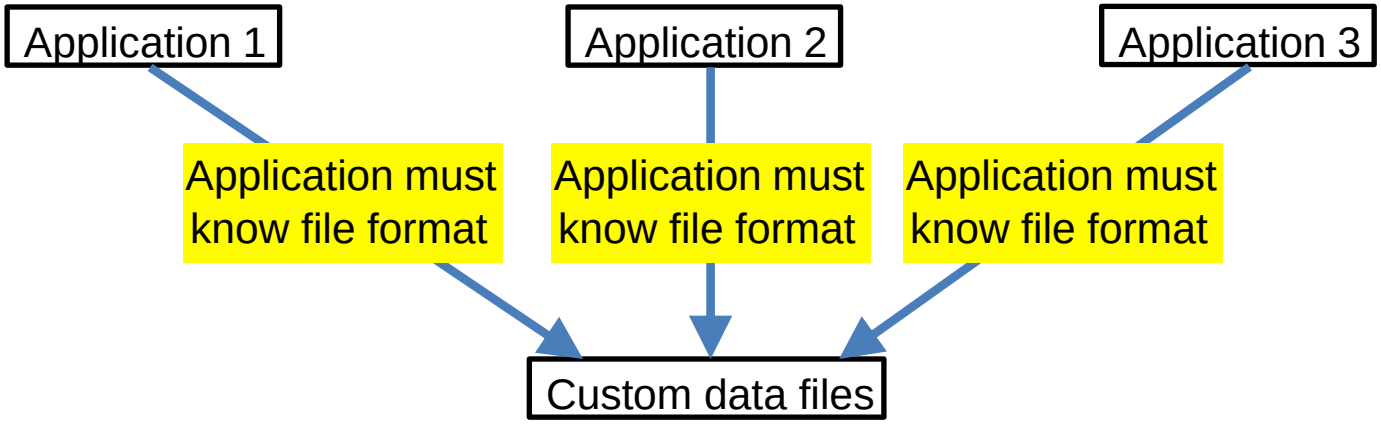
One copy of the data

- So, keep one copy of data…
  - All applications must know the file format
  - Data for all applications must be present



More problems

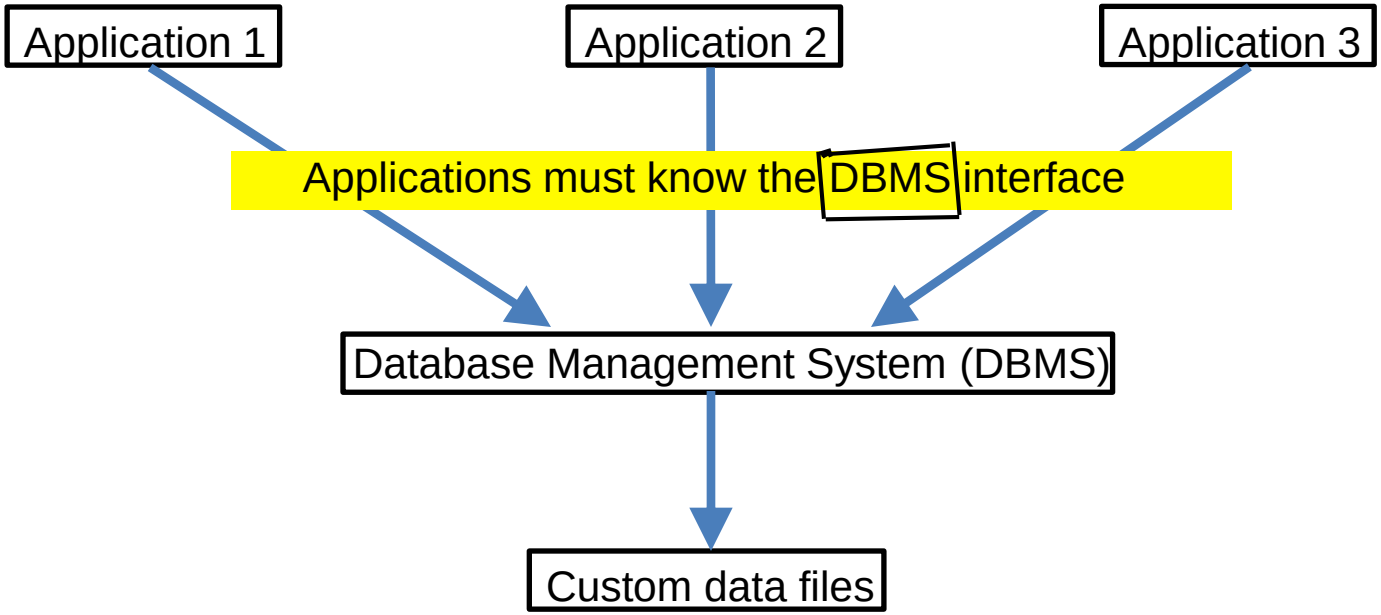
- There are still problems, for example:
  - Concurrency (multiple simultaneous changes?)
  - Security (everyone can see everything?)
  - Additions or changes to data format?
  - And others, as we will see



