Database Professional (DBTech Pro) Project

WP3 & WP4 & WP9 – Workshops Connected Work Packages

WP3 Specification of Teaching Practices & Laboratories

Jaakko Rantanen, Work Package Leader

Häme Polytechnic

Jaakonkatu 28

13100 Hämeenlinna

Finland

email: jaakko.rantanen@hamk.fi

WP5 Pilot Runs

Fritz Laux, Work Package Leader

Dept. Business & Information Sc.

Reutlingen University

Alterburgstr. 150

72762 Reutlingen

Germany

email: fritz.laux@fh-reutlingen.de

WP9 Workshops Planning and Organisation

Kari Silpiö, Work Package Leader
Helsinki Business Polytechnic, Helia
Ratapihantie 13
00520 Helsinki, Finland
email: kari.silpio@helia.fi

Table of Contents

Preface

This publication is part of the final report of the DBTech Pro project. The project tasks have been divided to nine work packages. Three closely interconnected work packages are reported here inside common covers.

The Database Technology Professional (DBTech Pro) project is an EU, Leonardo da Vinci Program Pilot project involving the participation of academic and industrial partners. The project is shortly introduced in the Introduction part of this book.

The DBTech Pro project defined a certain type of workshop as a collaborative working method. The "DBTech Pro Workshop" turned out to be a useful and productive working method. The project group believes that the method could be applied more widely. The authors of this publication try to introduce the method for a more broad audience than database specialists.

Part 1, the Final Report of the Work Package 9: " Workshops Planning and Organisation ", introduces the "DBTech Pro Workshop" method and makes a summary of the experiences using the method. The Part 2 may gain the most widest interest.

Part 2, the Final Report of the Work Package 3: " Specification of Teaching Practices & Laboratories ", introduces the four workshops that was ran during the DBTech Pro project. This part contains more substance related details than the previous part. However, there can be found examples how the workshops are described, evaluated etc. This report contains information that may be interesting to database educators and specialists.

Part 3, the Final Report of the Work Package 5: "Pilot Runs", gives a short description of experiences on piloting the workshop modules outside the original institution.

Introduction

The database is now the underlying framework of the information system and has fundamentally changed the way many organisations and individuals work. This is reflected within tertiary education where databases form a core area of study in undergraduate and postgraduate programmes related to computer science and information systems, and typically at least an elective on other data-intensive programmes (ACM/IEEE 2001; EUCIP 2003). The core studies are commonly based on the relational data model, SQL (the de facto language for relational DBMSs), data modelling and relational database design. This curriculum initially supported industry needs where the relational DBMS was the dominant data-processing software. At the same time, the field of computing has seen significant developments over the last decade with object-orientation becoming much more predominant and the Web emerging as a global communication and business medium. Database technology has encompassed these new developments with the emergence of object-relational DBMSs, object-oriented DBMSs, support for Web-database integration, HTML, XML, XQuery, Java, J2EE, data warehousing, business intelligence, and so on. These developments have significantly increased the breadth and depth of knowledge that a modern database professional requires to work effectively in industry.

The Database Technology Professional (DBTech Pro) project is an EU, Leonardo da Vinci Program Pilot project involving the participation of academic and industrial partners from the Database Technology Network (DBTechNet) initiative [1] that is investigating how academia can better support the needs of the database community. The partnership extends across five EU member states (Finland, Germany, Greece, Spain, UK) and involves seven academic institutions: Hochschule Reutlingen, Germany; Häme Polytechnic, Hämeenlinna, Finland; Helia (Helsinki Business Polytechnic), Helsinki, Finland; TEI of Thessaloniki, Greece; University of Macedonia, Thessaloniki, Greece; University of Malaga, Spain; University of Paisley, UK; and three companies: ALTEC S.A., Thessaloniki, Greece; Solid EMEA North, Finland; TietoEnator Public Sector, Espoo, Finland.

The project aims include:
      (a) the identification of professional roles and skills in relation to contemporary database technology practice in Europe;
      (b) for each one professional role identified, the specification of the knowledge and skills that are needed by the individual who wishes to achieve competency in the relevant job market;

(c) the development of pilot course and laboratory workshop material for training and, subsequently, testing the knowledge and skills in question;

(d) an investigation into the possibility of developing a vendor-independent European certificate of Database Technology Professional.

In this respect, the goals of the DBTech Pro project are along the lines of, and complementary in nature to, those of the broader industry-driven, vendor-independent professional certification and competency development initiatives (e.g. EUCIP) that target the database professional and practitioner. Direct beneficiaries of the project include:

- Students studying their degree in the partner institutions will obtain basic knowledge and skills as a database professional for employment opportunities both with national and international enterprises.

- Enterprises will be assisted to catch-up with the rapid pace of developments in database technology, assess and adapt to the existing databases. In this respect, they will avoid pitfalls and they can employ highly trained and competitive staff. The proposed European DBTech Pro Certificate for a database expert will guarantee the quality of knowledge and the competence in modern database technology of the employees. In a larger scale this will benefit the competitiveness of the whole European industry on worldwide markets, since professional use of databases is the foundation of all information systems.

- Database professionals in SMEs having the proposed European DBTech Pro Certificate can prove their knowledge and skills – and their value in the labour market.

- Educational institutions will combine their human and technical resources and thus get better results for all their staff members and students.

To enable the efficient and effective organisation of the collaboration across the partner institutions, the project has been decomposed into a number of work packages:

WP1: Project Management

– to oversee the effective and efficient organisation of the project.

WP2: Specification of Knowledge Areas

– to determine the knowledge areas required by a European Database Technology professional.

WP3: Specification of Teaching Practices and Laboratories, Pilot Tests

– to determine what teaching modules and laboratories exist across the partners, and to identify and develop four pilot workshops during the project.

WP4: Content Planning

– for each the knowledge areas identified in WP2, compile an initial version of educational/course content outline, indicative of the database technology knowledge and skills to be assessed for professional competence in the job market.

WP5: Pilot Run of the DBTech Pro

– to establish pilot runs of the four workshops that will be created during this project among the partner organisations.

WP6: Assembly of the DBTech Pro Certificate

– to make recommendations on the industry support for a vendor-independent certification scheme for a European Database Technology professional.

WP7: Evaluation and Assessment

– to evaluate the project overall and the four pilot workshops to be created during the project.

WP8: Dissemination

– on an on-going basis, to disseminate the results of the project.

WP9: Workshops Planning and Organisation

– to plan and organise the four pilot workshops to be developed during this project.

Part I - Final Report of the WP9; Workshops Planning and Organisation

I.1 Introduction to WP9

The aim of this Part I is to discuss the DBTech Pro workshop concept and format. The contents of the workshops, workshop-related teaching practices and labs are discussed in the WP3 report in Part 2. Workshop evaluation is discussed in the WP7 report.

I.1.1 Background, Aims, and Objectives

In February 2004 the project group agreed that it would be beneficial to coordinate DBTech Pro workshops planning and organisation without being tied up with any of the existing work packages. The new work package (WP9: Workshops Planning and Organization) was started in March 2004. Kari Silpiö (Helsinki Business Polytechnic, Finland) was named as the WP leader. The aims and objectives for WP9 were specified in detail as follows:

1. To plan and develop workshops in a productive and cooperative way.
2. To have successful DBTech Pro workshops with a new high-quality theoretical content and laboratories, reusable learning modules and innovative teaching practices.
3. To have a uniform workshop documentation.
4. To produce workshop modules ready to pilot right after the workshop.
5. To be able to utilize the workshop modules in all the DBTech Pro member institutes
6. To learn from the experiences in planning and organising the workshops.
7. To share pedagogical ideas and learn from the different teaching cultures and approaches.

WP9 was the provider of ready-to-pilot modules for WP5 (Pilots). Furthermore, WP9 cooperated with WP7 (Evaluation) in analysing the workshops from the workshop instructor's viewpoint and developing the workshop concept and method.

I.1.2 The Four Workshops

A total of four pedagogically monitored DBTech Pro workshops were developed and organised in four countries. The themes and modules for the workshops are shown in Table 1.

Table 1. DBTech Pro Workshops 2003-2005.		
Workshop	**Theme**	**Modules**
November 2003, Thessaloniki	Database Administration & Database Access	1. Backup and Recovery 2. Concurrency Control
May 2004, Malaga	Database Design	1. Requirements Elicitation and Database Design Methodology 2. Systematic Index Design
November 2004, Reutlingen	Data Warehousing & Data Mining	1. Data Warehousing 2. Data Mining
April 2005, Helsinki	XML and Databases	1. SQL Meets XML 2. SQL/XML and XQuery

Table 1. DBTech Pro Workshops

The project group named themes for the workshops based on the DBTech Pro content specified by WP4 (Content Planning). The main criteria for selecting the themes were the importance of the issue and the lack of teaching modules with laboratories and real-life cases on the specific issue. The themes were designated to the workshops and host institutes based on an estimation of which institute would get most benefit from the workshop organised by external experts.

In the workshops there have been from 30 to 50 participants attending to the theoretical part. The labs have taken place in one or two computer classrooms and there have been from 16 to 22 participants in labs (students, people from the industry and local teachers). In

addition to the actual lab participants there have been from 10 to 14 representatives of the project members present in the workshops observing and evaluating the workshop. Furthermore, external evaluators from academic institutes monitored the third and fourth workshop. The information on the workshop lab participants is shown in Table 2.

Table 2. Information on the Workshop Lab Participants.				
WORKSHOP	**Students**	**Participants from the industry**	**Local teachers**	**Degree Programmes & Related Database Courses**
WS 1, T.E.I of Thessaloniki	8 T.E.I 12 University of Macedonia	1	0	T.E.I: 7th semester (BSc), *Special Topics in Database Systems* University of Macedonia: 6th semester (BSc), *Database Systems II*
WS 2, University of Malaga	16	0	0	*Database Design and Use*, 5th semester (MSc)
WS 3, University of Reutlingen	12 University of Reutlingen 4 Other	0	2	Business Information Management, 1st semester (MSc), *Analytical Information Systems, Executive Information Systems*
WS 4, Helsinki Business Polytechnic	14 Helsinki Business Polytechnic 1 Häme Polytechnic 1 Other	4	2	Business Information Technology, 6th -7th semester, *Data Access Technologies*

Table 2. Workshop Lab Participants

I.2 The Workshop Concept

I.2.1 The Workshop Model

The model for developing DBTech Pro reusable learning modules is based on the ideas of simulating authentic real-life environments and problems in the labs, and the learner as an active subject in the learning process who gains knowledge through interaction. Most topics are dealt with in terms of the TEE approach: theory (T), evidence (E), and evaluation (E). That is, start with theory (T), move to evidence (E), and then conclude with evaluation (E). The TEE-approach based workshop design scheme shown in Figure 1 was developed when planning for the first DBTech Pro workshop.

