Advanced Database Management Systems

Introduction

Alvaro A A Fernandes
School of Computer Science
University of Manchester

Abstract

This document introduces the Advanced Database Management Systems course unit. It aims to characterize what is meant by advanced database management systems, what learning style the course unit adopts and, finally, to present the raw data about the course unit in the context of the MSc degree programmes in the School of Computer Science at the University of Manchester.

Contents

1 Introduction

1.1 What do we mean by Advanced Database Management Systems?

Topic (1) ::
What are classical database management systems good at?

- Classical database management systems (DBMSs) have been very successful.
- In the last four decades, they have become an indispensable infrastructural component of organizations.
- They play a key role in reliably and efficiently reflecting the transaction-level unfolding of operations in the value-adding chain of an organization.
- Each transaction (e.g., an airline reservation, a credit card payment, an item sold in a checkout) is processed soundly, reliably, and efficiently.
- Effects are propagated throughout the organization.

Topic (2) ::
Where do classical database management systems come short?

- Classical database management systems (DBMSs) assume that:
  1. Data is structured in the form of records : e.g., itemSold(id,price,amount);

2. Only on-line transaction processing (OLTP) is needed : e.g., to update stocks;
3. Data and computational resources are centralized : e.g., at the administrative headquarters;
4. There is central control over central resources : i.e., a DBA-team monitors performance and makes planning adjustments;
5. There is no need for dynamically responding in real-time to external events : e.g., most processing is for different forms of bookkeeping and enterprise resource planning;
6. There is no need for embedding in the physical world in which organizations exist : i.e., the database is an abstract model of the day-to-day operations, but does not monitor and respond to physical conditions.

- This is too constraining for most modern businesses.
- Classical DBMSs support fewer needs of organizations than they used to.

Topic (3) ::
How are DBMSs evolving?

- Most cutting-edge research in databases is geared towards supporting:
  1. Un- and semi-structured data too : e.g., web pages, web forms, documents, text;
  2. On-line analytical processing (OLAP) too : e.g., summaries and trends [*Which branch has sold less than the average of all others in the region for the last six months?*];
  3. Distributed data and computational resources : e.g., branch-level sourcing of products based on branch supply and demand patterns;
  4. Absence of central control over distributed resources : e.g., country-level scope often forced by different legal regimes;
  5. Dynamic response in real-time to external events : e.g., in supply-chain coordination, the unexpected shortfall in some product may require agile responses without paying a premium to supplier;
  6. Embedding in the physical world in which the organization exists : e.g., in supply chain management, perishable goods, like leaf vegetables, are sensitive to weather and storage/transportation conditions.
- DBMSs that exhibit these capabilities are **advanced** in the sense used here.

Relevance (1) ::
Why do they matter?

OLAP/DM Companies need to make more, and more complex, decisions more often and more effectively to remain competitive.

Text-/XML-DBMSs The ubiquity and transparency of networks means data can take many forms, is everywhere, and can be processed anywhere.

Relevance (2) ::
Why do they matter?

Parallel/P2P/Distributed DBMSs For both data and computation, provision of resources is now largely servicized and can be negotiated, or harvested.

Stream DMSs Widespread cross-enterprise integration means that companies must be able to respond in real-time to events streaming in from their commercial and financial environment

Sensor DMSs Many companies are aiming to sense and respond not just to the commercial and financial environment but to the physical environment too.

Novelty ::
What am I going to learn?

- This course unit is an introduction to some of the latest, cutting-edge research in the DBMS area.
- It aims to survey the research landscape with a view to understanding how researchers are responding to the challenges DBMS technology is facing.
- It provides:
  - A brief overview of the internal architecture of classical DBMSs;
  - A discussion of some of the ways in which that architecture is evolving to support more needs of organizations.
- Note: the emphasis is on **research**, not on technology, and not on how organizations use advanced DBMSs.

Usefulness ::
Why will it matter to me?

- DBMS technology has evolved as the result of a pull from customers as well as push from research labs for the last 40 years.
- If you plan to go on to industry, the course unit enables you to understand what strategic impact current research will have on the way companies will be able to use information and knowledge management for competitive advantage.
- There is also a thriving, exciting bleeding edge of start-ups in advanced data management.
- If you plan to go on to research, the course unit places you at the cutting-edge of data management research.

Syllabus ::
How will the topics be covered?