Put something in the middle…

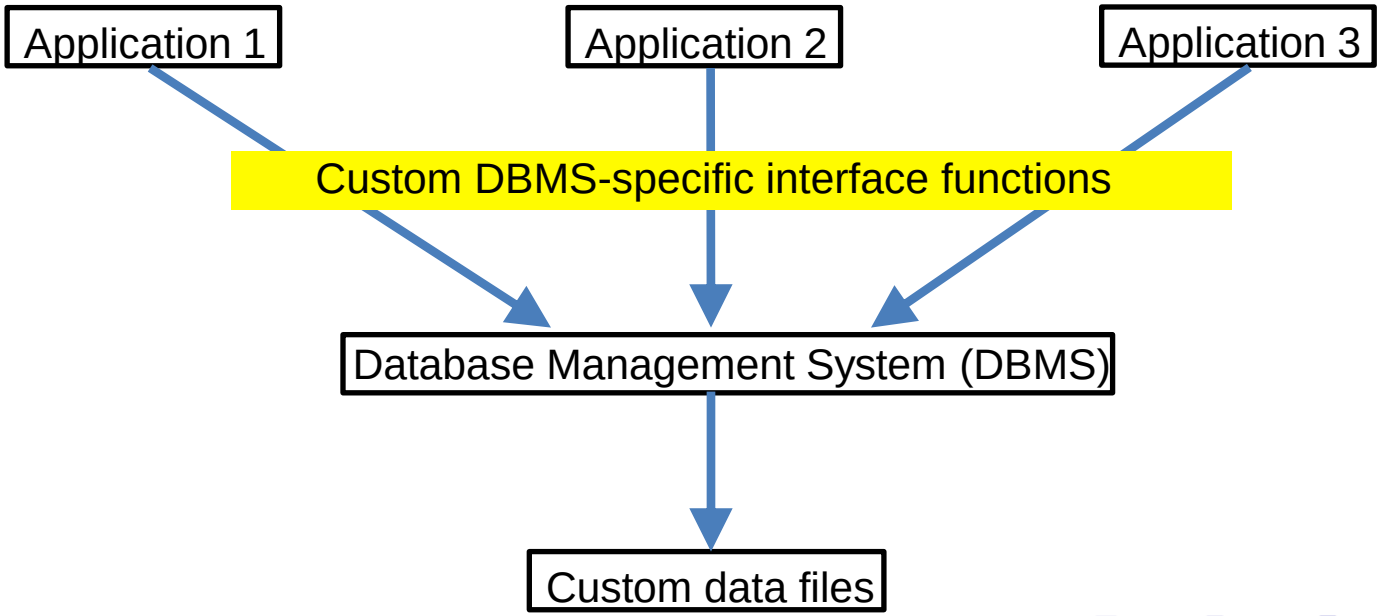
- A program in the middle can coordinate access
  - Preventing simultaneous access problems
  - Could provide extra integrity and security checks

