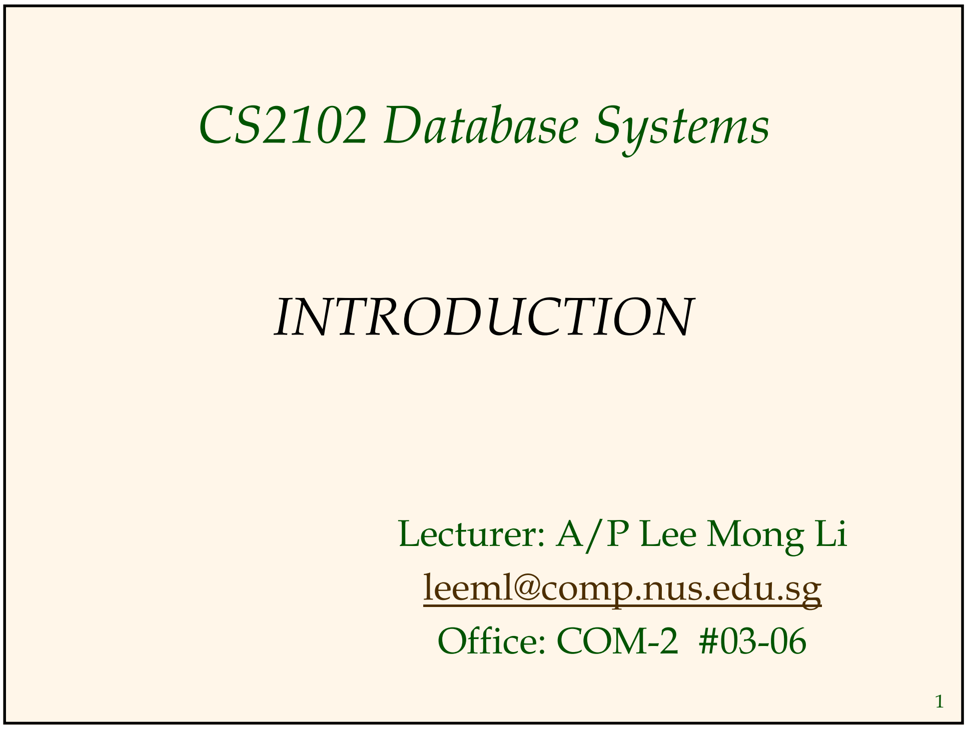
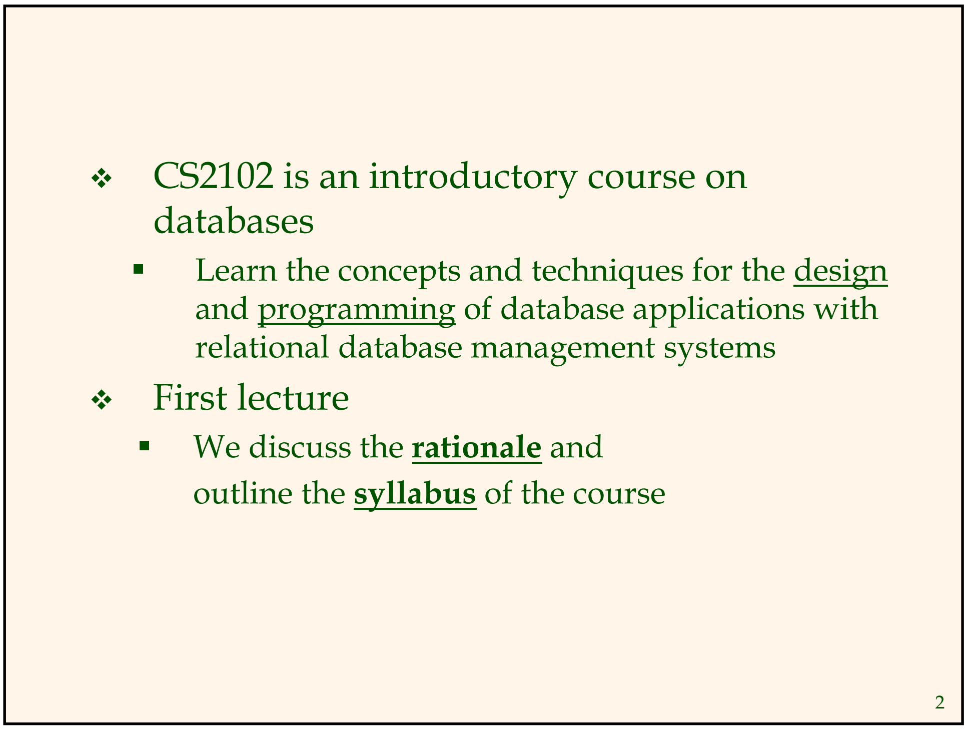


