UNIVERSITY OF
Southampton

UNIVERSITY OF Southampton

Introduction

COMP3211 Advanced Databases

Dr Nicholas Gibbins - nmg@ecs.soton.ac.uk

Lecturers

Dr Nicholas Gibbins
nmg@ecs.soton.ac.uk

Dr George Konstantinidis
g.konstantinidis@soton.ac.uk

Module Aims and Objectives

- Gain a better understanding of the nature of data
- Understand the issues to be addressed in writing database software
- Understand the variety of approaches taken so far
- Be able to select an appropriate database for an application
- Be aware of the latest developments in the use and application of databases

Learning Outcomes

You will be able to demonstrate knowledge and understanding of:

- The internals of a database management system
- The issues involved in developing database management software
- The variety of available DBMS types and the circumstances in which they're appropriate

Learning Outcomes

You will be able to:

- Choose appropriate approaches for data storage and access
- Demonstrate how a DBMS processes, optimises and executes a query
- Identify issues arising from concurrent or distributed processing and select appropriate approaches to mitigate those issues
- Select an appropriate DBMS for an application
- Implement components of a DBMS

Prerequisites: COMP1204

- The role of database systems in information management
- The concept of data modelling
- Entity-Relationship modelling
- The relational model and other models
- SQL
- Database management issues

COMP3211 vs COMP1204

In COMP1204, you learned how to build **databases**

In COMP3211, you will learn how to build **database management systems**

Course Structure

Three lectures per week:

- Monday 1700
- Tuesday 1500
- Wednesday 1200

All lectures will be online in Blackboard Collaborate; specific sessions have been set up for each lecture.

Assessment

- 75% examination (120 minutes, 3 questions from 5)
- 25% coursework (due Wednesday 14 April)

Books

Core Text

- Garcia-Molina H., Ullman J.D. and Widom J., Database Systems: The Complete Book, 2nd ed., Pearson, 2009.
  - Parts IV and V are the basis of this module

Background Texts

- Elmasri R. and Navathe S.B., Fundamentals of Database Systems, 6th ed., Addison-Wesley, 2010.
- Connolly T. and Begg C., Database Systems, 5th ed., Addison-Wesley, 2009.
- Date C.J., An Introduction to Database Systems, 8th ed., Pearson, 2004.