- Applications link with DBMS rather than data files



Early databases

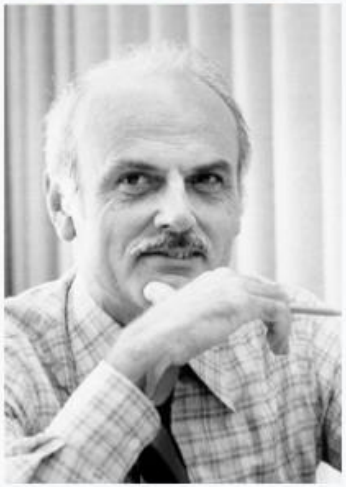
- Early databases were organised by the developer
  - New functions specifically created, not reusable
  - Adding new queries was complicated
  - No standards – database specific
  - Data duplication and data dependencies
  - Did not aid security, recovery, concurrency etc



Relational Databases

- 1970: E. F. Codd introduced the relational model
  - "A Relational Model for Large Shared Databanks"
- Information stored as records in relations (tables)
  - Sound mathematical basis
- Model covers data:
  - Structure
  - Integrity
  - Manipulation
- (Other database models exist, many are old (obsolete?), others are special purpose)

Edgar "Ted" Codd



Born	Edgar Frank Codd 19 August 1923[1][2] Fortuneswell, Dorset, England
Died	18 April 2003 (aged 79) Williams Island, Aventura, Florida, USA
Alma mater	Exeter College, Oxford University of Michigan
Known for	Alpha language Database normalization OLAP Relational model Codd's cellular automaton Codd's theorem Codd's 12 rules Boyce–Codd normal form

ANSI / SPARC Architecture

- Proposed a framework for DBMS in 1975
  - American National Standards Institute
  - Standards Planning Requirements Committee

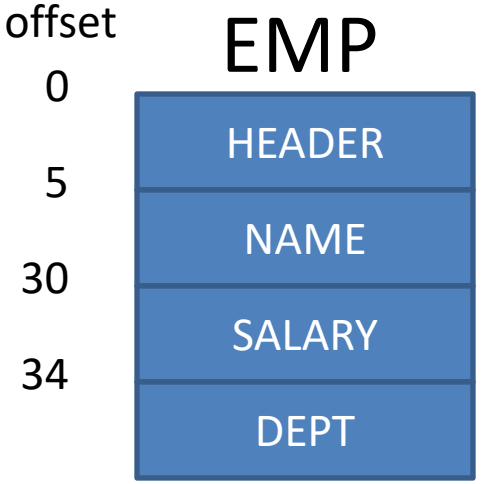
- Three tier/level architecture
  - Internal level - for systems designers
  - Conceptual level - for database designers
  - External level - for database users

Internal Level

- Deals with physical storage of data
  - Structure of records on disk - files, pages, blocks
  - Indexes and ordering of records

- Used by database system programmers

- Internal Schema example:
  RECORD EMP
  LENGTH=44
  HEADER: BYTE(5) OFFSET=0
  NAME: BYTE(25) OFFSET=5
  SALARY: FULLWORD OFFSET=30
  DEPT: BYTE(10) OFFSET=34



Conceptual Level

- Deals with organisation of entire database content

- Used by DBAs and application programmers
- Abstractions used to remove unnecessary detail
  - e.g. Internal level file formats
- Database holds metadata (data about data)
  - e.g. structure information for tables, primary, foreign keys etc

- Conceptual Schema Example:

```
CREATE TABLE Employee (
    Name VARCHAR(25),
    Salary REAL,
    Department VARCHAR(10)
)
```

