COMP1140: Database and Information Management

Lecture Note – Week 01

Dr Suhuai Luo

School of SEEC

University of Newcastle

This lecture...

- **Introduction to course**

- **Introduction to DBMSs**

- DBMS in Perspective

- Database Design Process

- More on Requirements Gathering

Introduction to course - Teaching Group

Coordinator/Lecturer/Tutor:

- Suhuai Luo
- Email: suhuai.luo@newcastle.edu.au
- Office & Phone: ICT3017, 49854508
- Contact times: Tue 12pm -1 pm (by appointment)

Other members:

- Ananda Kalutantiri (kalutantiri@newcastle.edu.au)
- Sharlene Von Drehnen <sharlene.vondrehnen@newcastle.edu.au>

Introduction to course

- Objective - Learning Outcomes
  - Design conceptual models encapsulating data requirements for business and organisational scenarios
  - Convert conceptual models to logical data models
  - Implement a database solution using contemporary database management systems (DBMS)
  - Develop expertise in database languages (e.g. SQL) including the ability to develop sophisticated queries to extract information from large datasets
  - Understand data security and data quality management

Introduction to course

www.datacentreknowledge.com

o What is database:

- Collection of related data and information

o Database examples:

Introduction to course

www.datacentreknowledge.com

○ Applications of database:

- Store and manage conventional data (e.g., bank statement)
- Store and manage multimedia data (e.g., audio, video of surveillance camera; YouTube)
- Store and retrieve Web data (e.g., Google)
- Store and analyse big data (e.g., customer purchasing behaviour)
- Monitor system, automatically take action when required (e.g., warning of bank transaction)

Introduction to course - Why Database

- DB is an important part of IT area skills
  - Technical
    - Systems analysis; Systems design; Programming;
    - Database design/management; Data warehousing
    - IT architecture/standards; Operating systems
    - Voice/data telecommunications;
    - Server hosting; Security; Desktop support/helpdesk
  - Business domain
    - Industry knowledge; Company specific knowledge
    - Business process design/re-engineering
    - Communication
  - Project management
    - Project planning/budgeting/scheduling
    - Project risk management; Project leadership
- DB is part of your degree footpath
  - COMP1140 is a core for all computing programs (BIT, CS, SE), …
- It is the market and direction
  - Job; research; products; big data; …

Tech-kid.com

Introduction to course –

What you'll get/learn out of the course?

- Theoretical and practical knowledge of database design approaches
- Job-ready skills in DB design
- Potential/capability to work in BI

Introduction to course.

- ## How to achieve objectives?
  - ### Lectures provide...
    - Necessary background knowledge
  - ### Pracs/Tutes provide...
    - Hands-on application of knowledge acquired
    - Will design, develop and implement a database for an application domain following the database design process

Introduction to course

The Purpose of...

assessment is to **INCREASE** quality.

evaluation is to **JUDGE** quality.

Too short and not enough leaves. C-

Pcrest2.com
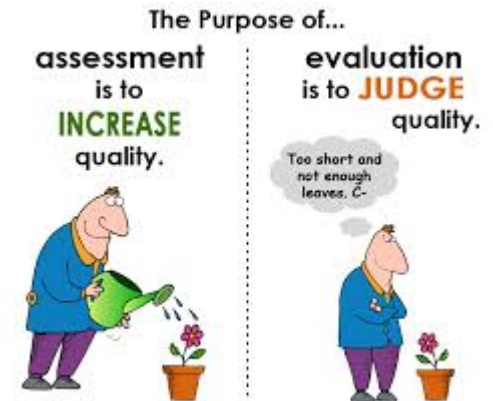
- **Evaluation of learning:**
  - **Assessments**
    - Assignment -
      - 3 assignments , subtotal 40%
      - Beware of lateness and plagiarism – read the university policies & course outline
    - Examinations –
      - Test – lab test in Week 9, 10%
      - Final exam, 50%
      - Note: Students must obtain 40% in the final exam to pass the course

Introduction to course...
Assignments

- SEEC Resource Access Project (individual work)
  - Assignment 1 – Description& EER model, due Wk4 , 15%
  - Assignment 2 - Logical Database Design, due Wk8, 10%
  - Assignment 3 - Physical Database Design, due Wk12, 15%
- Submission:
  - Both hardcopy & softcopy
  - Submission details as in assignment requirement document
    - Softcopy: submit via BB
    - Hardcopy: hand in to school office
  - You MUST be present at the lab of week 9 to take SQL test
  - You MUST be present at the lab of week 12 to have your A3 marked

Introduction to course…

- Lectures: Tuesday 1pm – 3pm, HD01
- Labs/Tutorials:

Tuesday	3pm - 5pm	ES105
Thursday	8am – 10am	ES105
Thursday	11am - 1pm	ES105
Thursday	2pm - 4pm	ES137
Friday	8am – 10pm	ES105
Friday	10am – 12pm	ES105
Friday	12pm – 2pm	ES105
Friday	2pm – 4pm	EFG14

- Download SQL Server from DreamSpark is arranged

Introduction to course...

