COMP283-Lecture 1
Applied Database Management

Introduction	
My Contact Details:	Phil Jimmieson
Room:	Ashton 1.20
Tel:	(+44/0) 151 795 4236
e-mail:	phil@liverpool.ac.uk
Comp283:	Applied Database Management 7.5 credits.
Lectures:	10 of 45 minutes.
Practicals:	10 of 2 hours.

COMP283-Lecture 1

Introduction: Please Note

- This is a practical module – you should make sure you attend all lectures, practical sessions and complete all the assignments
- Arrive at lectures on time please!
- **Do** ask questions

COMP283-Lecture 1

Introduction: Learning Outcomes

- Formal Learning Outcomes:
  - At the end of the module the student should be able to:
    - design and structure an efficient scalable database system,
    - implement and configure a database system,
    - maintain a secure and durable database,
    - describe, illustrate and explain the concept of linked databases and data migration techniques.
  - Informal Learning Outcomes:
    - The student should know how to produce timely and professional business plans, procedures, and reports.
    - Time management skills will be required for timely submission of assignments

COMP283-Lecture 1

Introduction: DB and DBMS

- What is a database?
- What is a **D**ata**B**ase **M**anagement **S**ystem (DBMS)?
  - A software application that helps define, store, retrieve and maintain data on computer systems.
- Is that all there is to a database?

COMP283-Lecture 1

Introduction: Data Models

- Databases and DBMS each follow a data model:
  - Flat
  - Hierarchical
  - Dimensional
  - Network
  - Relational
  - Entity-relationship
  - Graph
  - Object-oriented
  - Entity-attribute-value model

Usually a single file. One record per line with a field delimiter, e.g. csv file or fixed length (padded) fields.

COMP283-Lecture 1
Introduction: Data Models

- Databases and DBMS each follow a data model:
  - Flat
  - Hierarchical
  - Dimensional
  - Network
  - Relational
  - Entity-relationship
  - Graph
  - Object-oriented
  - Entity-attribute-value model

Data organised in to a tree-like structure, single parent – multiple child model. e.g. folder and file structure on a computer.

COMP283-Lecture 1
Introduction: Data Models

- Databases and DBMS each follow a data model:
  - Flat
  - Hierarchical
  - Dimensional
  - Network
  - Relational
  - Entity-relationship
  - Graph
  - Object-oriented
  - Entity-attribute-value model

Specialised Relational model tailored to data warehousing using Online Analytical Processing (OLAP) queries. Data is organised into numerical facts (measures) and categorised (dimensions). Quick and efficient for summing data or drilling down to specifics, e.g. a sales invoice could be used to indicate the total purchase price, but you could then query the items bought, where the item was stored, where it was manufactured…

COMP283-Lecture 1
Introduction: Data Models

- Databases and DBMS each follow a data model:
  - Flat
  - Hierarchical
  - Dimensional
  - Network
  - Relational
  - Entity-relationship
  - Graph
  - Object-oriented
  - Entity-attribute-value model

Data is organised into sets with a name, an owner and optionally a reference to another related set. The reference is a link to the owned set of data. e.g the World Wide Web is a networked database. A web page is defined by its URL, it contains information and may have links to other web pages.

COMP283-Lecture 1
Introduction: Data Models

- Databases and DBMS each follow a data model:
  - Flat
  - Hierarchical
  - Dimensional
  - Network
  - Relational
  - Entity-relationship
  - Graph
  - Object-oriented
  - Entity-attribute-value model

Consists of relations (tables) of structured data, where each tuple (record) has the same attributes and conform to the same constraints. The data can be separated into separate tables and linked using keys that uniquely define the record. Relationships between tables are enforced using primary and foreign keys that are pre-defined and ensure corresponding attributes of information maintain their integrity and conform to the constraints.

COMP283-Lecture 1

Introduction: Data Models

- Databases and DBMS each follow a data model:
  - Flat
  - Hierarchical
  - Dimensional
  - Network
  - Relational
  - Entity-relationship
  - Graph
  - Object-oriented
  - Entity-attribute-value model

Uses entities, relationship, and attributes. An entity will have one or more attributes and will be described by its relation to another entity. E.g The entity artist will have the attributes, name and date of birth, they will have the relationship of “Sang” to another entity called song which in turn has an attribute of title. The entity-relationship is a graphical way of defining datasets and their associations.

COMP283-Lecture 1
Introduction: Data Models

- Databases and DBMS each follow a data model:
  - Flat
  - Hierarchical
  - Dimensional
  - Network
  - Relational
  - Entity-relationship
  - Graph
  - Object-oriented
  - Entity-attribute-value model

Based on graph theory. Nodes contain properties (information describing an entity - the facts), edges describe the link and relationship to other nodes. A lot of the focus is on these edges describing the relationship, rather than the properties of the node. The edge can be queried to look at similar relationships – not so easy in a relational database.