DBA – Database Administrator

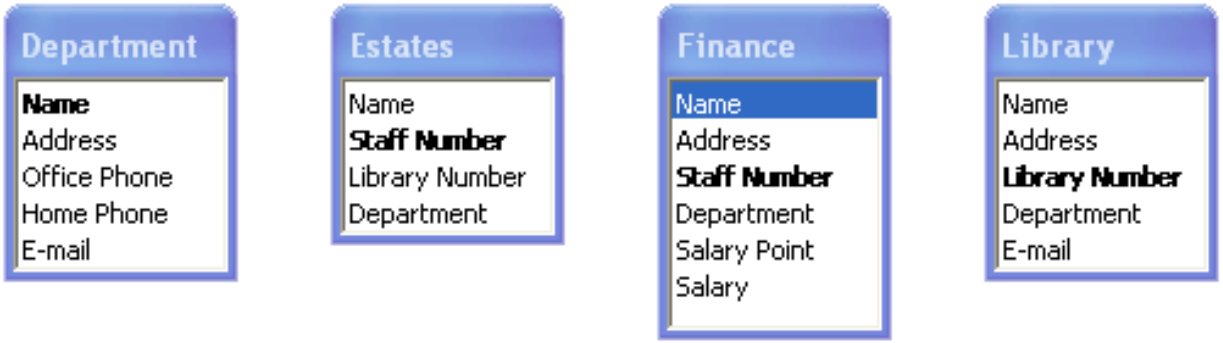
External Level

- Provides the view determined by the user
  - Data may be hidden
  - Data may be presented in a suitable form
  - Used by users and applications programmers

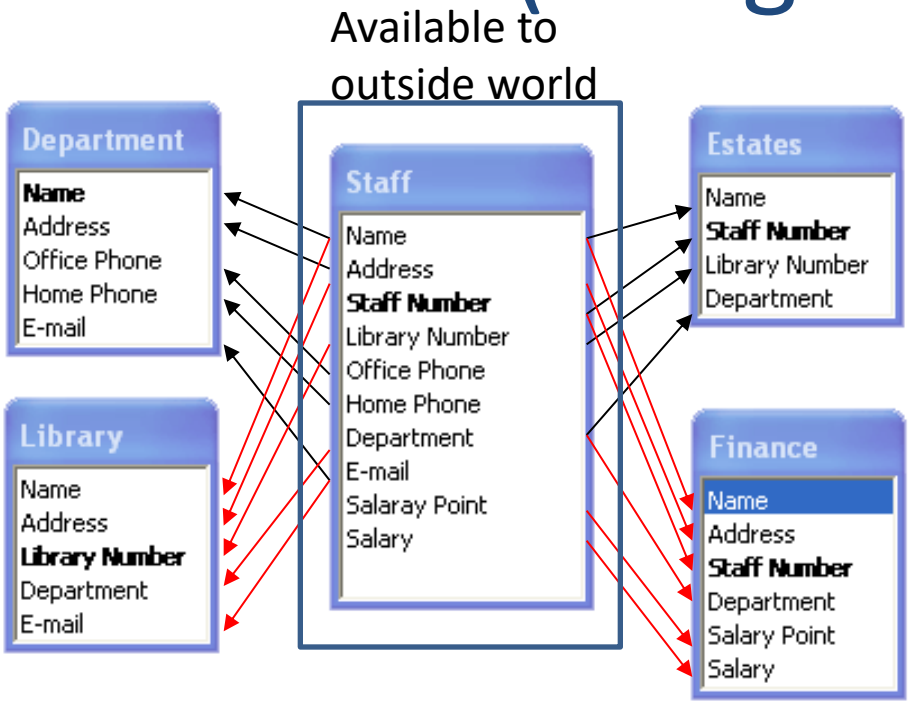
- External Schema example:

```
Create View myView as {
  SELECT Name FROM Employee
}
```