Week 1 Architecture/Components, Relational Databases, Query Processing

Week 2 Query Processing (cont.), Parallel QP

Week 3 Distributed DBMSs

Week 4 Massively Parallel/Distributed QP, Peer-to-Peer QP

Week 5 Stream and Sensor Network QP

NOT on the Syllabus ::
Other Kinds of Advanced DBMSs

- Data Warehouses and Data Mining
- Unstructured and Semi-Structured QP
- Spatio-Temporal DBMSs
- Scientific DBMS

1.2 What style of course unit is this?

Assumptions ::
What do I need to know already?

- The course unit focusses on DBMSs as **software systems**.
  - It is not about applications; it is about **software-system infrastructures**.
  - It is not about DBMSs as shrink-wrapped, off-the-shelf tools; it is about **peeking inside DBMSs, as glass boxes**
  - It does not focus on transaction and recovery, or on storage management, etc.; it does focus on **query processing**
- It assumes a good understanding of undergraduate-level material on:
  - database languages, database design and database application programming
  - core computer science (e.g., algorithms and data structures, etc.)
- The emphasis is on **systems**, not on theory (but a sprinkling of theory is needed)

Teaching v. Learning ::
What will I be told and what will I find for myself?

- The lectures, as well as the coursework, will centre around, and be driven by, research papers that will be assigned for advance reading.
- Conventional lab hours, with lab sheets for practicing skills, are not planned.
- **Not** centred around learning how to use tools.
- The material is not given a theoretical or a formal treatment.
- The course unit is **practical** (i.e., you will learn things that matter in industry and in research) but it is **not practice-based** (i.e., the kind of knowledge imparted in this course unit is not often embodied in usable tools that you could hope to get to grips with during your time here).

Assessment ::
How will I know how well I have learnt?

- Coursework mark has two components:
  - short tests (quiz-like) from the second teaching-day onwards
    - based on week-to-week revision, and assigned reading of research papers.
    - each test is set as if it were an exam question (in terms of structure, size, time to answer, difficulty, etc.)
    - required reading is tested
    - tests help students revise/assimilate material as we go, and are drills for the exam
  - a final report centred on reading, dissecting and comparing two related research papers on advanced DBMSs

1.3 What are the hard facts?

Advanced Database Management Systems (1) ::
The Basics

Structure	Level	Credits	Lectures	Coursework
	6	15	24 h	(5 * 12 h) + 20 h

Where and When	Period	Location	Day	Time
	1	2.19	Friday	9:00-17:00

Contribution (1)	Coursework	Exam
	50%	50%

Contribution (2)	Short Tests	Final Report
	(15)%	35%

Advanced Database Management Systems (2) ::
The history so far

- This course unit was first taught in the 2007/2008 academic year.
- The mean marks over 34 students for coursework, exam and final (out of 100) were, resp., 60.5, 63.9, and 62.2, with 18.3, 14, and 14.8 standard deviation.
- It is not one of the easiest course units (e.g., every year there are 5-10% exam resits).
- In 2008-2009, there were detected cases of academic malpractice, so make sure you know what you must and mustn't do. (See, among many other pieces of guidance, `http://www.campus.manchester.ac.uk/medialibrary/tlao/plagiarism-guidance-for-students.pdf`.)

Bibliography ::
What should I read?

- There is no adopted textbook for the course: you will get copies of the lecture slides for you to make notes on and you will need to read the assigned research papers (one or two every week).
- To brush up on the assumed background and to follow the more introductory material in the front-end of the course unit, any one of these books serve well the purpose (the second one below is probably the most suitable, by a whisker):

  [Garcia-Molina et al., 2002]

  [Ramakrishnan and Gehrke, 2003]

  [Silberschatz et al., 2005]

- Keep an eye on the course unit web page: `http://www.cs.man.ac.uk/~alvaro/teaching/COMP6036.html`

References

[Garcia-Molina et al., 2002] Garcia-Molina, H., Ullman, J. D., and Widom, J. (2002). Database Systems: The Complete Book. Pearson Education Limited, 1st edition.

[Ramakrishnan and Gehrke, 2003] Ramakrishnan, R. and Gehrke, J. (2003). Database Management Systems. McGraw-Hill Education - Europe, 3rd edition.

[Silberschatz et al., 2005] Silberschatz, A., Korth, H. F., and Sudarshan, S. (2005). Database System Concepts. McGraw-Hill Education - Europe, 5th edition.