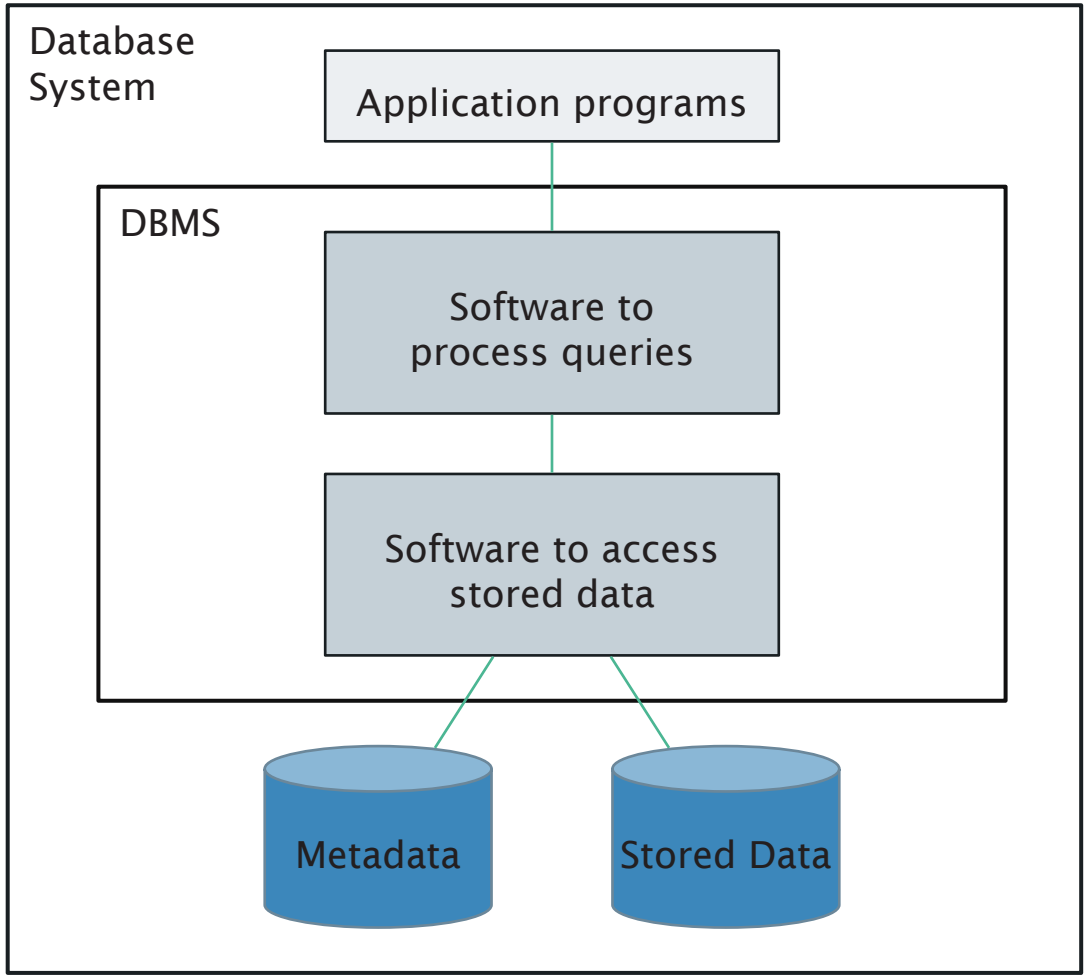
UNIVERSITY OF Southampton

Database Management Systems

What is a Database?

- Represents some aspect of the real world
- A logically coherent collection of data with some inherent meaning
- Designed, built and populated with data for a specific purpose
- Has an intended group of users and some preconceived applications in which these users are interested

Database System vs. DBMS



Database Management System

A DBMS is a set of general purpose software, that allows the user to:-

- Define the database
  - Specifying the data types, structures and constraints for the data to be stored
- Construct the database
  - Store the data on some storage medium that is controlled by the DBMS
- Manipulate the database
  - Querying to retrieve specific data, updating to reflect changes in the model of the real world, and generating reports from the data

What should the DBMS do?

- Store data (!)
- Control or eliminate redundancy
- Promote *program-data independence*
- Permit multiple views of the data
- Support sharing by multiple users
- Support sharing and integration of data between multiple applications
- Control concurrent access to data

What should the DBMS do?

- Offer various interfaces for data retrieval and manipulation
- Be self-describing / contain its own catalogue for metadata
- Support data abstraction
- Allow complex relationships between objects to be represented
- Enforce integrity constraints on the data
- Restrict unauthorised access
- Facilitate backup and recovery

UNIVERSITY OF Southampton

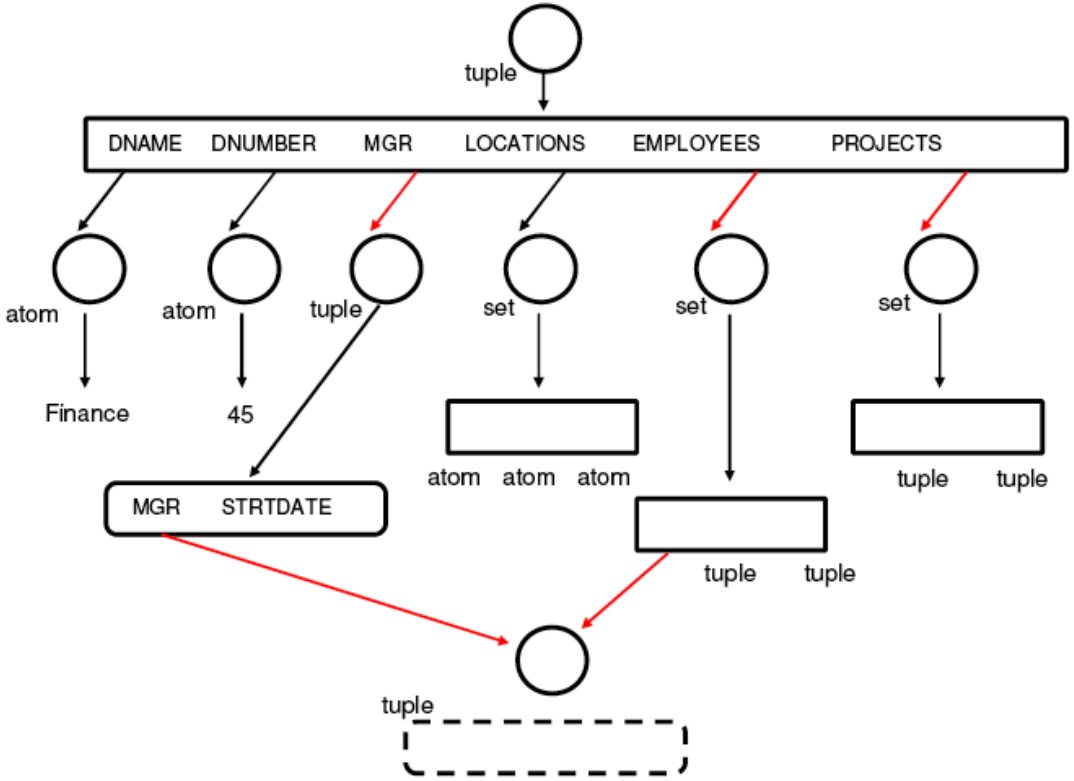
Topics

Datatypes

How does the type of data affect what we can do with it?