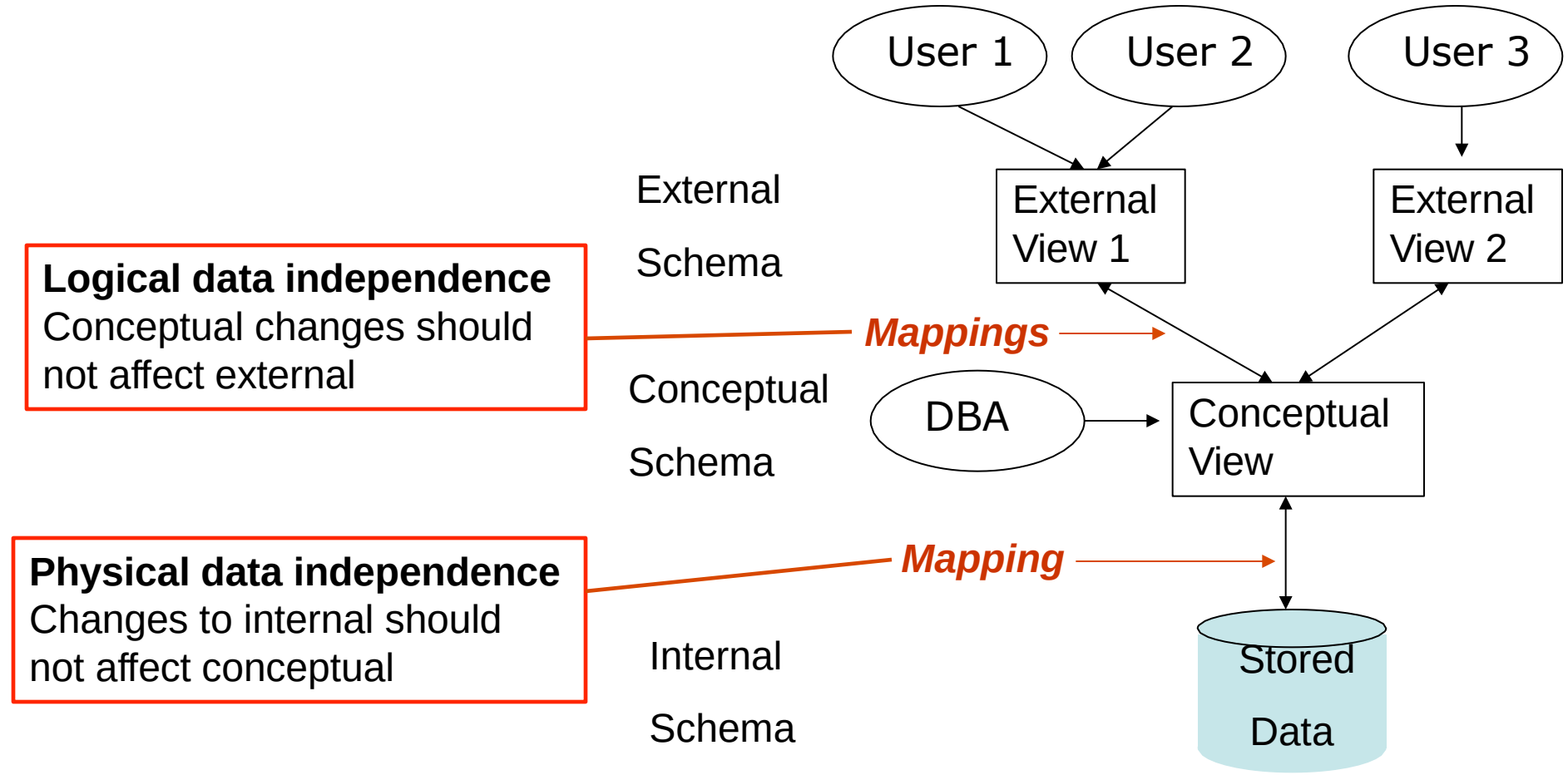
Conceptual Level



External Level (using views)



Mappings and users



Example Modern DBMSs

- Database Management System (DBMS): the software that implements a database

- Examples:
  - Oracle
  - DB2
  - MySQL
  - Ingres
  - PostgreSQL
  - Microsoft SQL Server

MS Access?

Excel?

DBMS User Facilities

- Allow users to:
  - Store data
  - Manage change (updates)
  - Organise data
  - Retrieve data
  - Retain privacy (security)
- And they expect it to always work!
  - i.e. recover from errors, avoid multi-user problems, inform if something goes wrong,…

