



INFO20003 DATABASE SYSTEMS



Complete Subject Notes
Semester 2 2016

Contents

Table of Contents

Contents	1
Data, Information & Knowledge	4
Data vs Information	4
Metadata.....	4
What is a Database?.....	4
Database Development.....	7
Lifecycle.....	7
Entity Relationship Diagrams	8
Data Dictionary	8
Data Types.....	9
Data Modelling.....	10
What Is Data Modelling	10
Entity Relationship Model.....	10
Business Rules as Possible Entitles	12
Single Entity Design.....	12
Multi-entity Design	13
Multivalued Attributes.....	13
Addresses	14
One-to-many Relationships	15
Binary One-to-one Relationships	16
Unary One-to-one Relationships.....	17
Unary One-to-many Relationships.....	18
Unary Many-to-many Relationships	18
Multiple Relationships	18
Ternary Relationships	19
Structured Query Language (SQL).....	20
What Is SQL?	20
How SQL Is Used	20
SQL Capabilities.....	20
SQL Queries.....	21
Ternary Relationship Querying	28
SQL Nested Queries	30
Aggregate Functions	31
Enhance ER Modelling	32

Super Subtype Relations	32
Extended Entity Relationship (EER).....	32
Identifying Super/Subtype Situations	36
Data Model Quality	37
Why Does Data Model Quality Matter?	37
Quality Evaluation Framework.....	37
Quality Factors	37
Web and Databases	40
SQL Problems	40
Web Applications	40
PHP and User Input	43
Miscellaneous PHP Function.....	45
Transactions	48
What Is a Transaction?.....	48
Why Use Transactions?	48
Defining Transaction	49
Transaction Properties and Features.....	49
Normalisation.....	51
What Is Normalisation	51
Normalisation Hierarchy	51
Normalisation Process	52
De-normalisation	52
Data Warehousing	53
Different Types of Databases.....	53
Data Warehouse	53
Big Data Analytics.....	56
What is Big Data Analytics	56
Big Data Analysis As A Social Process.....	57
Ethical Concerns of Big Data	58
Relational Algebra	59
What Is Relational Algebra.....	59
Relational Algebra Operators.....	59
Database Administration	64
The Database Administrator Role	64
Core Tasks for A DBA.....	64
Query Processing and Optimisation.....	69
Basic Steps in Query Processing.....	69
Measure of Query Cost	69

MySQL Execution Plan	72
Operation Heuristics	73
NoSQL Database.....	74
What is NoSQL?.....	74
Dominance of Relational Model	74
Types of NoSQL Database	74
CAP Theorem	75
ACID vs BASE	76
Future of Databases	77
Leading DBMS Venders.....	77
Cloud Storage.....	78
Object Orientation	78
Big Data	79

Data, Information & Knowledge

Data vs Information

Data

- Just fact stored and recorded
 - Numbers
 - Dates
 - Text
 - Images
 - Sound
 - Video

Information

- Data in context
 - Can be summarised
- Data that has been processed to increase its value, and user's knowledge about that data
- Makes data more useful
- Examples:
 - Put raw data in tables
 - Put raw data in charts

Metadata

- Data about data
 - i.e. Definitions of each column in a table
- Can include:
 - Structure
 - Rules
 - Constraints
- Why is metadata needed?
 - Consistency
 - Clarity
 - Meaning
- Metadata standards
 - Use XML, HTML, etc.
- **Data dictionaries** are generated as part of analysis of system requirements
 - Contains metadata

What is a Database?

- Collection of logically related data and its description
 - Usually in tables, spreadsheets

Database Management System (DBMS)

- A software system that enables users to define and control the database
 - Need to define, create, maintain and control database access
- Provides specific capabilities:

- **Query Language**
 - Data definition language (DDL): define and set up database
 - Data manipulation language (DML): maintain, alter and use database
 - Data control language (DCL): control access to the database
- **Other capabilities**
 - Integrity of data
 - Concurrency (shared access)
 - Recovery
 - Data descriptions

File Processing Systems

- Old method:
 - Multiple computers for each department, each contain flat file data of products and/or customer information
 - Problems:
 - Program-data dependence
 - If file structure changes, so does the program
 - Program knows too much low-level data structure
 - Duplication of data
 - Each department independently has to keep track of customers and products
 - Wasteful, inefficient, loss of integrity
 - Loss of metadata integrity
 - Limited data sharing
 - Data is tied to application in each department
 - Lengthy development times
 - Application has to do low-level data management, figuring out the file format each time
 - Excessive program maintenance
 - Up to 80% of development time is used for maintenance
- New method: Database systems
 - Advantages:
 - Data independence
 - Separation of data and program
 - Central data repository
 - Minimal data redundancy
 - Redundancy can be control using normalisation
 - Improve data consistency
 - Single store of data means to disagreement or update problems
 - Improve data sharing
 - Data is shared, not necessary for a single application
 - External users can be allowed access
 - Reduce data maintenance
 - Data structure can change without application changing
 - Increased productivity of application development
 - Data already collected and structures already known
 - DMBS have many tools to help access and maintain the data
 - Enforcement of standards
 - Centralised data management
 - Documented policy for data management
 - Data definition and dictionary for metadata

- Improved data quality
 - Constraints built into the database
 - Scrubbing data if necessary
- Data access “without programming”
 - SQL