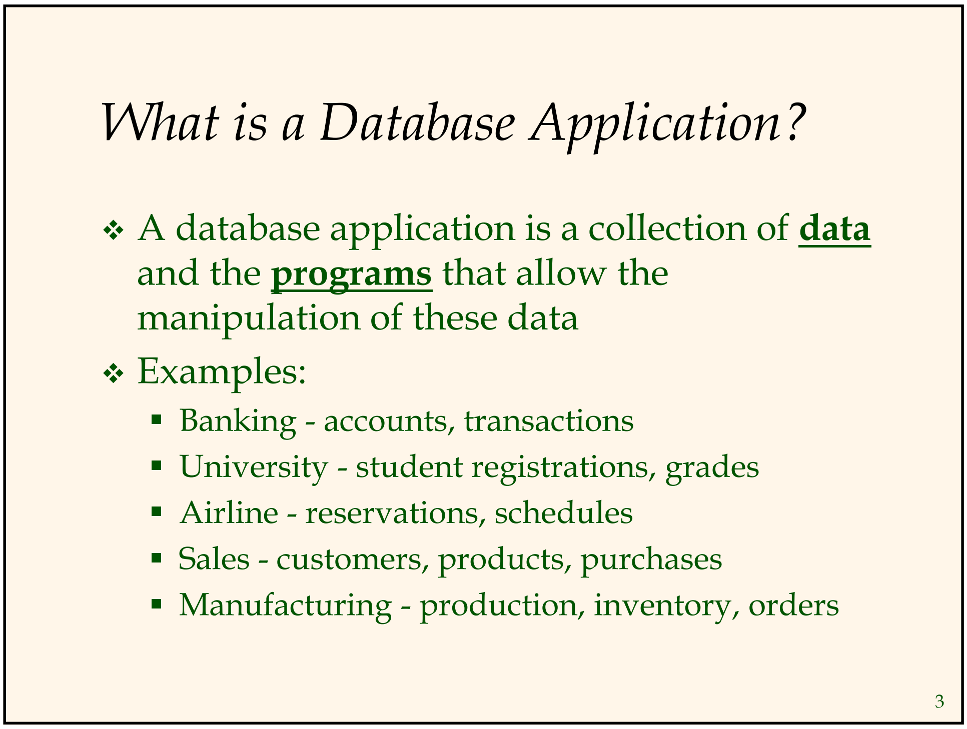
CS2102 Database Systems

INTRODUCTION

Lecturer: A/P Lee Mong Li
leeml@comp.nus.edu.sg
Office: COM-2 #03-06

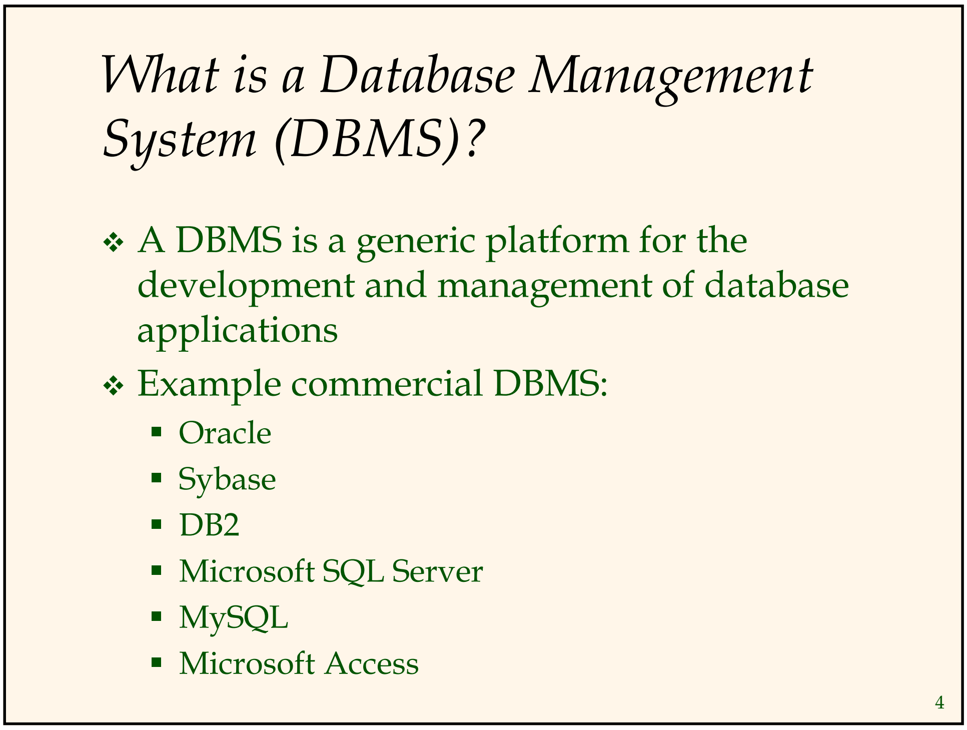


- CS2102 is an introductory course on databases
  - Learn the concepts and techniques for the design and programming of database applications with relational database management systems
- First lecture
  - We discuss the **rationale** and outline the **syllabus** of the course



What is a Database Application?

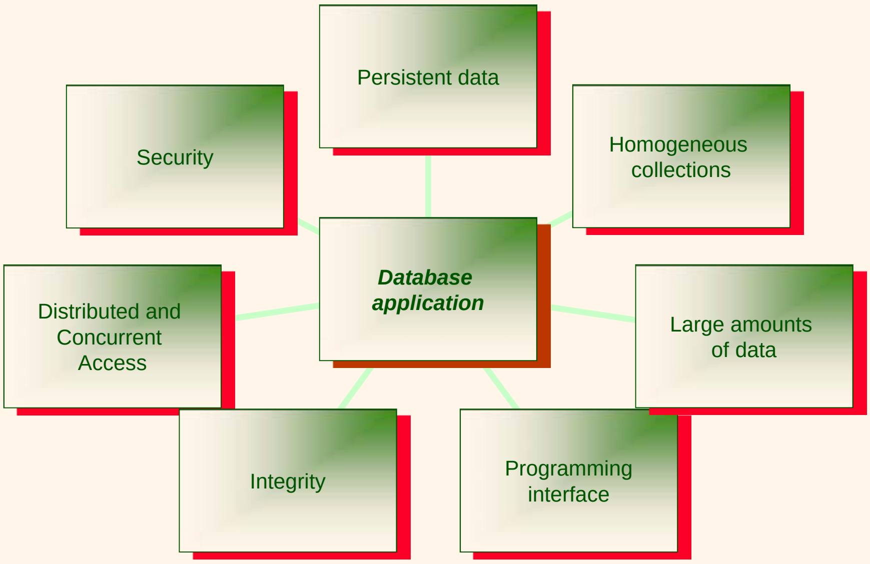
- A database application is a collection of **data** and the **programs** that allow the manipulation of these data
- Examples:
  - Banking - accounts, transactions
  - University - student registrations, grades
  - Airline - reservations, schedules
  - Sales - customers, products, purchases
  - Manufacturing - production, inventory, orders



What is a Database Management System (DBMS)?

- A DBMS is a generic platform for the development and management of database applications
- Example commercial DBMS:
  - Oracle
  - Sybase
  - DB2
  - Microsoft SQL Server
  - MySQL
  - Microsoft Access