(DBTech Pro project) Workshop Design Scheme

Theory
- Orientation to the theory
  - ✓ Primary concepts
  - ✓ Identifying problems

- Orientation to practicals/vendor approaches
  - ✓ Selected approaches
  - ✓ Other approaches

Hands-on Labs
- Recognising the concepts
  - ✓ Experiencing the problems

- Observing the behaviour
  - ✓ Understanding

Demonstrations
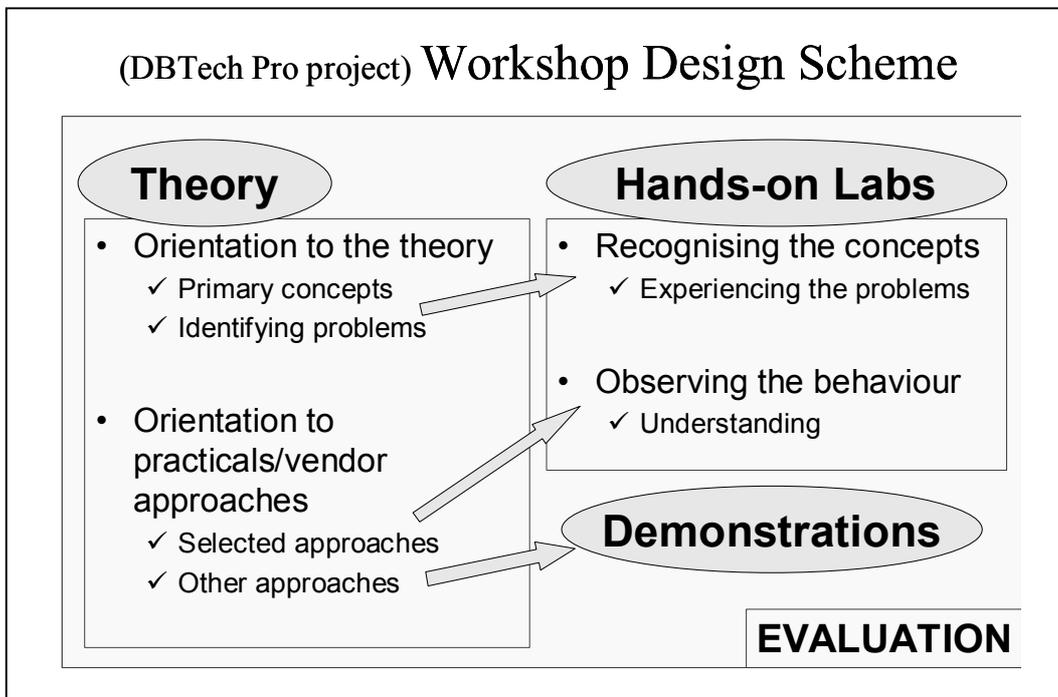
EVALUATION

Figure 1. The Workshop Design Scheme.

A two-day DBTech Pro workshop contains two tightly connected modules under a specific theme. The module content is derived from the substance focused by the DBTech Pro project. Due to the evaluation needs the primary participants of a workshop should be a well-defined group of students who participate to the whole workshop. The other

participants are people from the industry, staff members of the host institute and representatives of the educational and industrial DBTech Pro project partners.

From the teaching point of view a workshop contains the following elements: pre-study before the actual workshop, theoretical introduction and hands-on laboratories in the workshop, and the student reports after the workshop. The theory is connected to the practical part and the focus is on cases taken from the real working life. The workshop guide, study material, and references are put on a web site / virtual learning environment some weeks before the workshop. With the guidance of their local teacher the participants/students familiarize themselves with the incoming issue before the workshop. The outline of a DBTech Pro workshop organisation is shown in Figure 2.

1. Pre study
- The module descriptions, materials, and guidelines for the pre study are available on a web site / virtual learning environment before the workshop

2. Workshop (2 consecutive days, two modules under the same theme)
- Theory part / presentation
- Laboratory work
- Evaluation

3. Post study
- The students solve a problem related to the module content
- The students and teachers use a virtual learning environment for communication
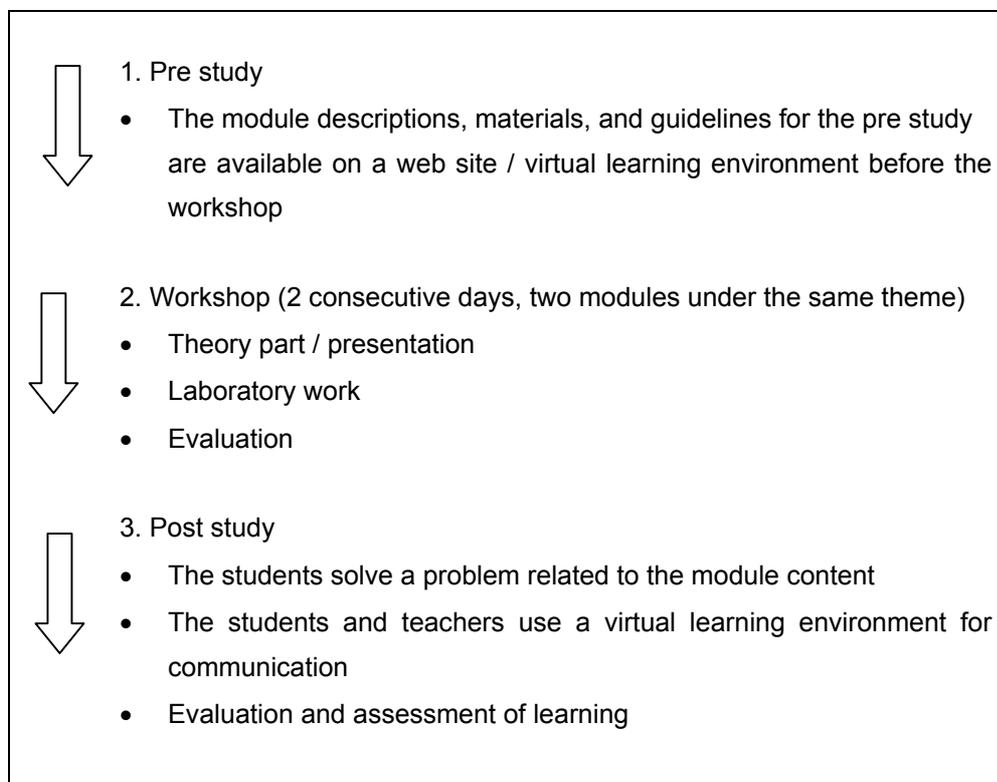- Evaluation and assessment of learning

Figure 2. Outline of a DBTech Pro Workshop.

I.2.2 Workshop Documentation

DBTech Pro workshops are documented in a systematic way. The main documents are the following: The Workshop Plan, Workshop Instructions & Study Material, Laboratory Environment & Test Data Documentation, and The Final Report for the Workshop.

The workshop developers plan and write the specification needed for laboratory with guidelines for a teacher. In the first run of the workshop the workshop developers set up the laboratory in the host institution where the workshop will be run. While setting up the lab environment they also test the installation instructions. The goal is that after the first run each of the partner institutions would be able to repeat the workshop autonomously.

The standard documentation of a workshop module consists of the following:

- Module description (motivation, perquisites, learning objectives, module content etc.)
- Teaching and study materials (presentation materials, task papers, instructions etc.)
- Instructions for the teacher
- Laboratory environment documentation and installation instructions
- A final report for a DBTech Pro Workshop.

The content of a final report on a DBTech Pro workshop is specified as follows:

- Notes on issues related to the technical environment
- Notes on practical workshop organisation issues
- Summary of the feedback from the participants
- Evaluation results for the workshop content
- Evaluation results for the teaching-learning process.

I.2.3 Workshop Planning and Organization Process

The project organised four workshops in T.E.I of Thessaloniki, University of Malaga, University of Reutlingen, and Helsinki Business Polytechnic. The WP4 (Content Planning) final report discusses the process of specifying the workshop themes and module topics [3].

After the workshop themes were specified the project group chose a pair of developers/instructors for each workshop. To make it possible to test the reusability of learning modules and the portability of lab environments the workshop

developers/instructors were not from the institution that hosted the first run of the workshop. This also supported sharing of special expertise. The developers planned the workshops and modules in tight co-operation with the host institution of the upcoming workshop in order to get the workshop to form a part of or an extension to an existing database course. During the planning process both the educational and industrial partners evaluated the workshop plan. The Internet was utilized in the communication. The following tools were used in the communication and sharing documents: e-mail, synchronous discussion board, synchronous voice-over-IP conferences, and the intranet.

The developers produced workshop materials including lab specifications with the teacher's guidelines. They also put up lab environments in the host institutions for the first runs of the workshops. At the same time they tested and evaluated the teacher's guidelines. The goal was that after the first run of a workshop any of the partner institutions would be able to repeat the workshop modules autonomously.

I.2.4 Workshop Evaluation

Evaluation is an important part of the DBTech Pro workshop method. During the project run the internal evaluation group carried on process evaluation, pedagogical evaluation, and content evaluation. The representatives of the industry evaluated the module contents and organisation practices from the viewpoint of the real-life challenges in the database field and gave feedback both during the module development and organisation phases. Students evaluated the modules in the workshops from the student's viewpoint. Furthermore, academic project members evaluated modules from the viewpoint of reusability and teaching and workshop developers evaluated the process and achieved results from the developer's viewpoint, focusing especially on lab reusability and portability issues. Selected workshop modules were also piloted in some partner institutions. All the pilots were documented and evaluated in a formal way. The results of the evaluation were also utilized in the iterative development of the workshop method itself.

The WP7 (evaluation) group collected all the evaluation data from the workshops and processed it in a systematic way. From the industrial point of view a DBTech Pro workshop can also be seen as dissemination and benchmarking activity, testing how well the workshop content relates to the challenging real-life problems in the database field.

Generally, the results of the evaluation were good [4]. The WP7 final report discusses the evaluation process in detail and summarises the results of the evaluation.

I.3 Experiences and Recommendations

The project group's own experiences of the workshop method have been very positive. The method proved to support collaborative work and cooperation between the educational and industrial partners. The workshop framework seems to have been a very motivating factor in the project. The project even got a generous amount of voluntary work contribution from external database experts from the industry who were really interested in participating in developing and testing the index design lab for the second workshop [2]. Thanks to the voluntary work it was possible to achieve significant results and such information that was valuable not only for the project use, but for the local industry.

Using the workshop method it was quite easy to set clear objectives for the development, testing and evaluation of reusable learning modules, and distribute work in a productive way during the project. The cooperation of teachers from academic institutes and database experts from the industry made it possible to combine challenging real-life cases and sound theoretical information in a practical and innovative way. Furthermore, students got valuable experiences when they attended the workshops that had both international teachers from foreign institutes and representatives from both the DBTech Pro industrial partner organisations and local industry.

Sharing of the workshop modules was quite easy within the project because all the academic project members were observing the first runs of all the workshop modules before they piloted those modules autonomously. On the other hand this has been a limitation on how the workshop method was used in the project. The easiness of reusing a workshop module was not tested in such case where the teacher had not observed the original workshop organisation. Another limitation of the method is that it does not cover the maintenance and distribution of updated learning modules. Especially in the information technology field the products and standards evolve so fast that a lab can be out of date in less than one or two years. We recommend that in the future the method should be developed further to cover also these issues.

The DBTech Pro workshop model emphasises the importance of pre study before a workshop for making sure that the initial level of workshop participant's knowledge is

suitable for the workshop. Nevertheless, on some occasions pre study instructions and materials were not available to the students before a workshop due to various reasons. In the DBTech Pro workshop framework the student is seen as an active subject in the learning process who gains knowledge through interaction. Generally, the workshop timetables have been quite tight. Due to some unexpected delays and technical problems it has been impossible to run most of the workshop modules on schedule. Furthermore, in most of the workshops there has not been time reserved for discussions during the labs. Although a virtual learning environment has been available for discussions after the workshops, more time should have been reserved for interaction during the workshop. We recommend that these issues should be taken into account when planning and organising workshops in the future.

Generally, setting up and testing the technical lab environment is quite time-consuming and error-prone work. For the workshop organised in Malaga the lab environment was set up on Häme Polytechnic's database server located in Finland. The students successfully used the lab environment over the Internet via a virtual private network. This kind of arrangement made it easier to set up and test the lab environment before the workshop. On the other hand, in this arrangement technical support had to be available also in Häme Polytechnic during the lab sessions. Technically this service provider model for a lab environment proved to be successful and beneficial for a single workshop run. However, it should be noted that as a continuous arrangement this model requires continuous investments on the technical environment, technical support, and lab administration in the provider organisation.

