**Introduction to Databases**

1. **The Modern Data Challenge**
2. **Data can be divided into three major categories**

* **Structured Data**
* **Relational Database Management Systems require data to be stored in a very structured way.**

**关系数据库管理系统要求以非常结构化的方式存储数据。**

* **These systems deal with data that has a repetitive pattern or format**

**这些系统处理具有重复模式或格式的数据**

* **Consider Student data stored in a University. While every student is different, the university want to store data in the same format for every student. Data Types are also specified for each piece of information**
* **Student ID-Numeric/Digits**
* **HomeAddress-Alphanumeric**
* **Student Name-Alpha**
* **Gender-Alpha**
* **DateOfBirth-Date**
* **PhoneNo-Numeric or Digits+space + brackets**
* **NextOfKin-Alpha**
* **Semi Structured Data**
* **Semi-structured data is information that doesn't match the requirements of a relational database.**

**半结构化数据是不符合关系数据库需求的信息。**

* **The data is organized / arranged that makes it easier to analyze.**
* **Examples of semi-structured data include XML documents and NoSQL databases.**
* **We will briefly deal with the topic of semi-structured data in future weeks.**
* **Unstructured Data**
* **Unstructured data is not organised in a pre-defined manner.**

**非结构化数据不以预定义的方式组织。**

* **The organization does not know the format, nor the content of the data in advance.**
* **Consider data sourced from social media, email, etc. The contents are unpredictable.**

**考虑来自社交媒体、电子邮件等的数据。内容是不可预测的。**

1. **Why use a DBMS?**

* **Data independence**
* **Applications should not be exposed to data representation and storage**

**应用程序不应该向数据表示和存储公开**

* **DBMS provides an abstract view that hides representation & storage**

**DBMS提供了一个抽象的视图，它隐藏了表示和存储**

* **Efficient access**
* **More efficient data storage and retrieval than flat files**

**比平面文件更有效的数据存储和检索**

* **Data integrity and security**
* **DBMS enforces data integrity constraints, access controls and govern user access**

**DBMS执行数据完整性约束、访问控制和管理用户访问**

* **Not reliant on just the operating system不仅仅依赖于操作系统**
* **Uniform data administration统一的数据管理**
* **Specialist skills in data management and administration**
* **Layer of expertise reduces risk to data and data owners**

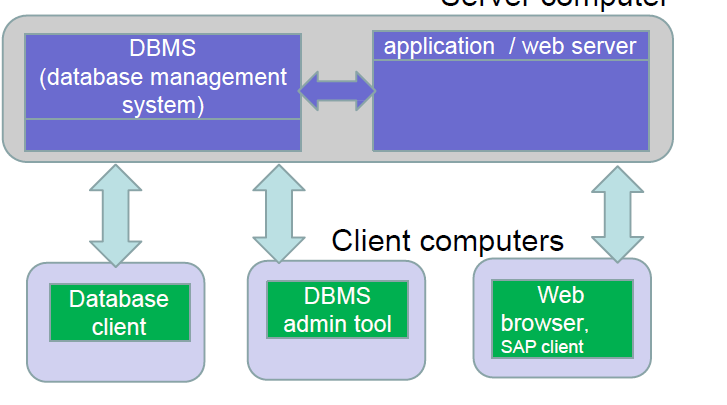
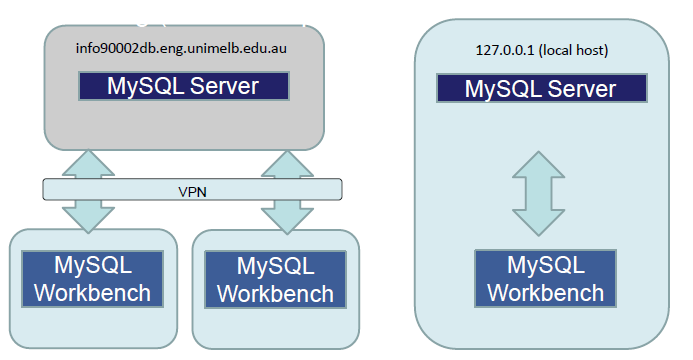
**专业知识层降低了数据和数据所有者的风险**

* **Concurrent access and crash recovery并发访问和崩溃恢复**
* **Schedules concurrent access. Protects data from system failures**