Features of a DBMS



Persistent data

Security

Homogeneous collections

Database application

Distributed and Concurrent Access

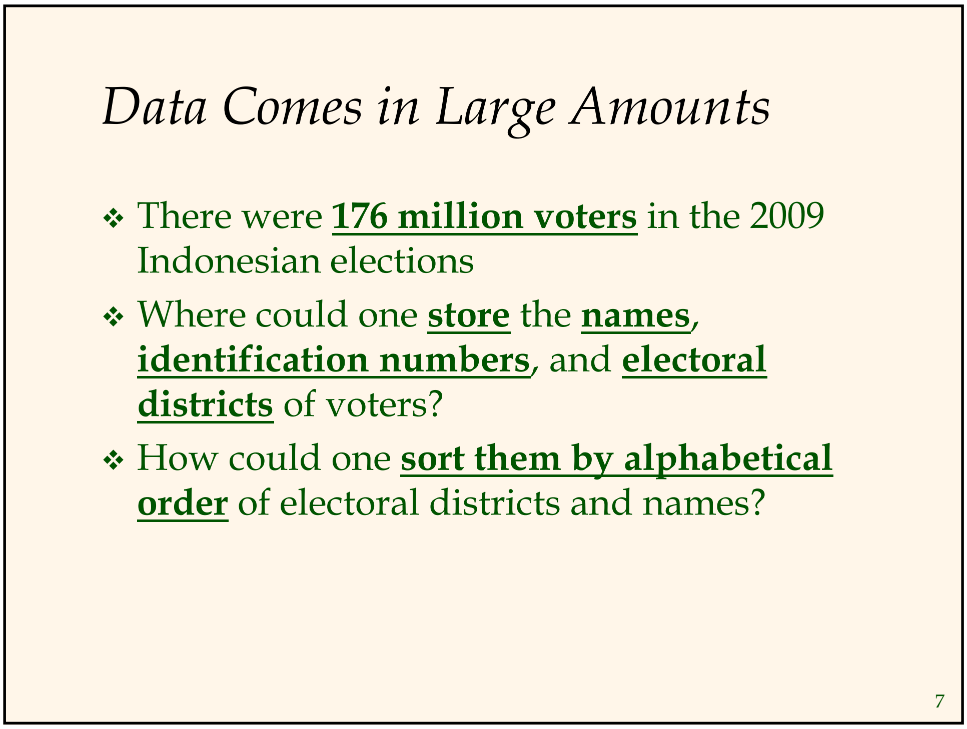
Large amounts of data

Integrity

Programming interface

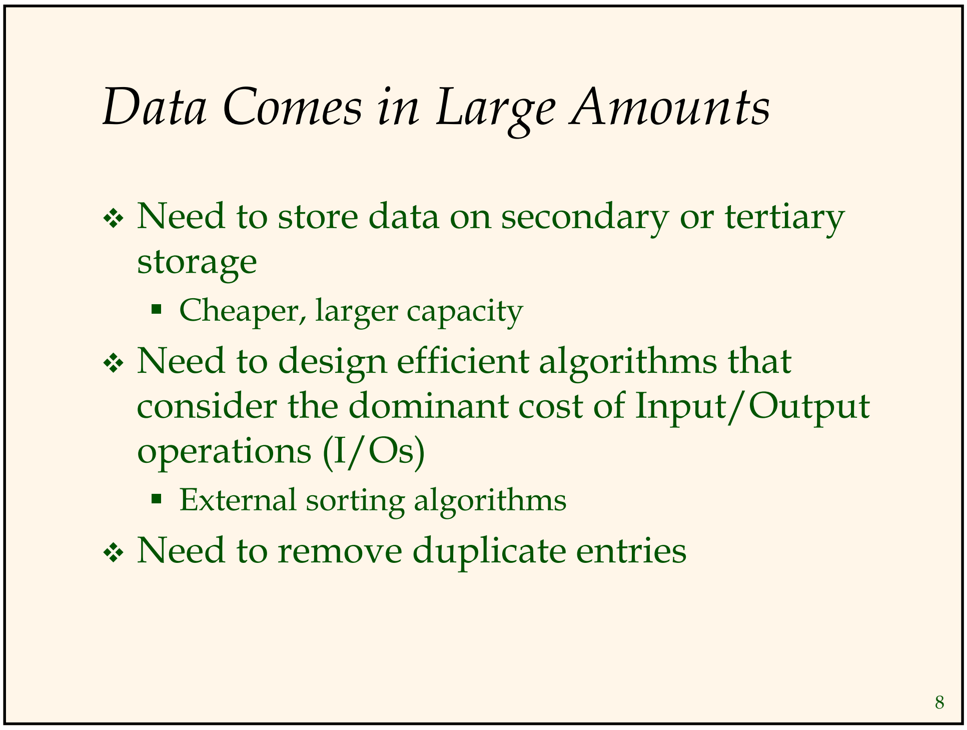
Data must Persist

- How can data **survive** the process that created it, and be **reused** by other processes?
- **Primary** memory is **volatile**
- **Secondary** and **tertiary** memories are **persistent**



Data Comes in Large Amounts

- There were **176 million voters** in the 2009 Indonesian elections
- Where could one **store** the **names**, **identification numbers**, and **electoral districts** of voters?
- How could one **sort them by alphabetical order** of electoral districts and names?