I.4 Summary and Conclusions

The project developed the DBTech Pro workshop method for producing reusable learning modules in a productive and co-operative way. The method consists of the workshop planning process, productization and documentation of reusable learning modules (i.e., workshop modules), general outline of a workshop organisation, evaluation, and piloting of modules in partner institutes and companies. The method is independent from the subject and can also be used in various other contexts.

The project produced eight reusable learning modules using the workshop method. The modules were tested in four workshops in different countries. Some modules were piloted in the member institutes [3]. Four modules were piloted in a partner company (TietoEnator, Finland). The students, industry representatives, external academic evaluators, all the project group members, and the internal evaluation group evaluated the modules. The results of the evaluation proved that the workshops and pilots were successful and the general quality of the reusable modules is good [4]. The DBTech Pro workshop framework proved to be useful in developing, testing and sharing reusable learning modules and teaching practices. In addition to sharing reusable learning modules the workshops and workshop pilots were found especially invaluable in testing and evaluating portability and reusability of labs and lab documentation.

In addition to the viewpoint of the workshop method as a tool used in developing learning modules DBTech Pro workshops can also be seen as collegial teaching studios providing an efficient way to disseminate teaching practices and new ideas. Our conclusion is that the DBTech Pro workshop method supported sharing of expertise, shared knowledge advancement, and collaborative building of knowledge in the project.

References

1. DBTech Pro. 2005. WP3 Final Report (Specification of Teaching Practices and Laboratories).
2. DBTech Pro. 2005. WP4 Final Report (Content Planning).
3. DBTech Pro. 2005. WP5 Final Report (Pilots).
4. DBTech Pro. 2005. WP7 Final Report (Evaluation).

Part II - Final Report of the WP3; Specification of Teaching Practices & Laboratories

II.1 Introduction to WP3

The task of Work Package 3 is expressed as "Specification of teaching practices and laboratories including pilot tests".

Start date	1 January 2003
End date	31 May 2004
WP leader	Jaakko Rantanen, Häme Polytechnic
Outcomes	Implementation outline and test reports

II.1.1 General Description

The main tasks was the analysis of different teaching practices, testing in different modules, specification and construction of laboratories, pilot tests for using those, co-operation with the pedagogic evaluator.

The definition and development of cases, exercises and laboratories took place in two phases. The first phase preceded the first workshop, the second phase followed thereafter.

II.1.2 Cases and Exercises

Requirement definitions were produced as a starting point. The initial definitions were too broad, so that only a part of it was put into practice during the DBTech Pro project. The requirement definitions were revised after the first workshop.

All educational partners were asked to present any existing cases & exercises they have used or are using. One important source is the Connolly & al. textbook [2] and its additional material.

II.1.3 Lab specs

A Workshop Plan describes what are the aims and learning goals, what is the prerequisite knowledge, what is the content, how the workshop will be implemented and what teaching practices will be used.

Workshop Instructions & Study Material includes notes and optional instructions for the instructors as well as instructions for the students. Which are the course texts, bibliography and links is described, what optional pre-study material is available. Texts for the laboratory exercises, cases etc. is the backbone of Workshop Instructions & Study Material.

Laboratory Environment & Test Data Documentation consists of the hardware and software environment needed in the workshop and its configuration and data. Instructions how to reinstall the workshop is included as well as instructions and tools for creating the test data needed in the workshop.

II.1.4 Pedagogical development

Planning the testing of cases, exercises and labs usually required a substantial amount of work. In several cases the exercises and especially labs could locally pre-tested before each workshop. A quick analysis was made so that any necessary improvements can be done before the workshops. Feedback from students & teachers will be collected in the workshops and it will be analyzed.

II.1.5 Workshops

The first workshop took place in November 2003 in Thessaloniki and the second in May 2004 in Málaga. The third workshop was held in November 2004 in Reutlingen and the fourth workshop in April 2005 in Helsinki.

II.2 Partners' Existing Lab and Study Work Material

II.2.1 Laboratory and Exercise Types and Topics

This chapter introduces a generic list, a simple categorization and description for the possible aims and topics of the laboratory exercises. The purpose is that the list would cover the most essential topics. The categories follow the DBTech Pro knowledge areas. Some of these demos/labs may exist in some form, maybe instructed in a native language and being perhaps available only in a specific (lab) environment. Each topic in the list is marked with a short label so that any referencing could be more convenient.

The next chapter presents a list of existing labs of educational partners, i.e. which of the following laboratory and exercise types and topics are implemented and which are not.

DB Administration Labs (AL)

- **AL-Concept:** learning to understand the key concepts: database, transaction log, checkpoint, point of consistency, roll-forward recovery, commit & rollback, backup & restore. The purpose is to simulate different types of failures and recovery.
- **AL-Config:** install (maybe not in practise) but at least to discuss the configuration – what is essential to achieve a feasible and fault tolerant db installation.
- **AL-DbObj:** administrative implication of "physical" database objects (the homonym "database", tablespaces, files/file groups, etc.) – what are the units of backup and reorganization etc.
- **AL-Security:** security: the concepts of authentication vs. authorization, planning and maintaining privileges, user roles and groups.
- **AL-Catalogue:** utilizing the system catalogue information – analyzing dependencies between database objects.
- **AL-Diagnostic:** collecting diagnostic information – data growth, data fragmentation, memory pool consumption, etc.
- **AL-Tuning:** performance tuning – the platform configuration, analyzing query/access plans, the effects of indexing and ordering of rows.

DB Design Labs (DL)

- **DL-Create:** creating "logical" database objects (tables, views, triggers and procedures) and constraints – loading data - testing the behaviour.

- **DL-Example:** observing a set of examples: the behaviour of proper designs and misbehaviour of bad designs (e.g. update problems of unnormalized data and inconsistency problems).

- **DL-Tools:** using design tools
    - database modelling & generation & reverse engineering tools, e.g. ERStudio or ERWin.
    - UML modelling tools and SQL DDL generation – critical observation of the possibly unsophisticated results, e.g. Rational Rose, TogetherJ, MS Visio, Platinum Paradigm, etc.

- **DL-Doc:** documenting the solution by utilizing the system catalogue information.

- **DL-Tuning:** affecting the performance – testing index access methods, effects of logical design alternatives.

- **DL-DbObj:** administrative implication of "physical" database objects ("database", tablespaces, files/file groups, etc.).

- **DL-Admin:** writing instructions for administrators: diagnostics, backup, reorganization, security.

DB Programming Labs (PL)

- **PL-ISQLq:** interactive SQL
    - a website for running queries against several dbms's to observe both syntactic and semantic differences between products.
    - local dbms installations for running updates and DDLs.

- **PL-ISQLu:** understanding the (atomic) behaviour of an updating SQL statement.

- **PL-Tx:** understanding the transaction integrity.

- **PL-Isolation:** understanding the transaction isolation levels.

- **PL-APIs:** demonstrating programming using different types of APIs, the basic program structure in each, mapping SQL data types to host languages.

- **PL-Exception:** handling exceptions, SQLCODE/SQLSTATE/native, the different techniques in the environments should not hide the clue (e.g. the "On error" or "Whenever error" approach / "try…catch" approach / ADA style exception part in a block / etc.).

- **PL-Sp:** writing and testing stored procedures, the fact that the sp-languages are native (the PSM standard is not de facto).

- **PL-SQLJ:** SQL embedded in Java, Java packages as database objects.

Data Warehouse Development Labs (WL)

- **WL-OLAP:** demonstrating the key concepts of OLAP as an user interface to DW by using demo applications:
    - ROLAP and the star schema: fact, dimension, measure, drilling
    - MOLAP and the hypercube: fact, dimension, measure, drilling
- **WL-Schema:** starting from two existing simple operational (OLTP) databases and a set of given requirements for a simple data warehouse (or a data mart) solution, developing a simple star schema and transferring and converting data from the operational databases to the star schema db.

Data Mining Labs (ML)

- **ML-Principles:** Knowledge discovery in large relational data repositories: conventional query tools (and statistics) do not suffice. Data vs. information. The process of performing data mining activities, from defining a business problem through providing recommendations for its solution.
- **ML-Stages:** Data organization, selection, cleaning and enrichment.
- **ML-Market Basket:** Frequent itemsets and the market basket analysis.
- **ML-Case Study:** A real life application is considered in the form of a case study. The data mining functions of a commercial system are utilized to process the input data and accomplish appropriate data mining techniques such as clustering, classification, and associations. Particular emphasis is given on the importance the initial preparatory/design phase has to the effectiveness of the subsequent data mining processing stages.
- **ML-Visualization:** The visual interface and its importance in the discovery of patterns in large data sets. The cases of at least two modern commercial environments are considered.

II.2.2 Summary of Educational Partners' Existing Labs

Here is a short summary of the aims of the existing demos, exercises and labs; data was collected from educational partner institutions in spring 2003. More detailed information is also provided. Each educational partner has created descriptions of existing laboratory exercises. The description forms are available in the Appendices.

Educational partners have given the information based on the legend presented below the summary table 3. Descriptions for the symbols are presented after the table.

	Reut-lingen	Mala-ga	Helia	Hamk	TEI	Mace-donia	Paisley
DB Administration Labs (AL)							
AL-Concept			g3	s2	s5	s3	s3
AL-Config			g3	g3	5	d5	s3
AL-DbObj			g3	g3,d3	i3	6	6
AL-Security			g3	g3,d3	5	s5	s3
AL-Catalog			g3	g3	g3	s5	s3
AL-Diagnostic			d3	d3	5	s6	6
AL-Tuning			g3	g3,g4	g	s3	g3
DB Design Labs (DL)							
DL-Create			s3	s3	s3	s3	g0
DL-Example			s3	s4,s5	i3	s5	g0
DL-Tools			s3	s3	5	s3	g3
DL-Doc			s5	s3	5	s5	g3
DL-Tuning			g3	g4	5	s3	g3
DL-DbObj			s3	g5	i3	s5	6
DL-Admin			s3	s4,g4	i3	d5	g3
DB Programming Labs (PL)							
PL-ISQLq			s3	s1	5	s3	g3
PL-Tx			g3,s3	s3	s3	s5	g3
PL-APIs			g3,s3	g4	g3	s3	g3
PL-Exception			g3,s3	g4	g3	s5	g3
PL-Sp			g3,s3	g4	i5	s3	g3
PL-SQLJ			d3	g5	g3	s5	6
Data Warehouse Development Labs (WL)							
WL-OLAP				s3	5	d5	g3
WL-Schema				g3	5	d5	6
Data Mining Labs (ML)							
ML-Principles				5	5	d5	w3
ML-Stages				5	5	d5	w3
ML-Market Basket				6	5	d5	w3
ML-Case Study				6	5	d5	w3
ML-Visualization				6	5	d5	w3

Table 3. Existing exercises and labs in partner institutions, spring 2003

About the symbols used in the table 3

About terms

The word "lab" is used here as a generic term, representing any type (see the list of types below). The concept "maturity" measures whether a specific lab exists and if so, the amount of work required if the lab would be provided for public domain of DBTechNet.

Type of the lab

        d   Demo

        g   Guided lab

        s   Self-study lab

        i   Independent practical

        w   Written exercise

Level of maturity

0) The lab is in use, ready to export, English instructions exist for setting up/configuring the infrastructure and English instructions for teachers and students exist.

1) The lab is in use, instructions exist for infrastructure set-up, for teachers and students instructions. Instructions are in native language only.

2) The lab is in use, instructions exist for teachers and students (in native language).

3) The lab is in use, student instructions exist (in native language).

4) The lab has been in use/trial, but not at present. The ideas could be re-used after a re-formulation in another infrastructure, with more man power etc.