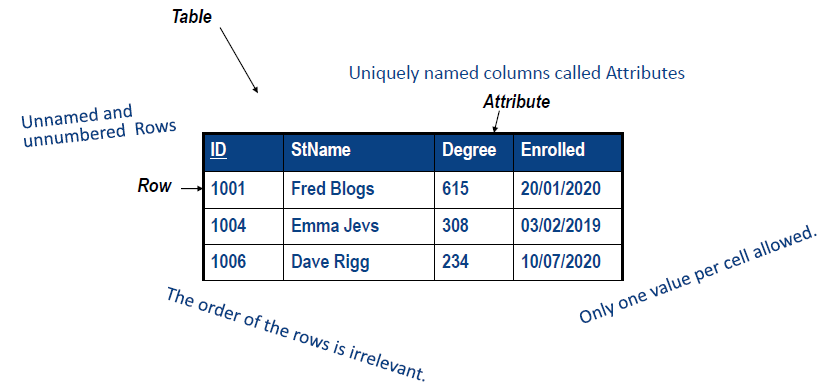
**日程安排并发访问。保护数据不受系统故障的影响**

* **Datasets increasing in diversity and volume数据集的多样性和体积不断增加**
* **Data Independence**
* **Logical data independence: Protection from changes in logical structure of data.** **逻辑数据独立性:防止数据逻辑结构变化。**
* **Physical data independence: Protection from changes in physical structure of data物理数据独立性:保护数据不受物理结构变化的影响**
* **Reduced Application Development time**

1. **Client-Server Architecture**
2. **Client Server-In Industry**

* ****
* ****

1. **Relational databases**
2. **RDBMSs are based on the Relational Data Model基于关系数据模型**

* **Data is represented in the form of two dimensional tables**
* **Each two dimensional table has the following properties（属性）:**
* **A set of uniquely named Columns / Attributes**
* **A list of unnamed/unnumbered Rows**
* **The order of the rows is irrelevant.**
* **A Row consists of a sequence of Attributes**
* **One cell for each Attribute**
* **Only one value per cell is allowed.** **每个单元格只允许一个值**
* ****
* **A relational database is a collection of related tables**
* **Student Table (stores data about students)**
* **Subject Table (stores data about university subjects)**
* **Enrolment Table (stores data about a student has enrolled into a specific subject)**

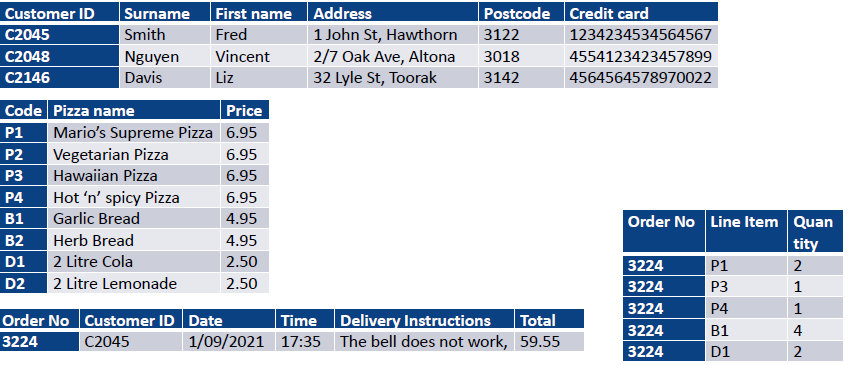
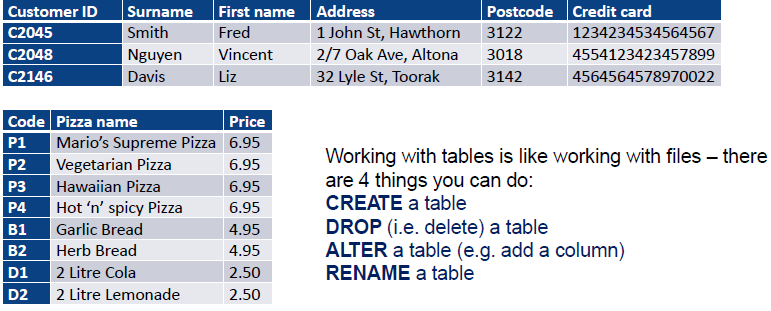
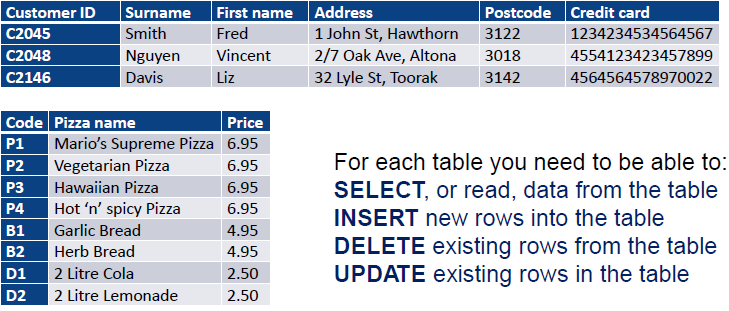
1. **RDBMS**