Data Comes in Large Amounts

- Need to store data on secondary or tertiary storage
  - Cheaper, larger capacity
- Need to design efficient algorithms that consider the dominant cost of Input/Output operations (I/Os)
  - External sorting algorithms
- Need to remove duplicate entries

Data Comes in Homogeneous Collections

Data Comes in Homogeneous Collections

- Structured data
- DBMS implements access methods and indexing for efficient storage, update and retrieval

Integrity of Data should be Maintained

- How to maintain the integrity of data in spite of possible **application**, **system**, or **media failures**?
- Restore data to a consistent state after failures
- A **consistent state** of the database is a state which complies with the business rules as defined by **integrity constraints**
  - E.g. "students who have not passed CS2102 cannot take CS3223"

Integrity of Data should be Maintained

- Recovery
  - **Atomicity:** all actions in a **transaction** happen or none happen
  - **Durability:** effects of successful transactions last

What is a Transaction?

- A **transaction** is a **logical unit of work** carried out by a user or an application
- Examples:
  - Booking of vacation
    - A transaction involves booking flight tickets, land transfers and hotel rooms
  - Transfer of money from one bank account to another
    - A transaction involves withdrawing the amount from the first account and depositing it to the second account

Distributed and Concurrent Access

- How can data be **shared** by users and processes that are possibly **distributed** over a network?

- Ensure **consistent** data access and updates

Distributed and Concurrent Access

- Concurrency control
  - **Isolation:** Transactions can be understood independently from each other
  - **Consistency:** If individual transactions would leave the application in a consistent state, a concurrent execution should do the same

Security and Access Control of Data is Critical

- How to protect the data and define **control access** to data?
- Prevent unauthorized data access
- DCL (**Database Control Language**) include statements to administer access privileges and transactions properties

Describing Data in a DBMS

- A DBMS allows users to define and query data based on a data model
- A **data model** is a collection of concepts for describing data
- A **schema** is a description of the structure of a database using a data model
- A **schema instance** is the content of the database at a particular time

Relational Data Model

- Most DBMSs today are based on the relational data model
- RDMBS vendors: IBM, Microsoft, Oracle, Sybase
- Data is modeled using **relations**
- A relation is a table with rows and columns

Querying in a Relational DBMS

- A DBMS provides a database language for users to retrieve data
- Formal query languages
  - **Relational algebra** (based on operators for manipulating relations)
  - **Relational calculus** (based on mathematical logic)
- Commercial database languages
  - **S**tructured **Q**uery **L**anguage (most widely used)
  - **Q**uery **B**y **E**xample (graphical)

To Summarize

A database application manages homogeneous collections of large amounts of persistent data that are shared among distributed users and processes, and whose integrity and security must be maintained.

Syllabus

- **Database Design**
  - Entity Relationship Model
  - Relational Model
  - Normalisation with Functional Dependencies
- **Database Programming**
  - Theory of Query Languages: Algebra and Calculus
  - SQL
  - SQL and Programming Languages

Texts & References

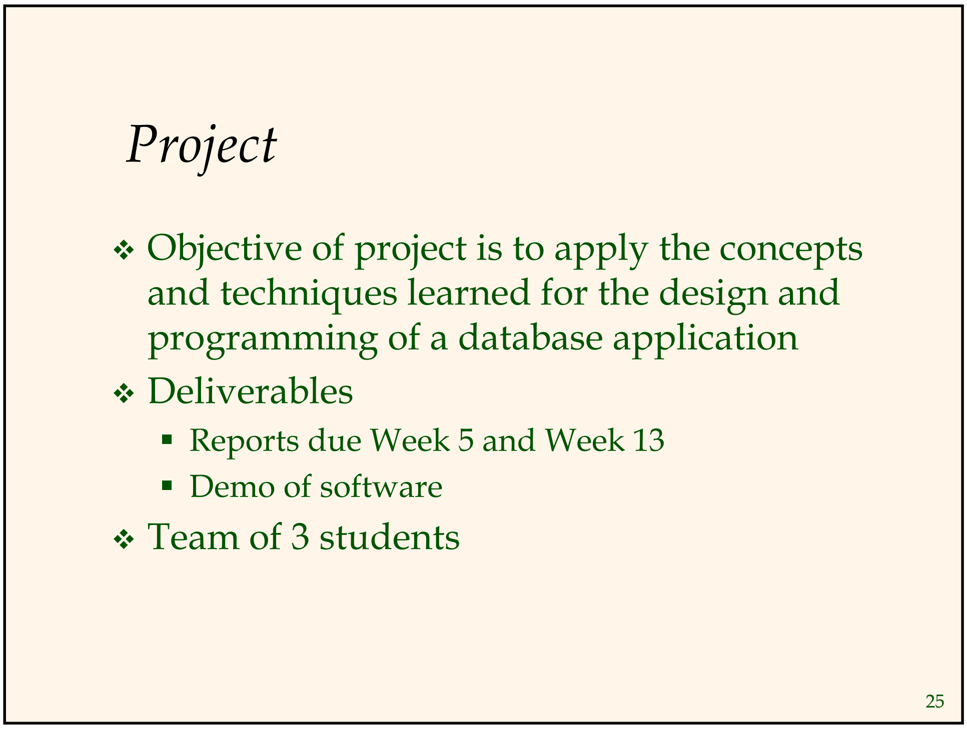
- Database Management Systems
  by R. Ramakrishnan and J. Gehrke
  McGraw-Hill, 3rd Edition, 2000
- A First Course in Database Systems
  by J. Ullman and J. Widom
  Prentice-Hall, 2nd Edition, 2002
- Introduction to Database Systems
  by S. Bressan, B. Catania
  McGraw-Hill, 2005