5) Is planned to implement or is under development

6) Seems interesting, but no plans exist

II.2.3 Different Ways to Use Hardware in the Labs

If a database server system is administered by student in a lab, this normally requires strong permissions even on operation system level. To grant any strong permission to students may be risky, if the environment has to remain intact otherwise. Even unattended accidental mistakes might jeopardize the operating environment. It follows that there is a clear need for isolated environments, if some administrative tasks are needed in the exercises. Some of the exercises can be performed using client software installed on

workstation environments. Some other exercises can be performed using web browser only.

Examples of hardware and database management system (DBMS) infrastructures:

1.   DBMS installed on each student workstation.
2.   DBMS installed on a removable disk to be used in a student workstation.
3.   DBMS installed on a server machine in school and accessed over LAN only.
4.   DBMS installed on a server machine in school and accessed over WWW.
5.   Student's home installations.
6.   DBMS installed on a removable disk to be borrowed for student's home use.
7.   DBMS lab environment provided by a DBMS vendor.

II.3 Descriptions of DBTech Pro Workshops

II.3.1 Introduction to Workshops

The project required a robust solution that would identify both contemporary and future knowledge areas. This task was assigned to Work Package 2. A survey-based approach was used to identify the current roles that were in use in industry within the partner countries but would also seek to identify the database skills that companies thought were lacking in both recent graduates and employees and their plans for future database development.

Four specialization areas have been identified to play a key role in determining job competence for the database technology professional today, namely:
–   Data Access Technologies
–   Database Administration
–   Database Design
–   Business Intelligence (OLAP, Data Warehousing, and Data Mining)

The topics of the four workshops were targeted to the knowledge areas according to the table 4.

Work-shop	Topics	DB Administration	DB Design	Data Access	Business Intelligence
WS1	Concurrency control and recovery	x		x	
WS2	Database design	x	x		
WS3	DW, OLAP and knowledge discovery			x	x
WS4	XML and databases		x	x	

Table 4. Workshop topics vs. Knowledge areas defined in WP 2

The 1st workshop on "Concurrency Control and Recovery" was held in Thessaloniki, Greece, in November 2003. The host organization was T.E.I. of Thessaloniki. The laboratories were built and conducted and the presentations were given by the Finnish educational partners. The author of the first workshop day was Jaakko Rantanen from HAMK and the author of the second day was Kari Silpiö from Helia.

The 2nd workshop on "Database Design" was held in Malaga, Spain, in May 2004. The host organization was University of Málaga. The author and conductor of the first workshop day was Thomas Connolly from University of Paisley. The author and lecturer of the second day was an independent consultant Tapio Lahdenmäki. The laboratory used a database server located in Hämeenlinna, Finland. The laboratory database was built and tested by a team of five persons, of which three persons were highly experienced professionals from industry. The laboratory session was conducted by Jaakko Rantanen, the creator and administrator of the server environment.

The 3rd workshop on "Data Warehousing, OLAP and Knowledge Discovery" was held in Reutlingen, Germany, in November 2004. The host organization was Reutlingen University. The laboratories were built and conducted and the presentations were given by the Greek educational partners. University of Macedonia and T.E.I of Thessaloniki. The author of the first workshop day was Georgios Evangelidis from University of Macedonia, assisted by Leonidas Karamitopoulos. The author of the second day was Dimitris Dervos from T.E.I. of Thessaloniki, assisted by Leonidas Karamitopoulos.

The 4th workshop on "XML and Databases" was held in Helsinki, Finland, in April 2005. The host organization was Helia. The laboratories were built and conducted and the presentations were given by Martti Laiho and Matti Kurki from Helia and Fritz Laux from Reutlingen University. Additionally, one extra seminar day was held, providing several specialist presentations.

II.3.2 Workshop Labs and Exercises

Here follows a brief description of each workshop. Every workshop has one covering topic and two sub-topics. The existence of two sub-topics indicate that each workshop consists of two modules. Each module takes one day that begins with theoretical lectures and continues with practical laboratory sessions.

It is possible that, for certain workshops, the students are given some pre-study material or assignments, maybe some weeks before the workshop. Any types of post-workshop assignments as well as some tests may also be used.

II.3.2.1 The 1st Workshop, November 2003 in Thessaloniki

<u>Title</u>: Backup and Recovery, Concurrency Control
The workshop consists of two modules, together covering a basic principle called ACID. The modules use different types ob laboratory exercises and have different target knowledge areas. Still, both module topics are tightly bound together because of a common backgrond principle.

The ACID Principle as a Common Background
The core features of a DBMS are targeted to

- avoid the problems of far too slow mechanical storage devices, i.e. disks.
- be able to recover to the last consistent state – not only to a last backup.
- allow large amounts of simultaneous access and updates from client processes.
- The objective is to observe the behaviour of the core features in practice and to examine their implementations in some commonly used products.

There are two key questions to be answered in the workshop. 1) What are short transactions in contrast to long-lived transactions. 2) Why is transaction integrity and ACID principle essential in a application designer's and programmer's point of view and for the recoverability of a database, in an administrator's point of view.

The ACID properties in a nutshell:

- *Atomicity*: All actions in the transaction happen, or none happen.
- *Consistency*: If each transaction is consistent, and the DB starts consistent, it ends up consistent.
- *Isolation*: Partial effects of incomplete transactions should not be visible to other transactions.
- *Durability*: Effects of a committed transaction are permanent and must not be lost because of later failure.

Module 1. Backup & Recovery

- The module gives a short introduction to ACID – what it is and why it is necessary.
- The module further concentrates on two ACID properties "Atomicity" and "Durability".
- The main focus is on database administration.

Module 2. Concurrency Control

- The module covers and explains two other ACID properties "Consistency" and "Isolation".
- The main focus is on application design.

Both modules consist of 2 - 4 hours of introductory lecture and circa 3 - 5 hours of guided hands-on laboratory exercises.

Some pre-study textbooks

There are at least a dozen of useful academic or college level textbooks written in English, covering the background information. Here are some examples.

Connolly & Begg: Database Systems [2]

19.1 Transaction Support

19.3 Database Recovery

19.5 Concurrency Control and Recovery in Oracle

Date: Introduction to Database Systems [7]

15. Recovery.

16. Concurrency

Elmasri & Navathe: Fundamentals of Database Systems [5]

17. Introduction to Transaction Processing Concepts and Theory.

19. Database Recovery Techniques.

Ramakrishnan & Gerke: Database Management Systems [6]

18 Transaction Management Overview

20 Crash Recovery

Description for Module 1: Backup & Recovery

Motivation

Computers today use two types of storage: volatile, fast accessible main memory and non-volatile but rather slow mechanical disks. One of the primary goals of a full-blown database management system is to hide this kind of complexity and provide a stable and effective storage system.

A full recoverability does not normally come out-of-the-box. Any proper DBMS installation requires that a database administrator is able to plan and define the complete database configuration.

The DBA needs to clearly understand some key concepts and principles: Why does the transaction log make up the most of recoverability? What is the purpose and precise behavior of data and log buffering as well as committing and check pointing? What are active and archive logs and how is backuping related to logging? What consist of a general, product-independent recovery strategy? What are the common failure types? What are the storage related structures and terms in some common products? Which are the principles to establish a recoverable environment that duplicates data, log and configuration information in e.g. Oracle and SQL Server?

A short discussion about some common open source products: MySQL and PostgreSQL; What is the current phase in development towards recoverability?

Prerequisites

The module is designed for university students who have theoretical knowledge of a recovery algorithm (ARIES). The module is also suitable for persons with some practical background, who do not know ARIES algorithm well, but are somewhat familiar with database server concepts like transaction, logging, buffering, etc.

Module content

1. Theoretical introduction
- Introducing the primary concepts
- Problems identified
- Backup & recovery strategies in a nutshell
- Overview on vendor approaches
- Orientation to the practicals

2. About the products
- Various terms/names being used
- Where are data files, transaction logs, configuration files
- Check pointing
- Different structures (database, tablespace, file sets,…) need to be studied
- Authority issues

3. Hands-on lab tasks.

Solid Embedded Engine was used. It is a full featured, small footprint database server system, having a special feature: it can be run as an user application on the operating system (as an alternative the normal way: as a system service).

This impes that no administrative permissions are required on operating system. Still the server process can be suddenly terminated without violating the operating system. Thus, even an electric blackout or a media failure can be simulated safely.

The students were given detailed instructions how to observe what happens on technical level in the following situations.

a)   Transaction integrity and logging. Observing the behaviour of Write-Ahead-Logging protocol (WAL).
   - What constitutes the commit operation?
   - How are incomplete transactions handled?
   - What constitutes the checkpoint operation?

b)   Observing the full recovery from a crash (simulation).

c)   Crash simulation during the roll-forward recovery process.

d)   Backup (online/offline/scheduled) and restore (manual, offline).

Description for Module 2: Concurrency control

Motivation

The module focuses on the concurrency control and traditional transactions (i.e. flat transactions) from the viewpoint of business oriented OLTP applications. A banking system with thousands of concurrent users or an eCommerce system on an Internet server, that has an unpredictable number of users, highpoint the importance of concurrency control.

In a multi-user environment a DBMS has to interleave the actions of several transactions. The aim is to preserve illusion that each user is executing alone. To fullfil the business requirements the DBMS should protect the database consistency without sacrificing performance. Traditionally, transactions are expected to satisfy the ACIDity properties. The concurrency control mechanism is responsible to provide the means to guarantee that a transaction is isolated from the effects of concurrently scheduling other transactions.

ISO SQL standards (SQL-92, SQL:1999, and SQL:2003) are important vendor-independent references on SQL and RDBMS-related issues. They specify the SQL transaction model and define four transaction isolation levels in terms of three phenomena and the concept of serializability. SQL-92 specifies flat transactions only. SQL:1999 adds savepoints to the model and specifies the concept of serializability more clearly. In addition to the ISO SQL isolation levels there are some other popular transaction isolation levels (Cursor Stability, Snapshot Isolation) in commercial systems.

The main DBMS vendors have two different approaches on implementing concurrency control. They use either strict two-phased locking (or variants, if a transaction requires a lower than serializable isolation level) or multiversion concurrency scheme. The latter provides the snapshot isolation level. For choosing a proper isolation level it is important to understand the basic ideas of how a DBMS protects consistency on each isolation level and what effect each isolation level has to the concurrency rate.

To write database consistency preserving transactions with good performance, among some of the important issues to understand are:

- The ACIDity properties of transactions
- The universal transaction correctness criteria (serializability)
- SQL-transaction isolation levels and possible anomalies
- Most popular other transaction isolation levels
- The cause and effect to both consistency and performance when choosing a transaction isolation level in a particular case
- The cause and effect to both consistency and performance when using any vendor-specific primitives that affect to transaction isolation (this is not discussed in this module).

Prequisite

For this module, familiarity with the following concepts is required:
- The concept of a transaction
- Atomicity, consistency, isolation, durability
- The SQL-92 transaction characteristics
- Basic ideas of concurrency control
- Some anomalies associated with interleaved execution
- The concept of locking, shared and exclusive locks, locking granularity, deadlock

Learning Objectives

TOPIC	LEVEL (Bloom)
The SQL transaction model The SQL transaction syntax & semantics	application
Basic approaches on concurrency control - locking / multi-versioning	knowledge application
Isolation levels (ISO SQL + other) vs. - degree of isolation - degree of blocking	application
Vendor implementations - behavioural differences	knowledge comprehension
The cause and effect to both consistency and performance when choosing a transaction isolation level in a particular case	application --> (later) analysis

Table 5. Learning objectives for Module 2: Concurrency control

Module content

1. Pre-study
   - See perquisite
   - Module web page (Student's guide + slides) to be browsed beforehand
2. Theoretical introduction (3+ hrs)
   - Motivation, orientation
   - Recapping main concepts
   - Discussing the main theme & orientation for the lab
   - Presentation content (see the slides):
   - Orientation
   - Motivation
   - The SQL transaction model
   - Serializability
   - Locking
   - The SQL-92 transaction isolation levels
   - Multi-versioning
   - Concurrency control schemes (Vendor implementations)
   - Concurrency conflicts
   - Phenomena
   - Phenomena vs. snapshot isolation
3. Lab (3 hrs)

   Verifying the theory

   At least two DBMS are accessed using the following tools:
   - An interactive SQL editor
   - Helia's TR_demo application

   Hands-on observations on
   - The phenomena: anomalies, concurrency conflicts
   - DBMS products' behavioural differences
   - The cause and effect to both consistency and performance when choosing a transaction isolation level in a particular case
4. Aftermath (not implemented in the workshop)

   Recapping the lab

   Exercises

   Summary

Feedback from workshops at Thessaloniki November 2003

General

Workshop days at Thessaloniki 28 – 29 November 2003 succeeded well. This was the first time during DBTech Pro –project we tried in practice multicultural learning situations. Teachers Jaakko Rantanen and Kari Silpiö came from Finland and students were from Greece, T.E.I and University of Macedonia. The object was that professors from different countries could take these modules as a part of their own teaching program.