COMP283-Lecture 1

Introduction: Data Models

- Databases and DBMS each follow a data model:
  - Flat
  - Hierarchical
  - Dimensional
  - Network
  - Relational
  - Entity-relationship
  - Graph
  - Object-oriented
  - Entity-attribute-value model

Information is stored as objects, in an Object Orientated database the DBMS is more closely incorporated in to the data model. Each object consists of Attributes and Methods. Attributes consist of information, methods define what can be done with the data. Object orientated databases can be used to store and define complex systems, e.g. Describe a plane, the modular aspects are wing, fuselage, tail, engine, etc.

COMP283-Lecture 1
Introduction: Data Models

- Databases and DBMS each follow a data model:
  - Flat
  - Hierarchical
  - Dimensional
  - Network
  - Relational
  - Entity-relationship
  - Graph
  - Object-oriented
  - Entity-attribute-value model

Used where there are potentially many different attributes but most of these are not relevant. The data for an entity only records the pertinent or “not-null” values. e.g. a careers advice interview may record some standard attributes: name, date and time, but other attributes may only be recorded for a small group of people – hobbies might be paint-balling and flower arranging, their training might include two years in Outer-Mongolia as a yurt builder, the related skills and discussion would be relevant to very few people. A well defined database would be able to link these activities with a set of related attributes.

COMP283-Lecture 1
Introduction: DB Systems

- This course will be focussed on Relational Databases, primarily on MySQL.
- We will highlight key similarities and outline notable differences between MS-SQL, Oracle DBMS and MySQL.

COMP283-Lecture 1 DB Hardware Design

Location: Where will the server be located – under a desk or in an ultra-modern, mains monitored, climate controlled, bomb proof server room.

Storage: How much disk space will be required.

Network: What is the network infrastructure. How much data can be transferred. What restrictions (firewalls) are in place. What network protocols will be required.

Processors: What type of processor will be suitable, 32 or 64 bit, PowerPC, Intel family...
How many functional processors will be required?

Cost: What is the budget.
What are the hidden costs – will it require additional cooling, electricity, support.
What are the direct costs – licences, hardware, commissioning, contractors.

Availability: How responsive does it have to be.
Should the database be disseminated to different locations – e.g. branch offices.

Reliability: What level of up-time is required, is 99% good enough or should it be 99.9999%

Security: Physical location.
On-line access.
Number of admin users.
Should some corporate databases reside on the same server or should they always be separated.
Confidentiality agreements and privacy policies.
Regulations.

DB Hardware Design: Location

- Location
- Storage
- Network
- Processors
- Cost
- Availability
- Reliability
- Security

- Where?
- Facilities
- Services (electricity, water, cooling)

DB Hardware Design: Storage

- Location
- Storage
- Network
- Processors
- Cost
- Availability
- Reliability
- Security

- Assess current storage capacity
  - Disk Space
  - Disk throughput
  - Location storage requirements
- Calculate future storage requirements

COMP283-Lecture 1

DB Hardware Design: Network

- Location
- Storage
- Network
- Processors
- Cost
- Availability
- Reliability
- Security

- Examine the network topology (layout).
- Identify bandwidth restrictions.
- Identify possible problem firewalls.
- Identify the networks for client access, server replication and data backups.
- How many server instances and databases can be supported?
- What network protocols must be supported?

COMP283-Lecture 1

DB Hardware Design: Processors

- Location
- Storage
- Network
- Processors
- Cost
- Availability
- Reliability
- Security

- A DBMS will be written for a specific operating system (OS)
- Both the DBMS and the Operating System will have some minimum and recommended system requirements.
- When choosing a DBMS you will also need to choose the OS and then determine which type of processor will be suitable.
- Be aware that a “normal” or recommended processor may not meet future requirements – you may need to over specify the processor.

COMP283-Lecture 1

DB Hardware Design: Costs

- Location
- Storage
- Network
- Processors
- Cost
- Availability
- Reliability
- Security

- Have to work to a budget
- Must be able to show cost effectiveness
- Evaluate all direct costs
  - Contractor fees
  - Licence fees
  - Location/Site costs
- Be aware of hidden costs:
  - Training costs
  - Warranty/Maintenance fees
  - Learning Materials
  - Location/Site costs
  - Location Preparation costs
  - Utility charges

COMP283-Lecture 1

DB Hardware Design: Availability/Reliability

- Location
- Storage
- Network
- Processors
- Cost
- Availability
- Reliability
- Security

- Availability = Responsiveness
- Reliability = Up-time
- Clustering is the process of providing some redundancy or high availability.
- Databases can be mirrored.
- Databases can be distributed.

COMP283-Lecture 1

DB Hardware Design: Security

- Location
- Storage
- Network
- Processors
- Cost
- Availability
- Reliability
- Security

- Who has physical access to the systems?
- Who is/are the System Administrators?
- Who will manage the DBMA - the Database Administrators (DBAs)?
- Who should have access to the data?
- Who does have access to the data?
- How secure is data communication?
- Will the data be encrypted when stored?

COMP283-Lecture 1

DB Design: Resource Considerations

- Consolidation is the process of combining hardware and/or software.
- A database server can run one or more database instances.
- A database instance can run one or more database applications.

COMP283-Lecture 1

Conclusion

- General introduction to this course
- Defined the Learning Outcomes
- Defined what a DBMS is.
- Introduced general Database design considerations.