Course Schedule

- Lectures
  - Tuesday, 10 am – 12 pm
  - LT19
- Tutorials and Labs
  - Mondays and Thursdays, 2 hours
  - Start on Week 3

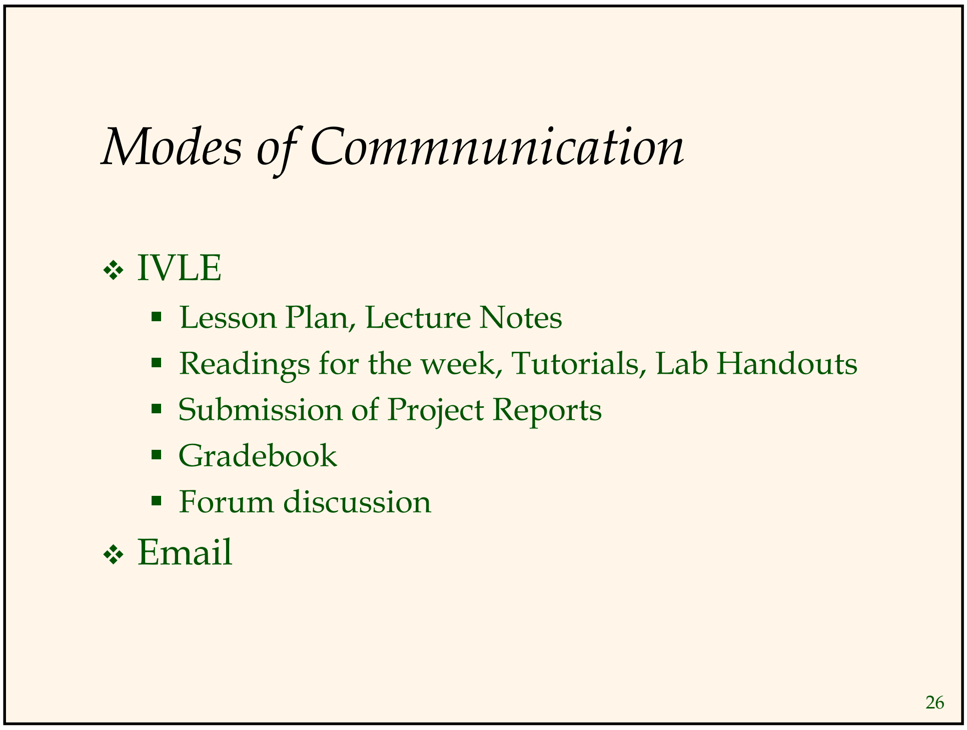
Assessments

- Final Exam (60%)
- Midterm Test (20%)
- Quiz (10%)
- Project (10%)



Project

- Objective of project is to apply the concepts and techniques learned for the design and programming of a database application
- Deliverables
  - Reports due Week 5 and Week 13
  - Demo of software
- Team of 3 students



Modes of Commnunication

- IVLE
  - Lesson Plan, Lecture Notes
  - Readings for the week, Tutorials, Lab Handouts
  - Submission of Project Reports
  - Gradebook
  - Forum discussion
- Email