The leaders of T.E.I mentioned in their welcome words that globalisation and becoming internationalized means that we need more and more information. Data and databases are the cornerstones of data processing.

Martti Laiho presented the history, aims and principles of DBTech-project. He also invited students "to think how and why". The analysis was successful. Only Martti's voice was too low in a large auditorium.

Backup & Recovery, Jaakko Rantanen 28 November 2003

Jaakko Rantanen began his teaching by lectures in the auditory. He shared the slides printed on paper. Jaakko Rantanen presented good and illustrative examples of his subject. The key words of his subject became clear for listeners.

Students observed with concentration but how could we get more interactivity in our lessons? Perhaps more eye contact would be needed. At the beginning there could be wider introduction to the subject. A common manner of Finnish lecturers is to go directly to the message. Also at the end of a presentation we often forget a short repetition and summary.

At the end Jaakko Rantanen compared different systems in very interesting and expert way. He used precisely the time reserved for the lecture.

The subject continued in the afternoon at the laboratory. All the students had not got the material that should have been delivered some days before workshop. Technical troubles at the beginning caused that we could not keep in time. Students' Solid user-codes did not

operate. It took more than half an hour to start. Jaakko Rantanen gave printed directions for laboratory and presented the tasks also with data projector.

After troubles at the beginning, students worked diligently and with concentration. At the end of evening Martti Laiho presented a demo of Solid recovering methods.
The feedback of students was very good.

Concurrency Control, Kari Silpiö 29 November 2003
Kari Silpiö began his 2 hours lecture with a slide of key issues. Kari also shared his material printed on paper and presented slides with data projector. Kari's lecture was excellent. He used illustrative examples and his way to keep lecture was compelling. There were new subjects not only for students but also for professors. Nutshell tables were a good work-sheet and comparison of DBMS SQL Server, DB2, Oracle 9 and Solid was informative.

At the end of the lecture there was an interactive moment where interesting questions and opinions were put to the lecturer. Unfortunately the ordinary students were not very eager to ask questions. Surely the other teachers will pilot this module in their own teaching.

In the afternoon we had 3 hours practical laboratory. In spite of Kari's careful preparation there were again big technical problems because of water catastrophe in the computer hall.

Both participating professors and the students liked Kari's workshop very much. The feedback of students was even excellent (see appendix 1).

Summary from the evaluation
According to the different kind of feedback the workshops in Thessaloniki succeeded well. Database expert areas need much work before the modules are ready for put into practice. This kind of cooperation can save teachers' time and even money. It is possible to run laboratories through the Internet. Product development can be shared in different countries and even these starting lectures could be seen through the Internet.

Students said that it was easy to follow Finnish teachers lecturing in English, much easier than lectures of native English persons. When keeping lectures in English it is important to remember to speak clearly and slowly enough. The teacher ought to reserve about 1,5 –2 times more time to keep lectures and workshops in English.

In laboratories there ought to be less tasks or considerably more time. It is not nice if half of the problems remain unsolved. There should be also time to ask and discuss after lessons and laboratories. Students in Thessaloniki were satisfied with the fact that there was a possibility to work and discuss with students from different collage. Particularly the students of university were satisfied with practical laboratories.

Many IT students have fast connections also at home so it is possible to distribute materials through the Internet. Unfortunately the connection time is quite expensive.

For next workshops we ought to remember to
- take care of paper and Internet material in advance (in time and distributed).
- prepare students so that the initial level is suitable for workshop.
- take time enough for practical labs or less problems to solve.
- make work-sheet of different preparing tasks and work according to that.
- make user-codes for all the participants.
- test labs beforehand (machines, connections, programs,...).
- make clear agendas and decision proposals for shorter meetings.

As a summary our students were very satisfied. "This course was a great experience for me. It gave us the change to get more information on the course with a very interesting way. Everything was very nice and so interesting." "I am very happy that I could have this opportunity to participate in DBTech_Pro."

II.3.2.2 The 2nd Workshop, May 2004 in Málaga

Title: Database Design Methods – Conceptual and Physical
The workshop consists of two modules. The first module concentrated on the start phase of the database design process, and the second module concentrated on a very late phase of the process: improving the performance of an implemented database.

Description for Module 1:
Requirements Elicitation and Database Design Methodology

Motivation (1) Importance of Database Design Methodologies
The database is now the underlying framework of the information system and has fundamentally changed the way many companies and individuals work. The developments in this technology over the last few years have produced database systems that are more powerful and more intuitive to use, and users are creating databases and applications without the necessary knowledge to produce an effective and efficient system.

To produce a system that will satisfy the organization's information needs requires a data-driven approach, which means we think of the data first and the applications second. For the system to be acceptable to the end-users, database design is crucial. A poorly designed database will generate errors that may lead to bad decisions being made, with potentially serious repercussions for the organization. On the other hand, a well-designed database produces a system that provides the correct information for the decision-making process to succeed, in an efficient way. In this module we present a methodology for conceptual and logical design as a series of simple-to-follow steps, with guidelines provided throughout.

Unfortunately, database design methodologies are not very popular, which may be a major cause of failure in the development of database systems. Due to the lack of structured approaches to database design, the time and resources required for a database project are typically underestimated, the databases developed are inadequate or inefficient in meeting the demands of users, documentation is limited, and maintenance is difficult. We hope the methodology presented in this module will help change your attitude.

Motivation (2) Importance of UML
Increasingly, companies are standardizing the way in which they model data by selecting a particular approach to data modeling and using it throughout their database development projects. A popular high-level data model used in logical database design, and the one we use in this module, is based on the concepts of the Entity-Relationship (ER) model. Currently there is no standard notation for an ER model. Most books that cover database design for relational DBMSs tend to use one of two conventional notations:

- Chen's notation, consisting of rectangles representing entities and diamonds representing relationships, with lines linking the rectangles and diamonds;

- Crow's Feet notation, again consisting of rectangles representing entities and lines between entities representing relationships, with a crow's foot at the end of a line representing a one-to-many relationship.

Both notations are well supported by current CASE tools. However, they can be quite cumbersome to use and a bit difficult to explain. In this module, we instead use the class diagram notation from the latest object-oriented modeling language called UML (Unified Modeling Language). UML is a notation that combines elements from the three major strands of object-oriented design: Rumbaugh's OMT modeling, Booch's Object-Oriented Analysis and Design, and Jacobson's Objectory. It is anticipated that the UML will become a standard and the Object Management Group (OMG) has adopted the UML as the standard notation for object methods. We believe you will find this notation easier to understand and use.

Pre-requisites

The module is designed for University students who have a theoretical knowledge of relational DBMSs as well as knowledge of ER modelling and normalization.

Objectives

After completing this module the student should be able to

- Discuss the stages of the database application lifecycle.
- Discuss the various fact-finding techniques that are used during the database application lifecycle.
- Select and use appropriate fact-finding techniques that are relevant in the early stages of the database application lifecycle.
- Produce a requirements specification and systems specification during requirements collection and analysis.
- Use appropriate UML notations to represent the organization in the early stages of the database application lifecycle.
- Discuss the three stages of database design (conceptual, logical, and physical database design).
- Perform conceptual and logical database design.

Knowledge Elicitation

In this module, students work in groups of 3-4 on a particular case study. It is suggested that prior to the day of the workshop, groups be provided with an overview of the case study. Each group will then prepare an appropriate fact-finding technique to obtain more detailed requirements from the client on the day of the workshop (for example, interview, questionnaire). One session of 30 minutes duration should be allocated for each group for this purpose after which the groups should be given time to produce an appropriate conceptual and logical database design. A further session of 30 minutes duration should be allocated for the afternoon session to allow each group to clarify any issues that may have arisen in the morning session.

At the end of the allocated period each group should be given about 15 minutes to present their design to the clients. Any clarifications or alterations requested by the client will be implemented after the workshop and a final design submitted for assessment.

Case studies and role playing as clients

To run this module, you need to find one or more people to act as clients for each case study. These people should familiarise themselves with the case study and be prepared to be asked various questions about the requirements. To help, two case studies have been provided: *StayHome* Video Rental case study and the *Wellmeadows* Hospital case study. However, other case studies can be used, as required (although it is suggested that a similar type and quantity of information be provided).

Some pre-study textbooks

- Connolly T.M. and Begg C.E. (2001). Database Systems: A Practical Approach to Design, Implementation, and Management, 3nd edition, Addison Wesley
- Connolly T.M. and Begg C.E. (2003). Database Solution: A Step-by-Step Guide to Building Databases, 2nd edition, Addison Wesley

Pre-study material

The students received the following excerpts form the first of above mentioned textbooks. The preparation materials for the students to read were:

- Chapter 10 (in file DS4_Chapter10.pdf), which covers fact finding techniques
- Section 25.7, which covers use of UML in the database design methodology

For completeness, the students might find useful the following files:

- Chapter 9, which covers the stages of the database application lifecycle
- Chapter 15, which covers conceptual database design

- Chapter 16, which covers logical database design.

Exercises

The workshop contains two interactive, role playing, sessions to allow students to interview "clients". 3 groups, each with about 3 students, worked on each case study.

Two case studies were used for the workshop: a Wellmeadows Hospital case study and a StayHome Video Rental case study. Each one outlines what the students have to do to prepare for the workshop and what's expected on the day of the workshop.

Teachers attending the workshop were acting as clients. There were more complete versions of the case studies available for the teachers, that defines the types of data held by the organization along with sample transactions. An element of the interviews was an ability to think on each teachers own feet, as students may ask questions that wasn't planned for.

Description for Module 2: Introduction to Systematic Index Design

Motivation

IT students often do not realize that performance is still a big issue with large database applications. They also tend to be surprised to learn that inadequate indexing is the most common reason for unacceptable response times.The labs in this module are exceptionally realistic to give the student a feeling of the dramatic difference between an inadequate index and a good index.

Content

Systematic index design consists of three tasks:
- Identify SQL calls that are two slow due to inadequate indexing
- Derive reasonable index candidates for a given SELECT with given predicate filter factors
- Choose the right candidate – taking into account hardware cost and the elapsed time for INSERTs, UPDATE and DELETEs.

B-tree indexes are very similar in the relational DBMSs. Therefore, the methods for index design are to a large extent product independent. However, as index design is always a

trade-off that involves the capacity, the cost and the speed of hardware components (processor, RAM and disk drives), many recommendations that are valid in one point of time are obsolete and harmful in ten years. This is an acute problem today as many often-repeated warnings are legacies from the 1980's when disk and central storage (RAM) were two or three orders of magnitude more expensive than today.