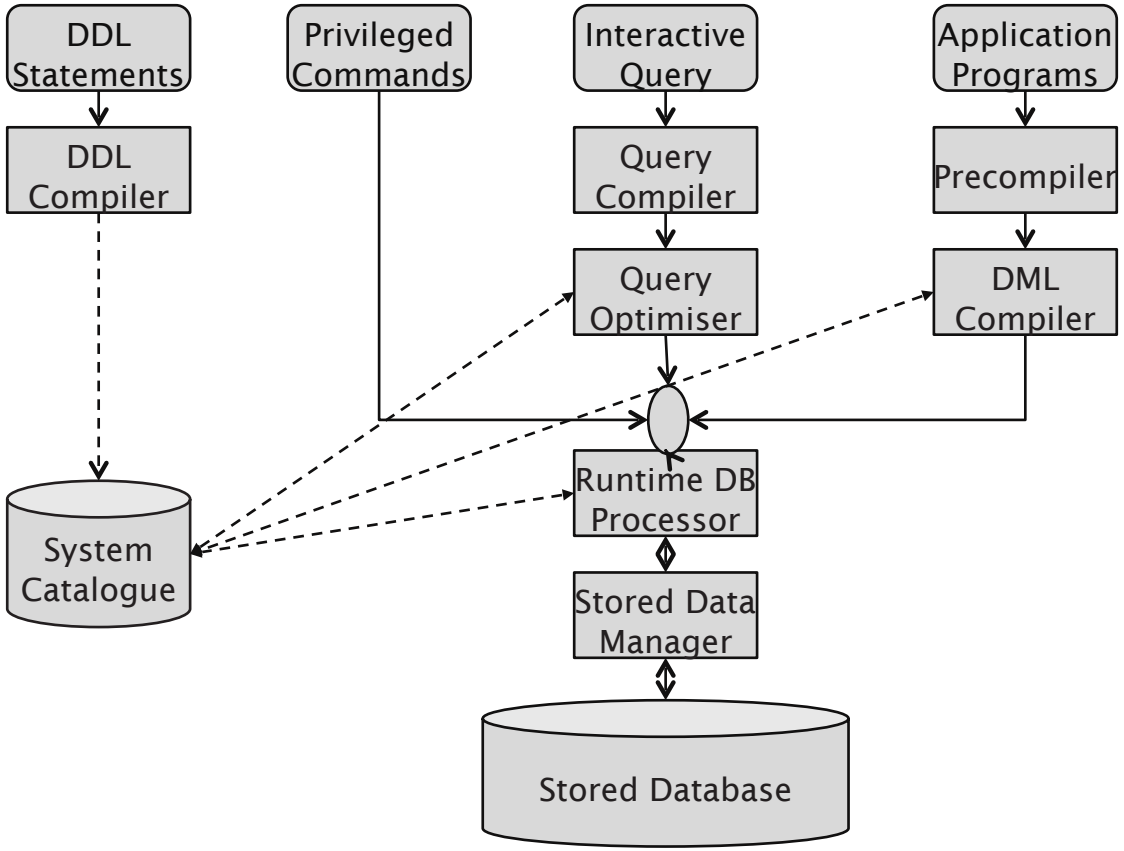
How do we model:

- Temporal data?
- Spatial data?
- Multimedia data?