- Blackboard (UoNline) Use
  - for communication, messaging and updates.
  - Lecture notes will be available on Blackboard (but details as per class lecture)
  - Announcements and Assignments will be available for download from Blackboard.
  - Assessments submission
  - Check it regularly along with your student email account. Do not use external email accounts.

Introduction to course...

- Reference Materials:
  - Text:
    - Thomas Connolly, Carolyn Begg, *Database Systems: A Practical Approach to Design, Implementation & Management*, Addison Wesley, 6th Edition (Global), 2015, ISBN-10 1292061189, ISBN-13 9781292061184
    - (for pre-order, contact The Co-op Callaghan, uoncallaghan@coop.com.au)
  - Other references in course outline
- Course outline provides details...
- Course content
- Qs?

Introduction to DBMSs

- What is data, information, databases, and DBMS?

- Requirements of today's information systems

- DBMSs and its features

- Files vs. DBMSs

- When to use DBMSs? When not to use DBMSs?

Data, Information, Databases and DBMSs

- **Data** are facts.

  E.g.   John Stuart

- Data are processed to create **information**.

  E.g.   John Stuart's grades

  Name: John Stuart
  Subject/Grades: INFT2040   A
                            INFT1004   A-
                            ...

Data, Information, Databases and DBMSs (contd.)

- **Database**

  Collection of related data and information.

- **Database Management System**

  a software package that is used to create, access and maintain databases

Organisations, data and information

- Every organisation collects and manipulates data and information
  - Examples.

- As the list illustrates, databases form an essential part of almost all enterprises/daily life today

Today's information systems

- Today's information systems collect and manipulate large amount of data.
  - Example

Requirements of today's Information Systems

- Store large amount of data

- Efficient (fast) access & manipulation of data

- Concurrent access by a large number of users

- Security

Requirements... (contd.)

- Reliability

- Data integrity

- Transaction processing

Files

- On computers, data are stored using different mechanisms.

- A popularly used mechanism is the Operating System's (OS) file system.

- You can create a file (text, word etc.) and store data

Files (contd.)

- Why don't we store data in OS files in an enterprise (e.g. bank database)?

- Files have certain limitations
    - **Sequential Access**: Files are accessed sequentially (hence not efficient for large amounts of data)
    - **Non-Concurrent access**: Files do not allow multiple users to update concurrently
    - **Security**: When you open a file, all the contents are visible to the user. Files have basic read, write, execute permissions.
    - **Cannot span disks**: Files cannot span multiple disks. Sometimes all the data might not fit on a single disk.

Files (contd.)

- The file age/ the importance of DB
  - Imagine time savings, storage space savings, information availability/updating in cases of:
    - to stay in touch and remember what was said/done in dealing with thousands of customers.
    - to keep the state of training or history of hundreds of employees.
    - to imagine how many people would have to be employed just making notes on paper, instead of being employed in creating new products and services

DBMSs

- DBMSs is a complex software package that creates and maintains databases.
- DBMSs provide features that are essential for today's information systems:
    - Storage of large amounts of data
    - Data independence
    - Powerful query capabilities
    - Efficient access
    - Security
    - Concurrent access
    - Easy data administration
    - Transaction Processing
    - Reliability
- Therefore, today we find DBMSs are used in almost every enterprise information system

DBMSs vs. Files

- With all its advantages, why do we use files then?
  - E.g.

  - Databases are expensive software packages
  - They take up resources (memory, disk space etc.)
  - Above all, the overhead of using a DBMS is not needed for these types of applications as...
    - Small amounts of data are stored (assignment or email is small)
    - Single user applications (not many people want to concurrently update your assignment!)
    - Overhead for concurrency control, query capabilities, security etc. are not needed for these types of applications

DBMS in Perspective - Evolution of DBMSs

- In early 1960s, Charles Bachman proposed the *Network Data Model*

- In the late 1960's, IBM developed Information Management System (IMS) based on the *Hierarchical Data Model*

- In 1970, E.F. Codd of IBM - San Jose Research Laboratory, proposed the *Relational Data Model*

- In 1990, came the **object-oriented** DB; **Data-warehousing**; **Web-enabled** DB

Evolution... (contd.)

- Benefits of the relational model was widely accepted & became the de-facto standard in commercial databases

- The System R Project at IBM led to the development of **S**tructured **E**nglish **Que**ry **L**anguage (Sequel) language which later was standardized to be SQL (Structured Query Language)

Evolution of DBMSs... (contd.)

