Database Architecture

Alberto Abelló Gamazo
Roberto García
Rosa M. Gil Iranzo
Oscar Romero
Marta Oliva

PID_00179806

UOC teaching material

Universitat Oberta de Catalunya

Alberto Abelló Gamazo

MSc and PhD in Computer Science from the Universitat Politècnica de Catalunya. He is associate professor of the IT Faculty of Barcelona. He is also a member of the Information Processing and Modelling Research Group (*Grup de recerca en Modelització i Processament de la Informació*) at the above university, specialising in software engineering, databases and information systems. His research interests include database design, data warehousing, OLAP tools, ontologies and reasoning. He is author of articles and papers given at national and international conferences and published in journals on these subjects.

Roberto García

PhD in Computer Science from the Universitat Pompeu Fabra, Master's in Electronic Commerce from La Salle Barcelona and MSc in Computer Engineering from the Universitat Politècnica de Catalunya. Currently, Roberto is associate professor at the Universitat de Lleida where he teaches courses and conducts research into Human-Computer Interaction, Web engineering, Semantic Web and software development.

Rosa M. Gil Iranzo

PhD in Computer Science by the Universitat Pompeu Fabra in 2005, she is currently associate professor of the IT and Industrial Engineering Department of the Universitat de Lleida. She also teaches as part of the Master's in Information Technology of the Universidad de Zaragoza. Her research interests lie in Human-Computer Interaction and open data.

Oscar Romero

With an MSc and PhD in Computing from the Universitat Politècnica de Catalunya, he is currently a tenure-track 1 lecturer at the Barcelona School of Information Technology. He is also a member of the Information Processing and Modelling Research Group at this university, specialising in software engineering, databases and information systems. His research interests mainly concern business intelligence, namely data warehousing, OLAP tools, multidimensional modelling, service-oriented BI, the semantic web, description logics and ontologies, NOSQL and query recommendation, among others. He has authored articles and papers for national and international conferences and in journals on these subjects.

Marta Oliva

PhD in Computer Science from the Universitat Politècnica de Catalunya. She is associate professor of the IT and Industrial Engineering Department of the Universitat de Lleida, assigned to the Escola Politècnica Superior. She has taught courses on databases as part of computer science degrees since 1992.

The assignment and creation of this teaching material have been coordinated by the lecturers: Alberto Abelló Gamazo, M. Elena Rodríguez González (2012)

First edition: September 2012
© Alberto Abelló Gamazo, Roberto García, Rosa M. Gil Iranzo, Oscar Romero, Marta Oliva
All rights reserved
© of this edition, FUOC, 2012
Av. Tibidabo, 39-43, 08035 Barcelona
Design: Manel Andreu
Publishing: Eureca Media, SL
Legal Deposit: B-20.494-2012

The texts and images contained in this publication are subject -except where indicated to the contrary- to an Attribution-NonCommercial-NoDerivs license (BY-NC-ND) v.3.0 Spain by Creative Commons. You may copy, publically distribute and transfer them as long as the author and source are credited (FUOC. Fundación para la Universitat Oberta de Catalunya (Open University of Catalonia Foundation)), neither the work itself nor derived works may be used for commercial gain. The full terms of the license can be viewed at http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode

Introduction

The goal of this subject is to first introduce a centralized functional architecture of a DBMS, to evolve it into a distributed one, outlining their differences.

First, we will present the limitations of the relational model to introduce objects and semi-structured data, giving rise to object-relational and XML extensions. The former will focus on the object features in the standard (i.e., large objects, user defined types, inheritance, references and multi-valued attributes). The latter will focus on XML standards like XSchema, XPath, and XQuery.

A classification of distributed databases will be explained. Then, depending on whether distribution is desired or imposed, we will talk, respectively, about design or integration. Design is related to fragmentation (i.e., horizontal, vertical or hybrid), while integration can be implemented as global as view, local as view or peer-to-peer, based on wrapper-mediator architecture.