DBMS Architecture

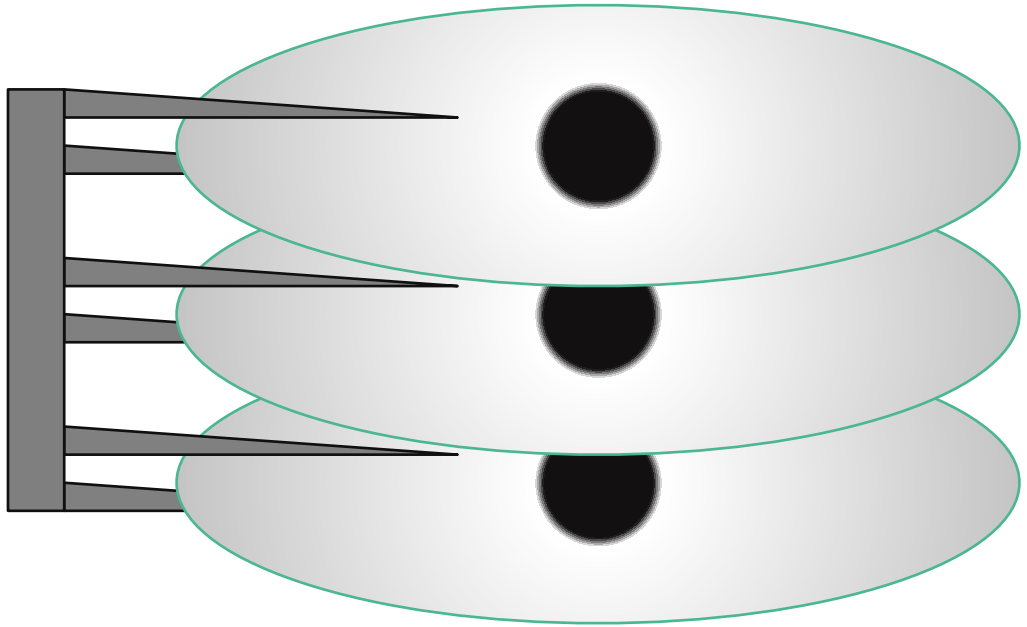
What are the functional units within a DBMS?



Data Storage

How does a DBMS organise data:

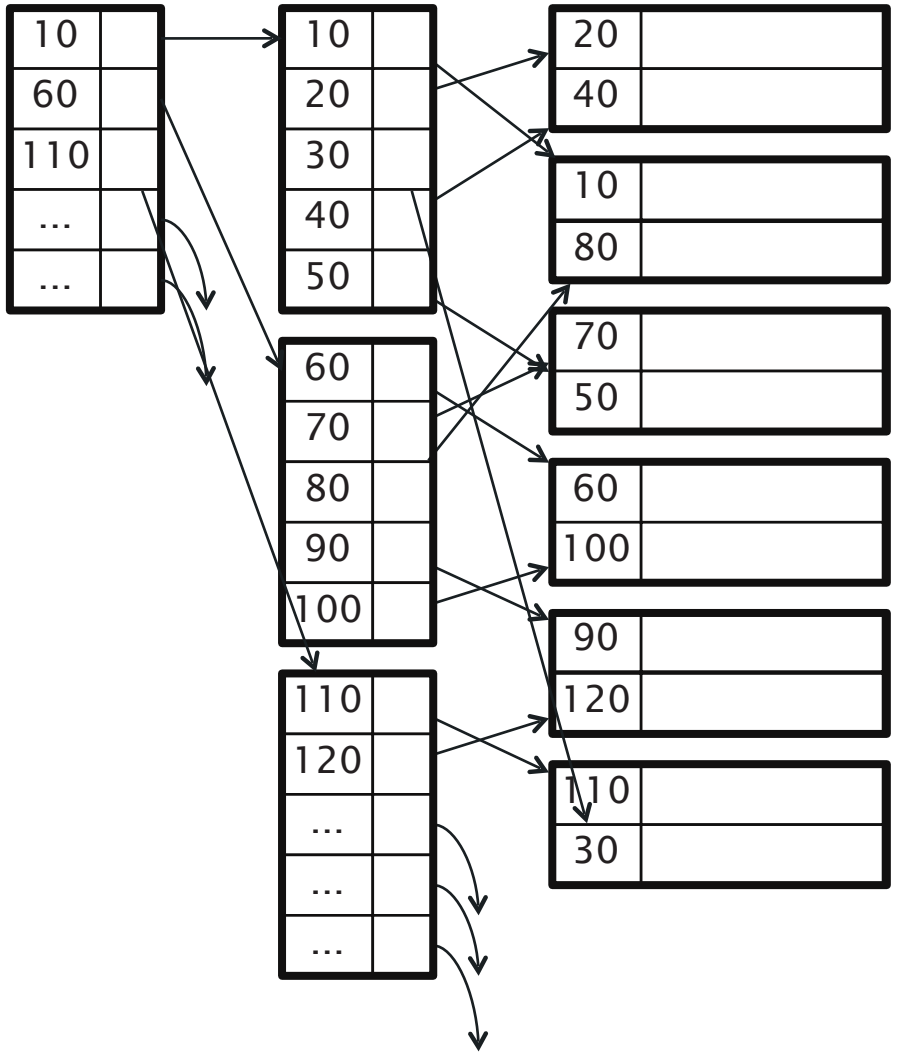
- On disc?
- In records?
- In fields?