DBMS Functions

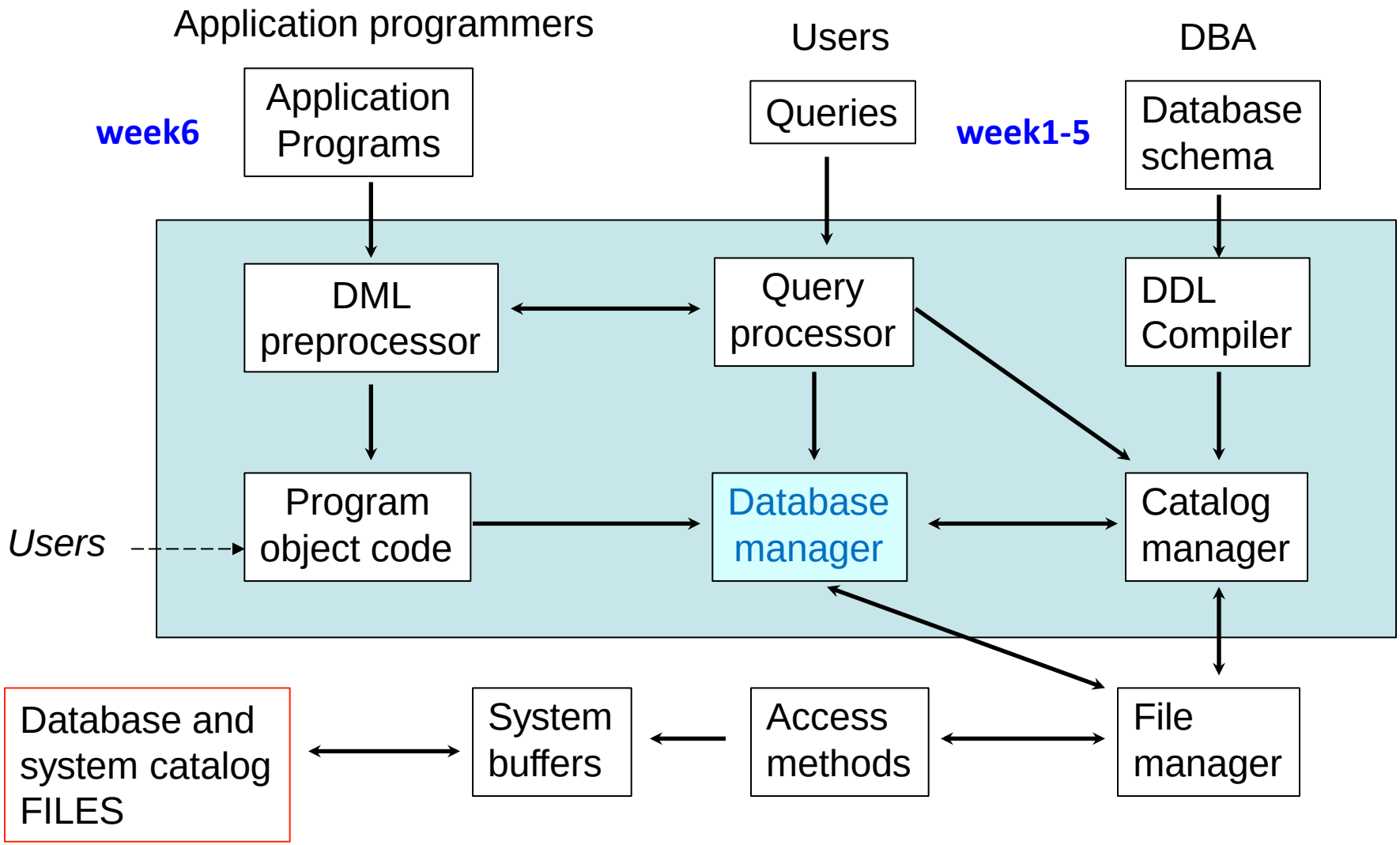
- Data storage, retrieval and update
- User accessible catalog
- Transaction support
- Concurrency control
- Recovery services
- Authorisation services
- Support communication software
- Integrity services
- Promote data independence
- Utility services (import/export, monitoring and logs, statistical analysis, consolidate files/indexes, reporting tools,…)

Provided Languages

- Data Definition Language (DDL)
  - Specify database format
- Data Manipulation Language (DML)
  - Specify and retrieve database contents
- Data Control Language (DCL)
  - Specify access controls