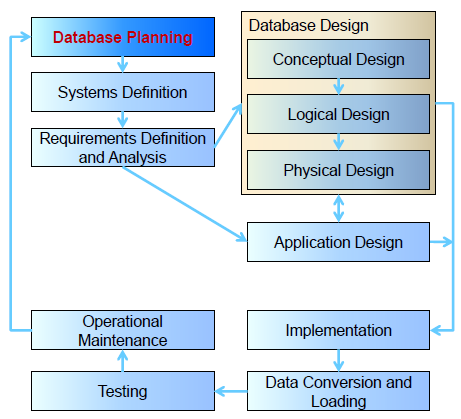
* **A RDBMS is a collection of programs that allow developers / users to store & retrieve data from relational databases**

**RDBMS是一组程序的集合，允许开发人员/用户从关系数据库中存储和检索数据**

* **It allows users to perform CRUD (create, read, update and delete) operations on data in the tables. E.g.** **它允许用户对表中的数据执行CRUD(创建、读取、更新和删除)操作。**
* **Create a student record**
* **Retrieve the student's details检索学生的详细信息**
* **Update the student's details**
* **Delete the student from a table**

1. **Setting up a RDBMS**

* **Tables**
* **Follow a 2 dimensional structure**
* **Each row of data is identified by a unique Primary Key**
* **No duplicates, e.g. Student ID**
* **Constraints can be added to validate data可以向验证数据添加约束**
* **Student ID is correct length**
* **Student type is PG or UG (post or undergraduate)**
* **Student is enrolled in a degree that actually exists**
* **Case Study-Overview**
* **Let's consider a business called Pizza OnLine**
* **It allows customers to order pizzas via the online portal.**
* **Customers have pizzas delivered.**
* **Only people who register with a credit card number can be an on line customer.**
* **Tables in the Pizza Online Database**
* ****
* ****
* ****
* **Database lifecycle**
* **Design the database**
* **data modelling E-R diagrams**
* **Implement the database**
* **data definition language DD L**
* **Data access / programming**
* **data manipulation language DML数据操纵语言**
* **Database administration**
* **data control language DCL**

1. **Database Development Lifecycle**
2. ****

* **Database Planning**
* **Planning how to do the project.**
* **How does the enterprise work**
* **Enterprise data model**
* **How can the stages be completed efficiently and effectively.**
* **Systems Definition**
* **Specifying scope and boundaries指定范围和边界**
* **Users**
* **Major user views**
* **Application areas**
* **How does it interact with other systems它如何与其他系统交互**
* **User views how the system operates from differing perspectives**
* **Requirements Analysis**
* **Collection and analysis of requirements for the new system收集和分析新系统的需求**
* **Conceptual Design**
* **High level, first pass model of entities and their connections**

**实体及其连接的高级、第一次传递模型**

* **Typically omits attributes\***
* **Could potentially be implemented in a non-relational database**

**可以在非关系数据库中实现**

* **Thus can include many to many relationships, repeating groups, composite attributes因此可以包含多对多关系、重复组、复合属性**
* **Logical Design**
* **Builds on the conceptual design**
* **Designing now for a relational database**
* **Includes columns and keys**
* **Independent of a specific vendor and other physical considerations**

**独立于特定供应商和其他物理考虑**

* **Physical Design**
* **Implements the logical design for a specific DBMS.**

**实现特定DBMS的逻辑设计**

* **Describes:**
* **Base tables**
* **Data types**
* **Indexes**
* **Integrity constraints完整性约束**
* **File organization**
* **Security measures**
* **Application Design**
* **Done in conjunction with database design 配合数据库设计完成**
* **Design of the interface and application programs that use and process the database设计使用和处理数据库的接口和应用程序**
* **Implementation**
* **Implementation of the design as a working database**
* **Data Conversion and Loading数据转换和加载**
* **Transfer existing data into the database将现有数据转移到数据库中**
* **Conversion from old systems从旧系统转换**
* **Non trivial task非琐碎的任务**
* **Testing**
* **Running the database to find errors in the design / setup**

**运行数据库以查找设计/设置中的错误**

* **Other issues also**
* **Performance**
* **Robustness稳健性**
* **Recoverability可恢复性**
* **Adaptability适应性**
* **Security**
* **Operational Maintenance**
* **The process of monitoring and maintaining the database following its commissioning数据库调试后的监控和维护过程**
* **Monitoring and improving performance**
* **Handling changes to requirements处理需求的变更**