Access Structures

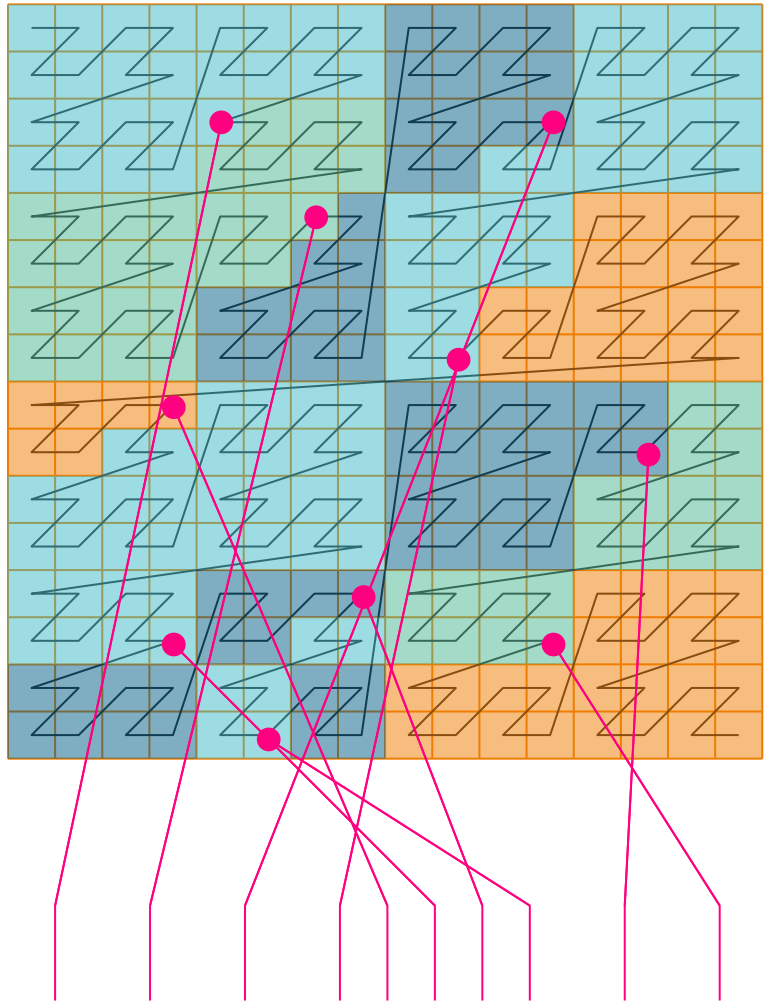
How can we improve the speed of access to data in a DBMS?

- Indexes, hash tables, B-trees



Multidimensional Access Structures

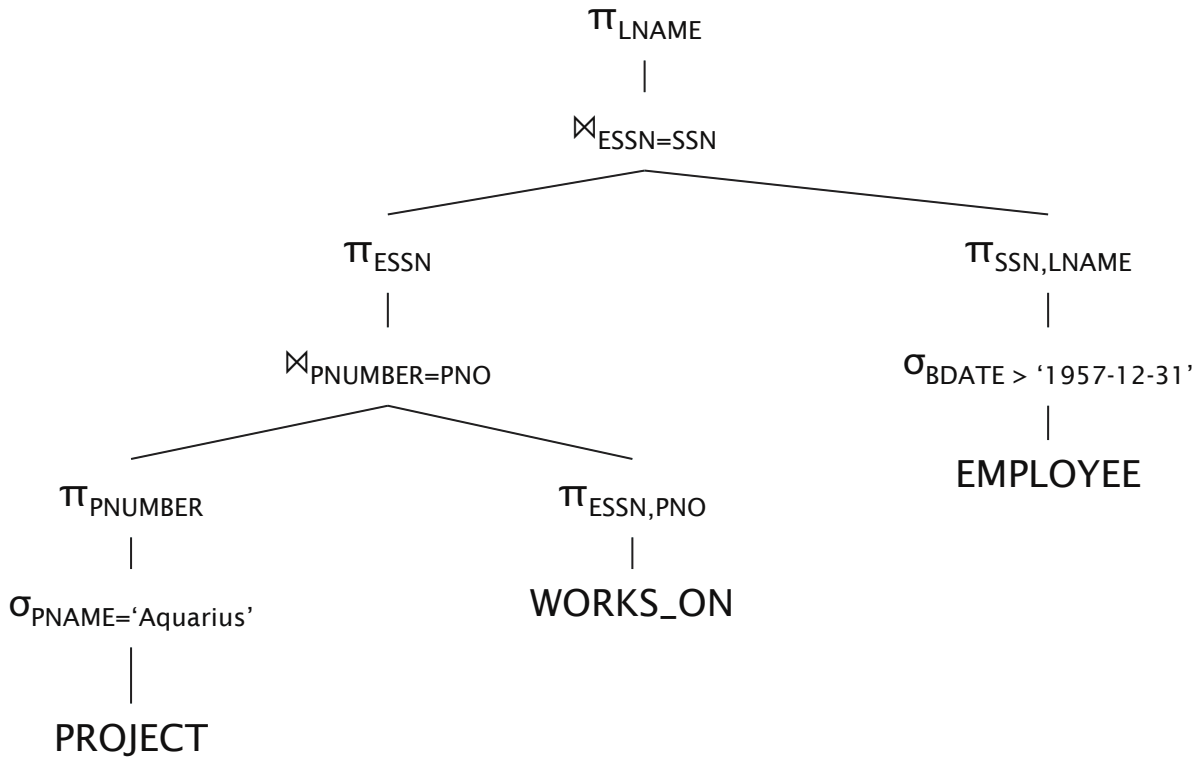
How do we improve the speed of access to multidimensional data in a DBMS?