- Which are often all one piece of software
  - e.g. SQL

DBMS Components

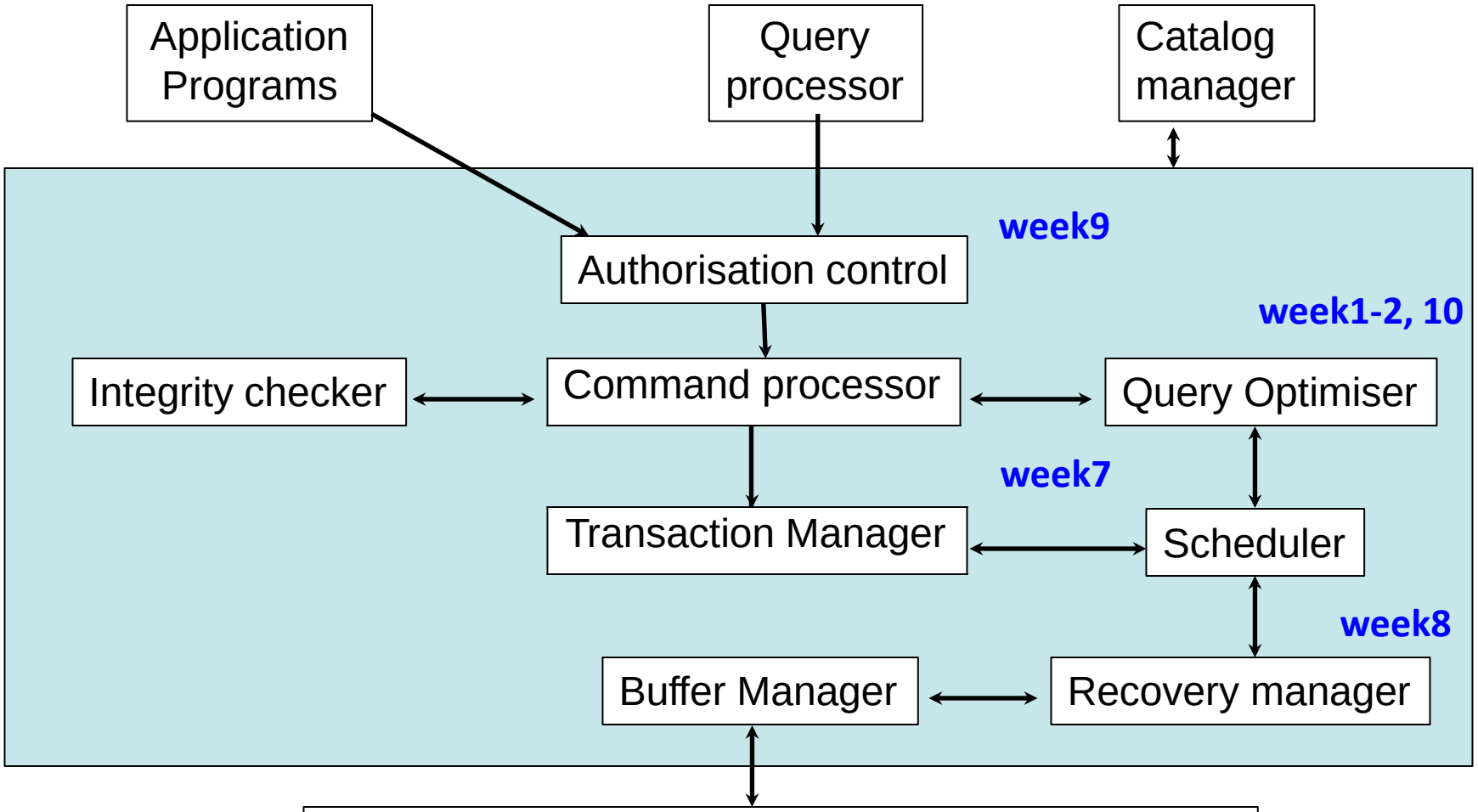


Connolly & Begg

Important DBMS Components

- Data dictionary / system catalog
  - Stores information about database objects
- Query processor
  - Interprets and optimises user queries
- DML preprocessor
  - Converts DML statements in application programs into standard function calls. Interacts with query processor to generate code
- DDL Compiler / Data Dictionary Compiler
  - Produces the data dictionary / system catalog
  - E.g. Converts DDL statements into tables
- Catalog Manager
  - Manages access to and maintains system catalog