- DBMSs are now an essential part of any Enterprise Information System

- Today, extending DBMSs in a variety of ways...
  - Storing complex data types (multi-media, spatial data etc.)
  - Adding tools and techniques to build vast data warehouses and mine data (Decision Support Systems)
  - ... etc.

DBMS Environment

- There are many different types of persons in developing, maintaining and using a database in an organisation
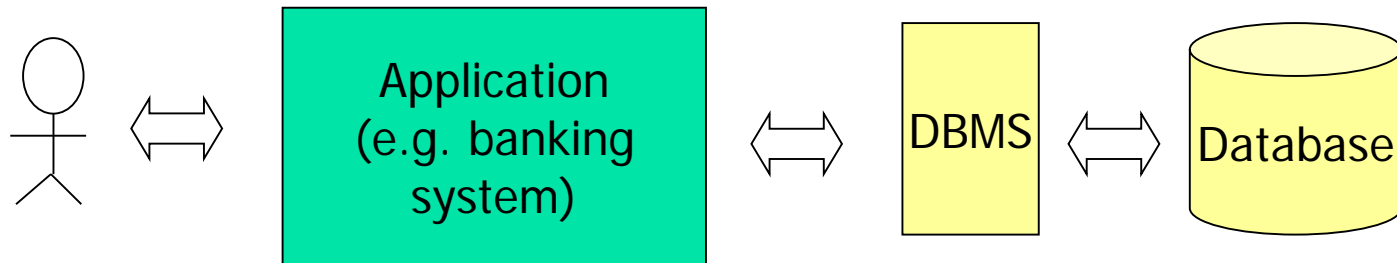
End-Users
Include all users for which the system was developed
(for e.g. customers of a bank, bank employees etc.)

Application Developers
Develop programs that access the database

Database Designer
Involved with designing the database for an application domain

DBA
Create and maintains database and system resources required for database

DBMS Environment (contd.)

- Database Administrators (DBA) are responsible to creating and maintaining databases and resources.

- DBAs use a variety of tools provided by the DBMS to monitor, manage, optimize and maintain the databases.

- DBA's duties include
  - optimize database for performance
  - plan, develop and implement backup and recovery procedures
  - ensure necessary resources such as disk space, memory etc. are met
  - implement security procedures

DBMS Environment (contd.)

- Database Designers are responsible for designing, the database

- Database Designers duties include
  - understand the application domain and data requirements
  - Design the database in a target DBMS, which includes:
    - Understanding application domain requirements
    - Modelling data requirements in a conceptual model
    - Mapping the requirements to a data model (logical database design)
    - Selecting structures and methods for storage and fast retrieval (physical database design)

DBMS Environment (contd.)

- Application Developers develop the programs that access the database providing the functionality of the information system.

- End-users include all users for which the Information System is built.

Database Design Process – Our course objectives…

- This course focuses on gaining knowledge and skills to design, develop and implement a relational database for an application domain

Development methodologies: What, Why and When?

- Developing large scale software is a complex process requiring careful design and management

- Certain methodologies have been proposed to ensure time, reliable and correct software products are built

Development methodologies: Why?

- Developing large scale software is complex and has had a high rate of failure (popularly known as the *Software Crisis*)

- Standish Report - The CHAOS Report
  - 24% of projects will be cancelled before they ever get completed
  - 44% of projects will cost more than their original estimates
  - 32% of software projects that are completed on-time and on-budget

Methodologies for software development

- To enable better productivity & manage large scale software development, software development processes are suggested and followed.

- A popularly known methodology is **Software Development Life Cycle** (**SDLC**) and its variants.