Query Processing and Optimisation

How are queries executed in a DBMS?

How can we modify queries to reduce their execution time?

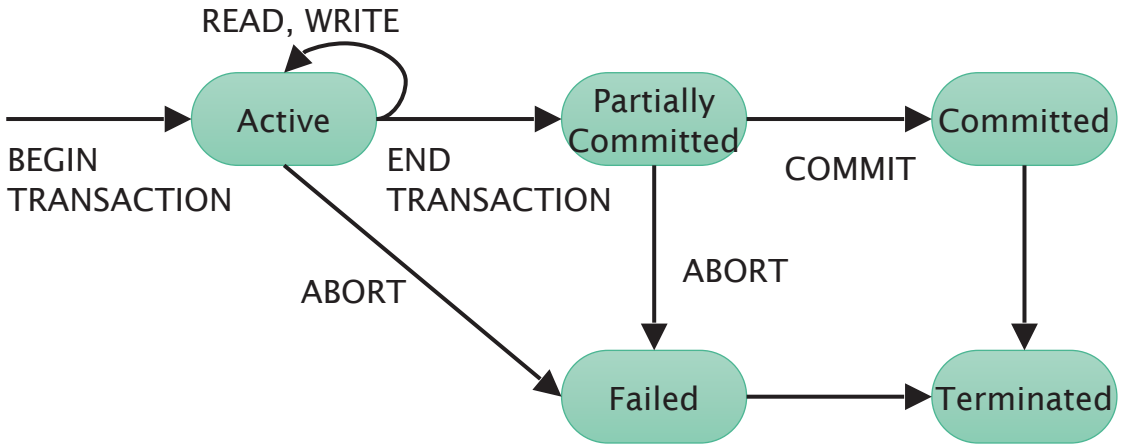


Transactions and Concurrency

How do we provide users with concurrent access to a DBMS?

What problems can arise?

How can we prevent or mitigate those problems?

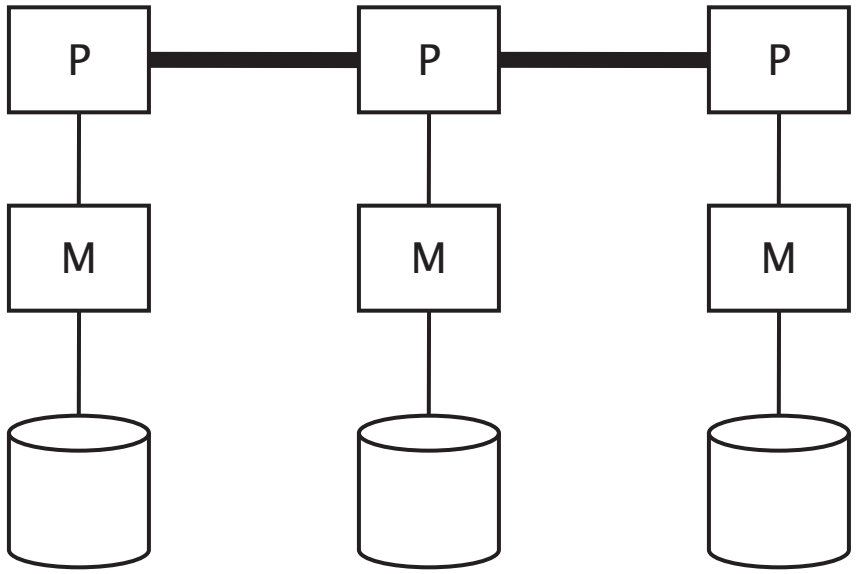


Parallel Databases

How can we distribute a DBMS across the machines in a cluster?

How does parallelism affect:

- Query processing?
- Deadlock detection?
- Reliability?