A good starting point for index design is an E/R model, which shows the primary key and the foreign keys of each table. Version 0 indexing should be derived from that model mechanically: primary key indexes, foreign key indexes and eventual candidate key indexes. Then the hard part begins: which columns must be added to these indexes and which indexes must be added to make all SQL calls fast enough (or actually: all programs fast enough).

If an application must perform well from day one, the elapsed time for each program should be estimated early. If it is not short enough with the current index version, according to a worst-input estimate, indexing should be improved. This is a feasible approach if the estimate formula is simple enough – as the Quick Upper-Bound Estimate (QUBE), which gives an "up to x seconds" estimate with a formula that has only two input variables (the number of random touches and the number of sequential touches). Experience shows that using QUBE adds 5 to 10% to the implementation cost of an application. However, it requires several days of training for application developers and that is why many companies do not use the method in this way. Basic Question (BQ) is a low-cost alternative that reveals many (but not all) index needs: the programmer must check whether all predicate columns of an SQL call are copied to one index. If not, an alarm is sent to a DBA.

It is much easier to detect inadequate indexing after the application is in production. Most RDBMSs include tools for producing exception reports based on the elapsed time of an SQL call – or, much better, the elapsed time of an interactive transaction. If the report includes data like the number of database pages accessed, it is relatively easy to find those SQL calls that are too slow because of inadequate indexing. As a valuable by-product, those SQL calls that are too slow because the optimizer chose a wrong access path are detected, as well.

As SQL calls have become more complex, deriving the best index for a SELECT is no longer a trivial task. Designing good indexes for a join can be particularly difficult – at least when several tables have local predicates. This is because the index designer should

make an assumption of the best join method and, in a nested loop join, the table access order. On the other hand, the optimizer's decision is, of course, strongly influenced by the indexes that have been created. There are now methods (and also tools based on these methods) that suggest a good – often the best – index for a SELECT with given predicate filter factors. These methods are not time-consuming even when used manually. The hardest part is making the decision: should we create the best possible index for a SELECT – often a new index – or should we just add new columns to an existing index to get a fairly good index (semi-fat or fat). Among the prerequisites for making wise decisions are an ability to quickly predict elapsed times (QUBE or sometimes Basic Estimates for better accuracy) and an awareness of current hardware costs and limits.

Objectives

After completing the modules, the student should be well prepared to study the product-specific implementations of B-tree indexes. After learning to use the tools for monitoring the optimizer and the execution of an SQL call, the student should be able to design good indexes and to improve existing indexes in a manner that improves the over-all performance of an application or a set of applications.

The student should also gain an appreciation of the importance of well-designed indexes as well as a critical attitude for unqualified recommendations.

Approach to the methodology

The trend in many organizations seems to be that indexes are proposed – or even implemented – by application DBAs or application developers. This is because it is often impossible for the top database specialists to know all applications well enough to design indexing. The database specialists tend to be busy with many other urgent tasks. Fortunately, with the right approach, an application developer who knows SQL can become a productive and successful index designer in about one week.

Part 1. Basic Concepts & Background

The part covers the basic concepts that are necessary for understanding index design – such as filter factors, optimizer issues, and basic CPU and I/O estimates. The students are also to research database literature for index design methods and recommendations.

Part 2. How to Design Good Indexes for Contemporary Hardware

The part concentrates on how to derive a good index for a SELECT with given predicate filter factors. A two-candidate algorithm for single-table SELECTs is introduced as well as the low-cost alternatives for the best possible index. The alternatives are evaluated with QUBE (Quick Upper-Bound Estimate). BE (Basic Estimate) is introduced as well.

After the students have designed an index for a single-table SELECT (and estimated the worst-input elapsed time), they will create the index and measure the elapsed time with an average input and with the worst input.

The last part of the workshop prepares the students for the post-workshop assignment: Design good indexes for an inner join with local predicates to two tables.

Part 3.Systematic Index Design

The part focuses and how to detect the SQL calls that need better indexes – at design phase and when an application is in production. Other topics include a discussion of pitfalls of measurements in a test environment and a review of the state-of-the-art tools for index design.

Part 1 is a reading and research assignment, which should take about 2 days. Part 2 is a workshop (8 hours) with a couple of short paper labs and a 2-hour machine lab. Part 3 (2 days) is a reading assignment with a machine lab (design indexes for an inner join).

Prerequisites

The modules are designed for university students who are familiar with SQL programming and table design.

The hands-on lab

The hands-on lab was provided for students to be utilized during the workshop day, and after the workshop as well. A part of the assignments were planned to be completed after the workshop. The database server to be used in tuning for performance was located in Hämeenlinna, Finland. The server system is a quite typical small or medium side company database server. The students accessed the database from Malaga using a virtual private network connection, for security reasons. Figure 3 describes the structure.

Partner institutions

Accessing the DBMSsthrough firewall

Internet

Server in Häme Polytechnic

Hardware

- 2 x Pentium III, á 1GHz
- 2 GB memory
- 4 disks in a RAID 5 array

OS: Windows 2000 Server

DBMSs among others:

- Oracle9 EE
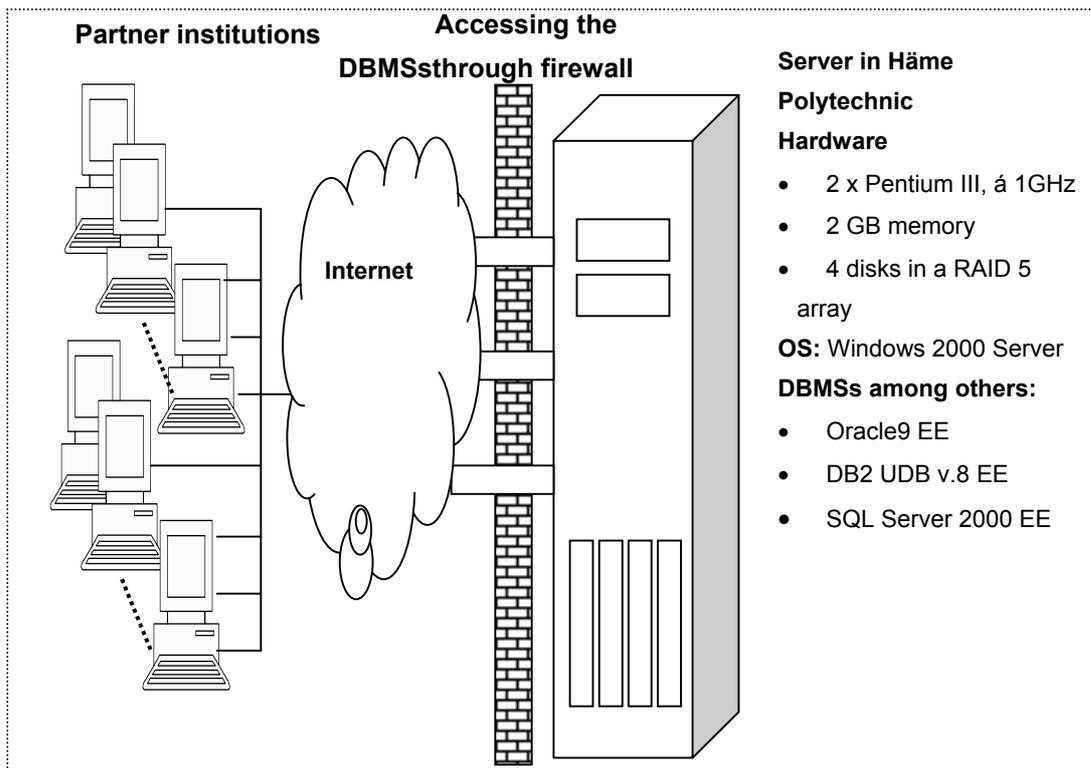- DB2 UDB v.8 EE
- SQL Server 2000 EE

Figure 3. Structure of provided hands-on lab

The server runs nine similar databases for a maximum of nine groups of students. Each database contains two tables: table "Client" containing 1 million rows and table "Invoice" containing 4 million rows.

The lab was used by six groups of three or four students in each. Each student group had its own schema, i.e. own database tables to use as follows: the group could create and drop indexes and run select commands to measure the elapsed time.

Any select could be executed so that no large result sets are transmitted over internet. Instead, the user retrieved only the following information: the elapsed time and the number of rows the DBMS was reading. The data volumes are such that it takes 1...5 minutes to create any index. Original "too slow" selections, that are to be tuned, may last a few minutes.

The database server environment was available about one month for the Malaga workshop participants, to repeat the same exercises and to do the post-workshop assignments. After this period the connection permissions were revoked for security reasons.

Feedback from the Workshop in Malaga, 27 – 28 May 2004

Summary

The workshop was a very good experience for both students and partners. The selected methodology was very appropriate, and students were very happy with the lessons they received. The workshop has satisfied student expectations and has given them a different vision of database and index design. The material for the students was sufficient and very well designed. The division of the material in pre, at and post material was appropriate, because students could study the main issues of the workshop topics before the lessons.

The general acceptation of the module demonstrated the adequacy of the contents to the workshop. Contents were clear, useful and appropriate. However, some students thought that contents provide too much information at the same time. The students also thought the contents will be very useful in the business market.

Unfortunately, the methodology of the second day didn't satisfy students and the majority complained about the duration of the theoretical part. More practical activities would be more interesting. The first day-, the practical part was more important than the theoretical one, and for this reason student interest was very big. However, it is very difficult to assimilate in only one day of the workshop all the previous material and the new material of the presentation. In consequence the attention span decreases very easily during the session. In general, students consider the second day to be a very hard day.

II.3.2.3 The 3rd Workshop, November 2004 in Reutlingen

<u>**Title**</u>**: Data Warehousing and Knowledge Discovery**

The workshop consists of two modules. The first module concentrated on data warehousing and OLAP technologies. A data warehouse is a federated database that contains data that represents the big picture of business data. OLAP technology uses data warehouses for online analysis that has to do with iterative complex analytical queries.

The second module concentrated on Knowledge discovery from databases: Data mining. The purpose is to find new, not obvious, relationships and patterns in the data.

Description for Module 1: Data Warehousing and OLAP

Aims & Objectives

The main objective of this one day workshop is to introduce the audience to two important components of modern Decision Support Systems (DSS) or Business Intelligence Systems, namely, the data warehouse and OLAP technologies. A data warehouse is a database that contains data that represents the business history of an organization. OLAP technology uses data warehouses for online analysis that has to do with iterative complex analytical queries.

The above concepts will be presented at a theoretical level together with a demo presentation. In the laboratory, the participants will have the opportunity to acquire hands-on experience by working on data from a real application (a given data warehouse) and will attempt to play the role of an analyst by searching for answers to given questions and exploring (navigating) complex business data relationships.

Content Outline (Theory)
- Data warehousing overview
- The database of a data warehouse
- Historical data
- Interaction, response times
- Data preparation, loading, synchronization
- Semantic integration of data
- Heterogeneous sources
- Loading, refreshing, purging of data
- Metadata management
- Multidimensional data model
- OLAP overview
- OLAP vs. OLTP
- ROLAP, MOLAP, HOLAP
- Dimension hierarchies
- OLAP queries (aggregations)
- Comparison with SQL

- CUBES
- Design Issues (Star schema)
- Implementation issues (bitmap indexes, join indexes, compression)
- Online aggregation

Content Outline (Lab)
- Software: MS-SQLServer 2000 Data Analysis Tools
- Participants will utilize real data and play the role of an analyst.

Timetable

Theoretical Introduction – 2hrs

Hands-on Laboratory Workshop Session – 3 hrs (plus time for breaks)

Description for Module 2: Data Mining

Aims & Objectives

The main objective is to introduce the audience to the new, qualitatively different, way of using the modern database system, i.e. how to advance from the (present) verification-driven approach to the (new) discovery-driven one. Data mining is to be presented not to the detail of its (numerous) algorithms and techniques, but rather at the higher (business intelligence) level, i.e. how one can take advantage of the new tools that conduct data analysis in a way that makes possible the discovery of new information (knowledge) that is beyond reach for the conventional SQL-type query data probing and processing.