Requirements → Design → Implementation → Verification → Maintenance
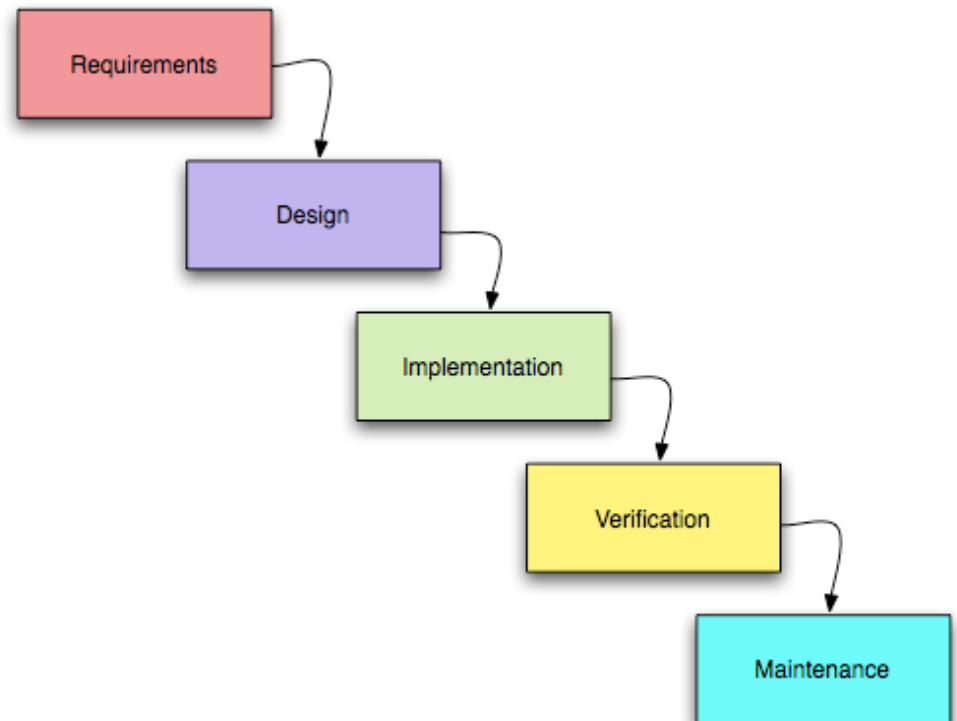
Database Design Process

- Database is a fundamental component of any enterprise information system.

- Information Systems development incorporates design and development of the database

- This course focuses on the proposed set of steps for developing databases, known as the **Database Design Process**

- We will follow these steps in completing this semester's assignment

Database Design Process

- Database design process consists of the following main steps:
  - Requirements Analysis
  - Conceptual Database Design
  - Logical Database Design
  - Physical Database Design

Requirements Analysis

- In this phase, you will understand and produce documentation outlining the detail requirements for the database system

- Fact finding techniques: interviews, research on existing systems etc. are used at this stage

- This step answers the following basic questions:
  - What types of data are stored in the databases?
  - What are types of services are implemented?
  - What types of business rules exist?

Conceptual Database Design

- In this step, a high-level conceptual model is designed encapsulating the data requirements

- Entity-Relationship (ER) Model is a very popular & widely used conceptual model

- ER Model is enhanced with object-oriented features & known as the Enhanced Entity-Relationship (E-ER) Model

- We'll use EER model in developing a conceptual model for the retail database

Logical Database Design

- At this stage, we convert the conceptual model developed in the previous step to a data model

- We will learn the relational data model and convert the EER conceptual model of database to relational data model (known as the **relational database schema**)

- At this stage, we further refine the relational database schema eliminating any redundancies that may exist. This process is called normalisation.

Physical Database Design

- In this stage, the logical database design is realised in a physical implementation targeting a specific DBMS

- Storage structures and access methods are considered for efficient access

- Design considerations for protected access are also considered

Requirements Analysis

- Fact finding techniques...
  - Examining documentation

  - Interviewing

  - Observing the enterprise in operation

  - Research

  - Questionnaires

Requirements Analysis (contd.)

- Results in a document which outlines
  - What data to be stored in the database and their relationships?
  - What are the operations that the database needs to support? (i.e. inserts/updates/queries)
  - What are the rules that the data must follow? (business rules)

Requirements Analysis – The DreamHome Case Study

- DreamHome specializes in property management, taking an intermediate role between owners who wish to rent out their furnished property and clients of DreamHome who require to rent furnished property for a fixed period. It currently has about 2000 staff working in 100 branches.

  (reference: chapter 11)

'The purpose of the *DreamHome* database system is to maintain the data that is used and generated to support the property rentals business for our c ients and property owners and to facil tate the cooperation and sharing of information between branches.'

Assignment

- Assignment 1 handout outlines the main features of the system. It is on course Blackboard, will be discussed next week.

- Sample requirements documents are available in Appendices A and B of the textbook (placed on Short Loans in library)

Summary

- **Introduction to course**
  - Objectives
  - Strategy
  - Lecturer info
  - Assessments
- **Introduction to DBMS**
  - Data, information, database and DBMS
  - Data needs of today's information systems
  - Files and their limitations
  - DBMS features
  - When and when not to use DBMSs

Summary (contd.)

- DBMS in Perspective
    - Evolution of database technology
    - Roles of persons in a database environment
    - Revisit course objectives
- Database Design Process
- More on Requirements Gathering

Summary Qs

- 1. What are the names and definitions of the four database design processes?

- 2. List nine features that a DBMS can provide that are essential for today's information systems.

- 3. Name and briefly describe the four types of persons in developing, maintaining and using a database in an organisation.