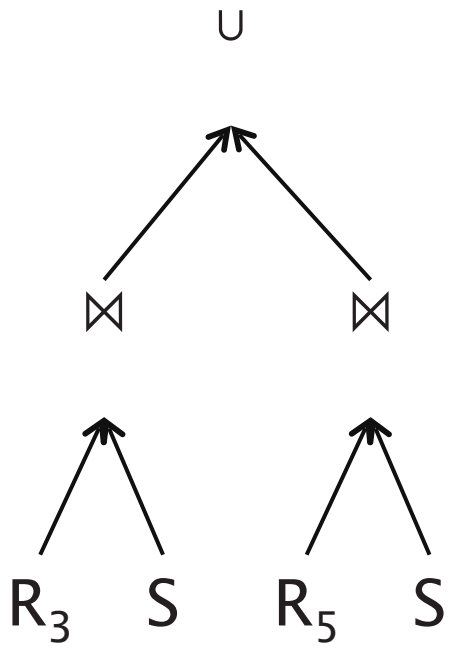


Distributed Databases

How can we distribute a DBMS across a WAN?

How does distribution affect:

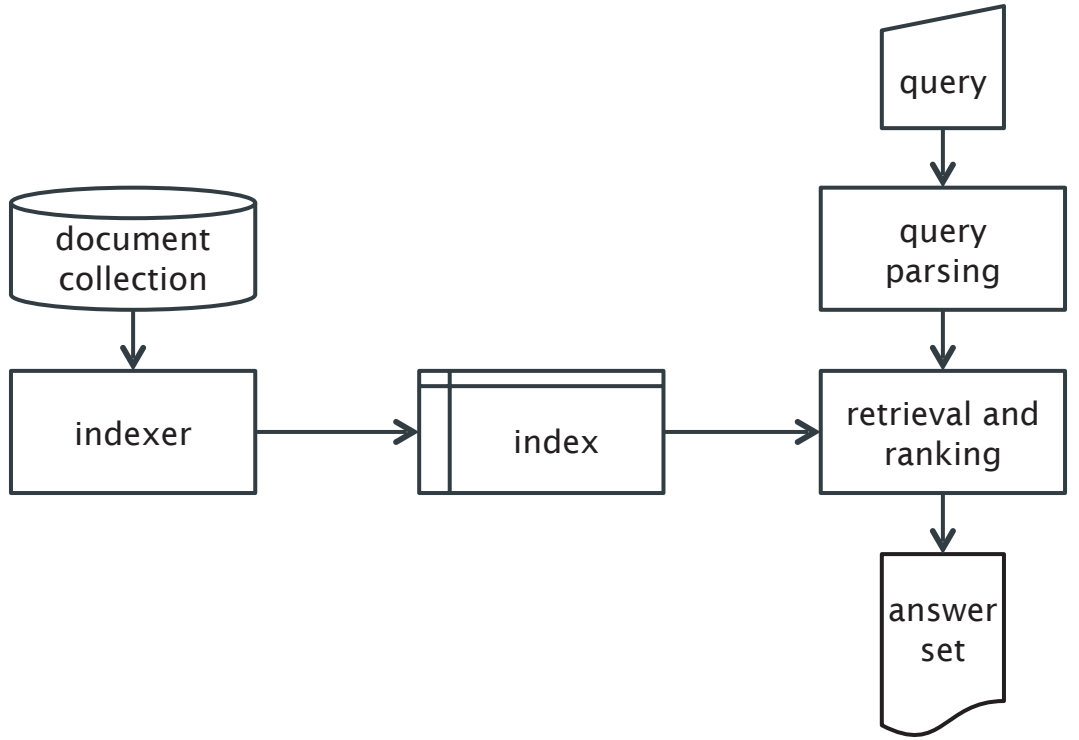
- Query processing?
- Concurrency control?
- Reliability?



Information Retrieval

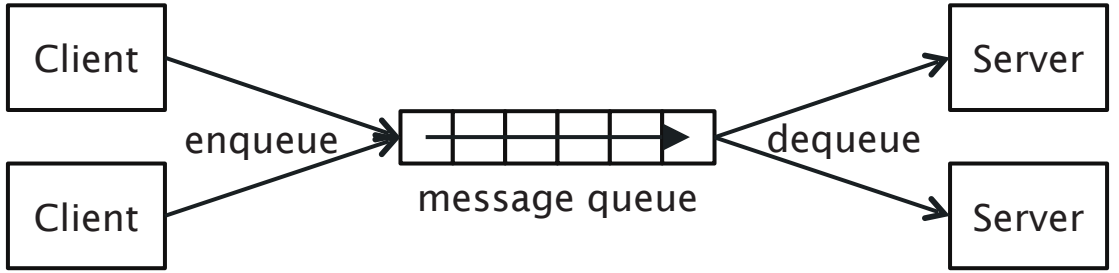
How do we support queries over free text data?