Content Outline (Theory)
- Compare: data vs. information
- Need to interpret data, identify patterns & trends, categorize cases
- From machine learning (1950-) to knowledge discovery in DBs (2000-)
- KDD stages
- KDD: A multidisciplinary field of research
- KDD and data mining
- Data mining vs. SQL
- Problems relating to KDD applicability
- The empirical cycle of KDD applicability
- Real life applications that call for knowledge discovery
- Data mining: application examples

- Typical data mining techniques (an overview)
- Association rules: basic concepts
- Support, confidence, expected confidence, lift
- Association rules: an example
- The apriori algorithm: key observations, method, pseudo-code
- Clustering: basic concepts
- Distance functions
- A simple clustering algorithm: BIRCH
- Data mining vs. data warehousing & OLAP

Content Outline (Lab)
– Software: IBM's Intelligent Miner for Data v.8.1
– It is asked to conduct data analysis and propose strategies for tackling real-life problems, like: 'What is the target audience on which to focus my advertising campaign for new product X, based on the analysis of the existing customer data?'

Timetable
Theoretical Introduction – 2hrs
Hands-on Laboratory Workshop Session – 3 hrs (plus time for breaks)

Feedback from the Workshop in Reutlingen 2-3 November 2004

General
A pre-pilot for the Reutlingen workshop was held in Helsinki in September 2004. So it was interesting to compare the results of these two feedback inquiry. In general students at Reutlingen University gave very few comments or arguments in their feedback. Also conversation and dialogue during these two workshop days was quite small.

As a whole the workshop succeeded well. Teachers Georgios Evangelidis and Dimitris A. Dervos both gave first lesson about three hours and after that there were 3 hours laboratory on both days. The subject of Georgious Evangelidis was Data Warehousing and OLAP and Dimitris A. Dervos lecture handled Data Mining. So the agenda was about the same as in Finland a few weeks earlier. Only some changes in contents were made.
Data Warehousing and OLAP by Georgios Evangelidis

The lab succeeded well. If we try to find some criticism so the material was not available before and during lesson and laboratory works there were very few questions and conversation.

The average of answers both in Helsinki and in Reutlingen was quite similar. There were no noticeable differences between answers. The feedback altogether was excellent.

Data Mining by Dimitris A. Dervos

The most noticeable difference between Finnish and German students was that in Reutlingen students were more critical concerning the Data Mining workshop.

Remarkable was that the project members and partners estimated Data Mining workshop considerably better than the students. Perhaps this really was a little bit too strange and difficult subject to learn during this short time. The initial level of students seemed to be insufficient and they haven't got material beforehand to become acquainted with the subject. That's why they estimated their learning process not to be the best possible.

Summary

Just as in Helsinki, the real dialogue during both lessons and labs was quite small. How could we get these learning situations more interactive? Another question is the basic level of students. Is it possible to train our students for a workshop day? One answer might be a training material which would be distributed early enough for students.

Perhaps we can repeat the apophthegms from Helsinki pre-pilot:

1. Take care of paper and Internet material in advance (in time and distributed).
2. Prepare students so that the initial level of everyone is suitable for workshop.
3. Take more time for practical labs and a little bit less for lessons.
4. Test even more the labs beforehand (machines, connections, programs,...).
5. Give more time for possible home works after workshops.

II.3.2.4 The 4th Workshop, April 2005 in Helsinki

Title: XML and Databases

The workshop consists of one seminar day and two additional hands-on workshop modules.

The Seminar Day

Evolving SQL and XML standards

Hierarchical and Semistructured Data, XML and Databases, Standards and Trends, prof. Thomas Connolly, Paisley University

XML and Relational Databases, Standards and Implementations, Antoni Wolski, Solid

XML support by RDBMS Vendors

Oracle 10g and XML, Petri Savolainen, Oracle

DB2 UDB 8 and XML, Ari Rahkonen and Pav Kumar-Chatterjee, IBM

Microsoft SQL Server 2005 and XML, Shankar Pal, Microsoft

XML DB Support expected by the Software Industry

XML, Databases and Applications, Lasse Akselin, TietoEnator

Panel discussion, chaired by prof. Fritz Laux, Reutlingen University

Description for Module 1: SQL meets XML

Aims & Objectives

The main objective of this one day workshop is to introduce the audience to current XML technologies used for accessing data in relational databases starting from XML in the middleware as available for data access technologies in distributed, multi-layer architectures like J2EE and Microsoft .NET.

The above concepts will be presented at a theoretical level together with a demo presentation. In the laboratory, the participants will have the opportunity to acquire hands-on experience by working with XML data access technologies of middleware in the .NET architecture as well as XML supported by the SQL Server 2000/2005 databases.

Pre-requisites

The participants are expected to know basics of relational databases, SQL and XML.

Content Outline (Theory)
- Hierarchical Data Model
- Disconnected Data Processing
- Why XML in databases? Applications?
- Family of XML languages related with database access
- Standards and implementations
- SAX and DOM
- XML Schema
- SQL and XML Mappings
- XPath
- XSLT and XSL
- SOAP and Web Services

Content Outline (Lab)
- Software: SQLXML 6.0 + IIS MS SQL Server 2005
- Exercising with XPath and XML Schema
- URL Access: XPath Query, SQL Query, Template Query
- XSLT transformations
- Security issues
- XSD Schema
- Data type mappings and conversions
- XML in .NET Data Access Programming
- XML view of data in relational tables
- Relational view on XML documents
- OpenXML mapping of XML data into relational view
- Insert, update and retrieve XML data in database, updategrams

Description for Module 2: SQL/XML, SQL/XPath and XQuery

Aims & Objectives

The main objective will be to introduce the audience to the integration of relational and hierarchical data models in modern databases provided by new XML extension of the ISO SQL:2003 Standard and the new XML query language of W3C as implemented by the mainstream relational database products.

Arguments of having a database as part of the service infra for XML storage and processing will be covered: security, integrity, transactional processing, and performance tuning with indexes. The various storage methods as well as the benefits and problems of the database services for XML will be covered.

Pre-requisites

The participants are expected to know basics of relational databases, SQL and XML.

Content Outline (Theory)
-   SQL/XML standard: XML type and functions, XML Schema facets
-   SQL/XML an implementations in Oracle and DB2
-   XML storage options and access technologies
-   Integrating XML and Relational Data
-   XQuery 1.0 and XPath 2.0
-   XQuery tools.

Content Outline (Lab)
-   Software: Oracle 10g, DB2 UDB 8.2, Altova XQuery
-   Participants will learn the basics storing and retrieving XML data in the
-   Mainstream relational DBMS systems, how to implement XML view on SQL
-   Data using the SQL/XML functions, and how to implement SQL view on
-   XML data
-   Participant will experiment with issues of performance and transactional
-   Access of the XML data and the various XML storage and access options in the database.
-   Participants will learn by doing the basics of using the new W3C Xquery language.

Feedback from the Workshop in Helsinki 19 – 21 April 2005

The feedback from these workshop modules was collected from both students and the DBTech Pro partner persons. Nearly all the respondents had received much new information, only a few of them were somewhat familiar with some of the topics before. The timing was very tight. The material was not provided beforehand. Clearly there were difficulties to follow and take in so large amount of information.

The DBTech Pro partner persons were more often able to adopt and apply the information in practice than the students. The students evaluated the technical implementation of the laboratories to be better as what the partners did.

II.4 Summary and Conclusions

The main results of the WP3 consist of eight modules produced for and tested in the four workshops. Additionally, some pilots of the modules have been implemented, either before or after the workshops. In addition to the workshop modules, an other type of result is the collected information on the existing laboratories and exercises that are in used by all educational partners. An attempt for categorization of the laboratory and exercise types is also made.

The present manuscript reports on the outcomes of the DBTech Pro Work Package 3. The eight modules have been developed and tested in practise, at least in the four workshops, some of the modules even more often.

The amount of work invested in developing the modules has not been systematically measured. The subjective judgements of many of the developers strongly implies that the DBTech Pro workshops have become a challenge, to develop and provide some new and useful theoretical material as well as laboratory exercises.

As a start, the first workshop introduced new results from an intensive study, and provided also new useful software tools. In practice, it became maybe even an expectation to put an effort of several man-weeks or even man-months to set up a workshop laboratory module. For the second workshop, the indexing laboratory, that can be used as a benchmark as well, was built in collaboration with highly experienced specialists. In the industry, there is a high demand for this kind of solutions and know-how.

The amount of work in developing the 3rd and 4th workshops grew up to several man-months for each workshop. The workshops have turned out to become highly beneficial and productive ways of collaboration.

Interesting differences in organizations and cultures among the educational partners have brought up many smaller and bigger surprises, and sometimes caused unexpected trouble

and extra work. Some bad accidents have come close, but DBTech Pro has been a lucky project.

The purpose in DBTech Pro project has been to produce rather short and compact modules, which are easiest to adopt by educational partners. One specially advanced module, the Systematic Index Design Method and laboratory, requires at least one week work of studying. A module of this size is rather problematic to fit into existing courses. During the DBTech Pro project, any changes in curricula have not been possible.

A constant threat is hidden behind scenes when one is developing specialized laboratories using complicated and rapidly evolving technology. The threat is the fact that high technology solutions become obsolete very quickly. The laboratories have to be constantly maintained and renewed, which requires a substantial effort.

A fact is that the outcomes of DBTech Pro should require additional development effort, if the produced solutions should be provided more widely to the European academic society.

II.5 Credits

What is being reported in the present document, has been the product of collaboration of the following DBTech Pro partners (listed alphabetically, by last name):

> T.M. Connolly, University of Paisley, U.K.
> D.A. Dervos, Technology Educational Institute (T.E.I.), Thessaloniki, Greece
> G. Evangelidis, University of Macedonia (U.o.M), Thessaloniki, Greece
> L. Karamitopoulos, T.E.I. and U.o.M., Thessaloniki, Greece
> M. Kurki, Helsinki Business Polytechnic (Helia), Finland
> T. Lahdenmäki, www.Tapio1.com, Ljubljana, Slovenia
> M. Laiho, Helsinki Business Polytechnic (Helia), Finland
> F. Laux, University of Reutlingen, Germany
> J. Rantanen, Häme Polytechnic, Hämeenlinna, Finland
> A. Rusanen, TietoEnator Corporation, Finland
> K. Silpiö, Helsinki Business Polytechnic (Helia), Finland
> M. Ståhl, Ilmarinen Mutual Pension Insurance Company, Finland

DBTech Pro has been exceptionally lucky to enjoy the innovative, voluntary and generous work in developing and testing the indexing lab (workshop 2, module2) that T. Lahdenmäki, M. Ståhl and A.Rusanen were willing to provide, for public domain.

II.6 References and Some Essential Bibliography

The below listed bibliography serves as list of references, and in addition it represents a collection of information sources for the workshop developers. It is a brief list of necessary and general sources. Additionally, each workshop material includes a more detailed list of references.

The books [2], [5], [6], [7], [8] and [9] are general database textbooks used in the educational partner institutions. In these books some special issues may be covered so briefly that more deep material would be welcome.

Some additional material is included to strengthen two knowledge areas. The first area is physical database design and the second is data warehousing and data mining. The books [3] in English and [4] in Finnish, introduce the method of systematic index design, that gives necessary background information for the indexing lab (Workshop 2, module2). The books [10], [11] and [12] are some primary handbooks of data warehousing and data mining.

Finally, the database technology standard [13] is quite broad and complicated. It needs a substantial amount of material to be explained, e.g. [14], [15], [16], [17] and [18] are helpful on this manner. The last references [19] and [20] are reminding about the complex marriage of SQL and XML.