The Database Manager



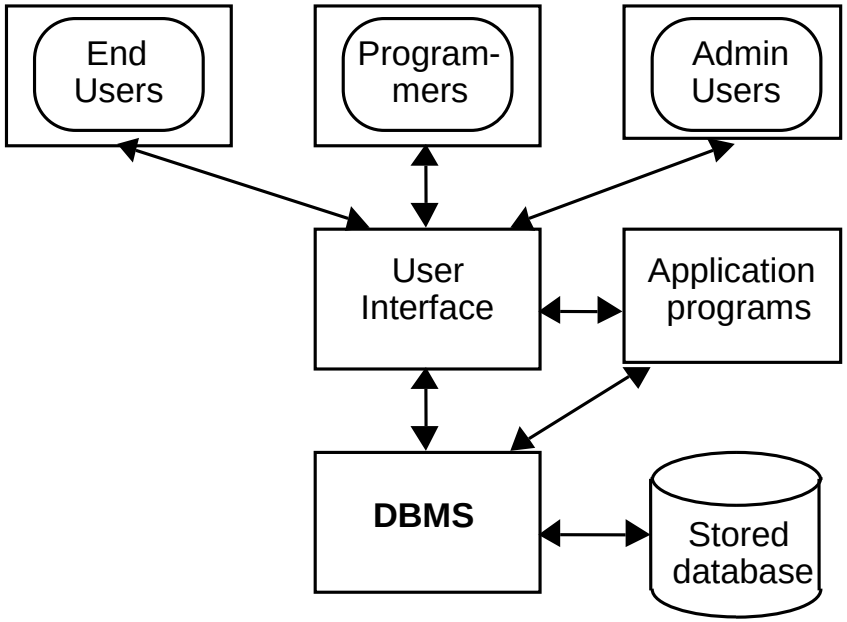
Internal Mechanisms, Algorithms etc

Connolly & Begg

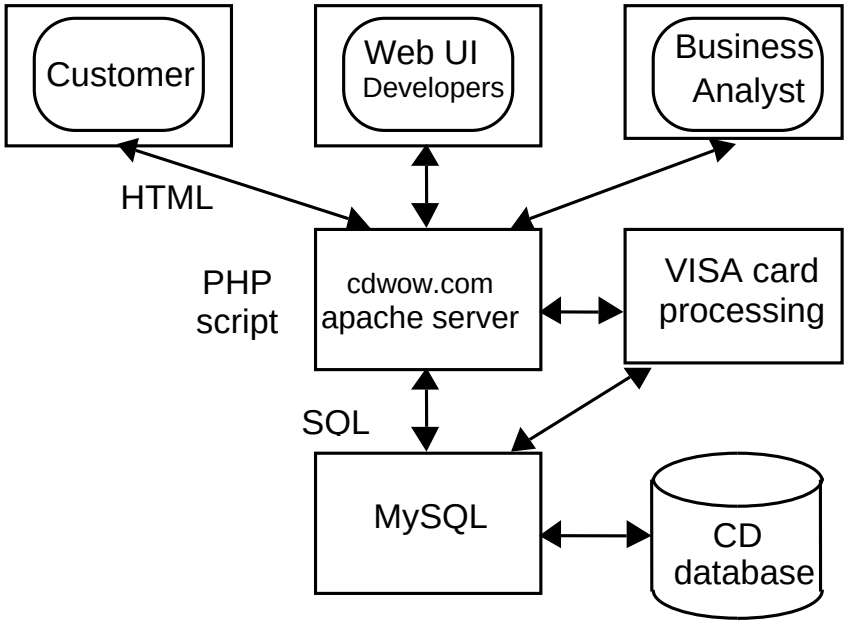
More Detailed DBMS Components

- Authorisation Control
  - Confirms whether user has correct authorisation
- Command Processor
  - Handles commands once system has confirmed authorisation
- Integrity Checker
  - Checks whether database changes meet integrity constraints
- Query Optimiser
  - Determines optimal strategy for query execution
- Transaction manager
  - Communicates with recovery manager and scheduler to ensure data consistency in multi-user environments
- Scheduler
  - Ensure that concurrent operations do not conflict, and orders them
- Recovery manager
  - Ensure database stays in consistent state after failures
- Buffer manager
  - Transfer data between main memory and secondary storage

Example Use of DBMS



CW2 a taste of PHP, MySQL, Apache



Example: Online CD Store