How do we evaluate the effectiveness of an IR engine?



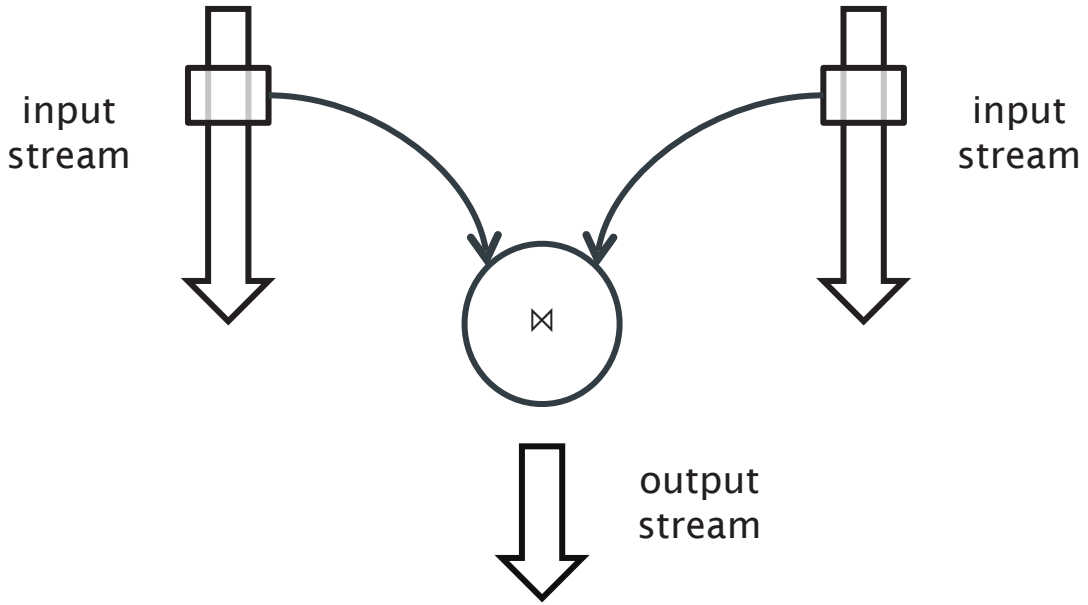
Message Queues

How can we use asynchronous communications for reliable distributed DB applications?



Stream Processing

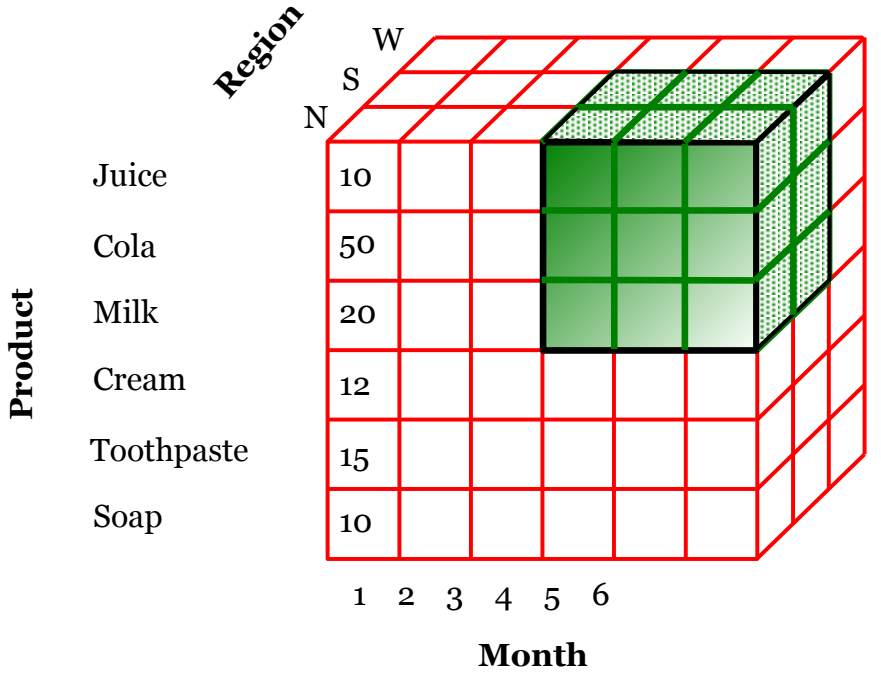
How can we query data when there's more data than we can store?



Data Warehousing

How can we best support the analysis of complex, multidimensional data?

- OLAP vs OLTP



Non-Relational Databases

What's out there apart from RDBMS?

- Hierarchical, XML
- Network, Object
- Graph
- NoSQL

UNIVERSITY OF Southampton

Next Lecture:
Data Types