Database security encompasses three aspects: confidentiality, integrity and availability. So, we will review topics related to access control, application access, vulnerability, inference and auditing mechanisms. We will focus on access control systems and describe the discretionary and mandatory access control models, as well as the role-based access control (RBAC) model. We will also discuss security for advanced data management systems, and cover topics such as access control for XML.

We will also recall the concept of transaction. Based on this, we will distinguish ACID and BASE transactions, to focus on the former. We will discuss multilevel locking, multi-version locking and timestamping. The first two have been chosen because of their wide use in centralized commercial systems, while the latter has been chosen because of its wide use in distributed environments. Also related to transactions, we will discuss the concepts related to recovery.

Finally, we will extend the concepts related to query optimization to be able to deal with distribution and parallelism. We will first analyze syntactic optimization to later deal with the physical phase. Intra- as well as inter-operator parallelism will be analyzed.

Objectives

The teaching materials associated to this subject aim at facilitating students to reach the following objectives:

1. Know two extensions to the classical relational model, which: a) capture more semantics (i.e., object-oriented extension) and benefit from them, and b) represent semi-structured data (i.e., XML), validate and retrieve them.

2. Understand the benefits and means to partition and distribute data.

3. Know the different ways to integrate schemas and data.

4. Understand the responsibilities of a database administrator from the security point of view.

5. Understand confidentiality threats and solutions available in relational DBMSs.

6. Learn new concurrency control techniques beyond simple locking (i.e., multi-granule locking, multi-version locking and timestamping).

7. Understand the benefits and means to work with in-memory data.

8. Know how a data manager works and guarantees recovery (i.e., logging and backup).

9. Understand new transaction models and mechanisms appropriate for highly distributed systems.

10. Understand the benefits of parallel execution of queries.

11. Know the modifications that must be introduced in a query manager to deal with the distribution of data and execution.

Contents

Module 1
Database Management Systems
Alberto Abelló Gamazo
1. Relational Extensions
2. Functional architecture of a DBMS

Module 2
Relational Extensions
Oscar Romero, Roberto García and Rosa M. Gil Iranzo
1. The Object-Relational Extension
2. The XML Extension

Module 3
Distributed Databases
Òscar Romero and Marta Oliva
1. Distributed Processing
2. Brief History and (Tentative) Classification of Distributed DBMSs
3. Distributed DBMS Architectures
4. Distributed Database Design
5. Database Integration

Module 4
Security
Marta Oliva
1. Database Security Concepts
2. Security Models
3. Statistical Databases
4. Security for Advanced Data Management Systems
5. Data Protection Act

Module 5
Transaction Models and Concurrency Control
Alberto Abelló
1. Interferences and Isolation Level
2. Centralized Concurrency Control
3. Transactions in Oracle
4. Distributed Concurrency Control
5. Replica Management

Module 6
Data Management
Marta Oliva Solé
1. Loss of Data Due to Errors or Failures

2. Data Manager Architecture
3. Logging Recovery
4. Media Recovery
5. Further Aspects in Distributed Databases

Module 7

Distributed Query Optimization

Oscar Romero and Alberto Abelló Gamazo

1. Basics of Query Optimization
2. Distributed Query Optimization

Bibliography

Abiteboul, S.; Manolescu, I.; Rigaux, P.; Rousset, M.-C.; Senellart, P. (2011). *Web Data Management*. New York: Cambridge University Press.

Castano, S.; Fugini, M.; Martella, G.; Samarati, P. (1995). *Database Security*. Addison-Wesley.

Garcia-Molina, H.; Ullman, J. D.; Widom, J. (2009). *Database systems: The complete book, second edition*. Pearson Prentice Hall.

Liu, L.; Tamer Özsu, M. (Eds.) (2009). *Encyclopedia of Database Systems*. Springer.

Tamer Özsu, M.; Valduriez, P. (2011). *Principles of Distributed Database Systems* (3rd Ed.). Prentice Hall.