[1]   Database Technology Network (DBTech Net, http://www.dbtechnet.org)
[2]   T. Connolly, C. Begg: Database Systems, 2004
[3]   T. Lahdenmaki, M. Leach: Relational Database Index Design and the
       Optimizers, 2005
[4]   J.Huotari, A.Hovi, T.Lahdenmäki: Tietokantojen suunnittelu & indeksointi,
       2003
[5]   R. Elmasri, S.B. Navathe: Fundamentals of Database Systems, 2003

[6]   R. Ramakrishnan, J. Gehrke: Database Management Systems, 2003

[7]   C.J. Date: An Introduction to Database Systems, 2003

[8]   H. Garcia-Molina, J. D. Ullman, J. D. Widom: Database Systems, 2001

[9]   D. Kroenke: Database Processing, 2004

[10]   R. Kimball: The Data Warehouse Toolkit, 2002

[11]   W.H. Inmon: Building the Data Warehouse, 2002

[12]   M.H. Dunham: Data Mining: Introductory and Advanced Topics, 2002

[13]   ISO/IEC 9075:2003 Information technology - Database languages - SQL

[14]   J. Melton, A. R. Simon: SQL:1999 - Understanding Relational Language
        Components, 2001

[15]   J. Melton: Advanced SQL:1999 Understanding Object-Relational and
        Other Advanced Features, 2003

[16]   P. Gulutzan, T. Pelzer: SQL-99 complete, really, 1999

[17]   SQL standards related readings:
        http://www.wiscorp.com/SQLStandards.html

[18]   The SQL Standards Maintenance Portal:
        http://standards.iso.org/iso/9075/

[19]   SQL/XML standard: http://standards.iso.org/iso/9075/2003/sqlxml

[20]   XML documentation: http://www.w3.org/

Part III - Final Report of the WP5; Pilots

III.1 Introduction to WP5

The following report presents the achievements found by repeating the workshops (called pilot runs) developed by members of the DBTech Pro. The results are presented from the instructor's point of view. Comments by the students and their pedagogical evaluation was done in WP7 and reported separately.

The following workshops have been piloted:

- WS1 "Backup & Recovery, Concurrency Control"
- WS2 "Database Design Methodology"
- WS3 "Data Warehousing and Data Mining"
- WS4 "XML and Databases", Module "XQuery"

All workshops proved to be repeatable by people other than the authors. The lecture material is excellent and the practical part with questions and hands-on lab is well guided with instructions and hints. Students are able to solve the tasks in a consecutive manner obtaining gradually more understanding.

The preparation and setup of the labs exhibited the dependency on certain products or even versions of a product. This is unavoidable since we want the students the possibility to experiment with commercial products.

The main insight from the pilots was that the more technologically dependent a workshop is, the more challenging is the setup of the environment. We propose to separate conceptual from technical issues as much as possible in order to identify the neuralgic spots and to make the workshop as modular and reusable as possible.

WP5 Definitions and proposed tasks

Definition: Pilot run

We defined the term "pilot run" as the execution of the workshop material by instructors others than the authors. (See project meeting in Thessaloniki 29.11.2003)

WP9 defines four workshops which have been implemented during the project phase of DBTech Pro. The piloting comprised these 4 workshops.

Tasks defined in the DBTech Pro WP5

According to the project plan we have defined and executed the following task under WP5, shown in the table 6.

WP-leader	Description	Application Outcomes	Defined outcomes & date of decision
F. Laux Reutlingen	Pilot runs of training in partner organizations in three semesters, follow-up and documentation control	Reports (DE, EN, ES, FI, GR) Feedback report from students and companies (20/country)	Thessaloniki 16.05.2003: ✓ Collection of reports and evaluation forms implemented at each institution ✓ Report on modifications to pilot implementation dictated by the local environment of each partner institution Thessaloniki 29.11.2003: ✓ Piloting mean piloting of the workshop materials and labs Malaga 29.05.2004: ✓ The local lecturer writes a short report on the process and results Helsinki 18.04.2005: ✓ WP5 report will be combined to WP3 report

Table 6. WP5 tasks

The above task have been executed and reported in the reports of WP3, WP5, and WP7.

Pilot runs reports

In the following chapter we describe the details of our experiences of each pilot run. There have been 16 pilot runs on workshop modules (see table 1 below) between May 2004 and April 2005 consisting between 6 and 43 students for a single module.

The students have been requested to answer a list of questions on each workshop they participated in. The evaluation of these questionnaires was done in WP7.

The instructors provided free format comments about their impressions which are recorded in this chapter. The following table 7 collects all the pilots.

Work Package 5 "Pilot runs of DBTech Pro Workshops"				
Workshop	**Institution**	**Date**	**# of Students**	**Comments**
WS1 "Backup & Recovery"	TietoEnator, FI	24 and 25 March 2004	11	
WS1 "Concurrency Control"	TietoEnator, FI	4 and 7 May 2004	6	
WS1 "Backup & Recovery"	Reutlingen University, DE	11 May 2004	43	
WS1 "Concurrency Control"	Reutlingen University, DE	18 May 2004	43	
WS2 "Database Design"	Univ. Paisley, UK	24 April 2003	10	
WS2 "Database Design"	Univ. Paisley, UK	14 Nov 2004	10	
WS2 "Database Design Methodology"	Reutlingen University, DE	11 and 15 June 2004	35	*) ERP
WS2 "Database Design Methodology"	Reutlingen University, DE	14 Dec 2004	30	
WS2 "Database Design Methodology"	Reutlingen University, DE	18 Jan 2005	35	*) ERP
WS2 "Database Design Methodology"	Reutlingen University, DE	6 and 13 April 2005	25	
*) we used different business domain: Enterprise Resource Planning (ERP)				
WS3 "Data Warehousing"	Helia, FI	24 Sep 2004	23	
WS3 "Data Mining"	Helia, FI	25 Sep 2004	23	
WS3 "Data Warehousing"	University of Macedonia (UoM), GR	18 Nov 2004	15	
WS3 "Data Mining"	UoM, GR	25 Nov 2004	15	
WS4 "XML & Databases"	TietoEnator, FI	5 April 2005	11	XPath, XQuery Oracle XML DB
WS4 "XML & Databases"	Reutlingen University, DE	28 April 2005	11	only XQuery
WS4 "XML & Databases"	Univ. Paisley, UK	4 May.2005	10	

Table 7. WP5 Pilot runs of DBTech Pro Workshops

III.2 Pilots of Workshop 1

III.2.1 WS1 at TietoEnator on Module "Backup & Recovery" (24 and 26 March 2004)

Participants

The participants were 11 database specialists from TietoEnator, all quite experienced. Lecturer and lab instructor: Jaakko Rantanen. The experiences were reported by Jaakko Rantanen.

Setting up the technical lab environment

TietoEnator used its own PC labs in the headquarter building. The normal routine was explained to be as follows: the workstation disks do not contain any permanent setup. The whole content is cloned from a disk image that contains the tested environment. This kind of technical service was not available for DBTech pilot.

I proposed to go and check the class environment beforehand. I was not allowed. The only possibility was to go two hours before the lecture and install whatever is necessary on a Windows XP basic installation.

Just to make sure, I prepared an auto run CD. This task took nearly a whole working day. In practice, without the well tested install CD, the installation on 12 workstations would have not been successful. Some of the participants kindly assisted in the installation. Otherwise I had missed the lunch.

The impression after the discussions

Although the participants should be well aware of the subject, they seemed to be well motivated. Obviously the motivation was to have an opportunity to experience and comment new methods of learning. One indicator on the motivation was the fact that all except one of the participants came to run the exercises, after the theoretical lecture day. Beforehand the lecturer believed that many of the participants might not come. The topic must have been familiar enough.

Many of the people were not familiar with Solid DBMS. Because the software is Finnish and has a good reputation, maybe people were partly interested on the tool. Some of the participants finished the exercises quite fast and continued doing some own testing with Solid.

During the course days there was a continuing discussion on the subjects going on. The feedback message was encouraging: the module content is professionally relevant and it certainly should help university students to understand database recovery better that any purely theoretical material (as many the participants remembered the nature of their own university studies).

III.2.2 WS1 at TietoEnator on Module "Concurrency Control" (4 and 7 May 2004)
Participants
The participants were 6 IT professionals from TietoEnator who work with databases frequently.

Setup
The environment was installed following the instructions on the installation guide and using the scripts provided with the CCLAB software package. No modifications were needed on the scripts or the installation procedure. The existing TietoEnator's DBMS's (Oracle, Solid) were used. Unfortunately, there was some data communication configuration problem that prevented the use of SQL Server.

Presentation material
The presentation material was used as it is. The material was covered in the 3,5 h theoretical introduction.

Running the lab
The theory part was covered on 3,5 h lecture and the 3,5 h lab was given on another day. The lab was run with no modifications. The lab went as planned. As instructed on the lab guide and task texts, the participants made notes and discussed in pairs on their observations on the concurrency issues and DBMS' behaviour. The participants gave the following feedback:

- These concurrency issues are very important and interesting. We really need this information when working with databases. I have tried to tackle these issues myself for a long time.
- It is very useful to be able to learn the differences of the main DBMS's.
- This is a very useful workshop. The lab supports the theory very well.

III.2.2 WS1 at Reutlingen University on Module "Backup & Recovery" (11 May 2004)

Setup

During the Setup the needed Software for Module "Backup and Recovery" was not on the T.E.I. Blackboard. The Software "Solid Server and Client" was provided by the author. The client "Flow Control" runs only with JDK 1.3.1_01. We used "TE_InstallSolid.zip (created Monday, 22 March 2004, 11:13:24)" provided by Jaakko Rantanen to install a working environment.

A setup procedure was not given, but the setup was strait forward. As this was Finnish software you need to know at least the Finnish word "asenna". Now the license is expired.

Presentation material

It was comprehensive and a good supplement to the own slides used during the
running course. The students found it clear and useful.

Running the lab

The assignment was step-by-step and therefore easy to follow. Sometimes the
students executed the tasks without reflection on what they were doing. The students appreciated the hands-on lab as experience they usually cannot gain in ordinary courses or labs.

III.2.3 WS1 at Reutlingen University on Module "Concurrency Control" (18 May 2004)

Setup

There was no procedure for setting up the environment. Java knowledge is required. If JRE (or the JDK) is installed you need to edit the run-scripts *.cmd to adjust the PATH and CLASSPATH settings. If you do not have the predefined databases running you need to install it as well. However, there is no instruction on how to do it. An alternative would be to change the client configuration and supply other drivers to make other databases (already installed in your institution) accessible.

Presentation material

The presentation material was excellent; it covered more topics than necessary for
the lab part. It could easily be integrated in a database course.

Running the lab

Running the lab motivates the student because it is only clicking und
discussing the results. However I decided to add a matrix as a deliverable
for the different concurrency situations and asked the students to fill in their observations
and comments.

III.3 Pilots of Workshop 2

III.3.1 WS2 at University of Paisley on Module DB Design
(24 April 2004 and 14 November 2004)

Setup

This WS part required no labs - it was the workshop on database design with role play (so
no particular difficulties there).

Presentation Material

We used our own book and the provided slides.

Running the lab

The lab was executed as a role play; we did not use any modelling tools or database.

III.3.2 WS2 at Reutlingen University on Module DB Design Methodology
(14 December 2004)

Setup

The WS is mostly conceptual. This allows drawing UML-diagrams by hand and
you can implement the schema on any relational database you like.
We have chosen Access to implement the schema.

Presentation Material

We did use the book of Th. Connolly and the provided slides. We found the
material too detailed for our courses of 2 respective 4 h/week.
It is useful as additional material and for self study.

Running the lab

The students found the description of the situation too vague. During the discussion I had to limit and precise the scenario. The fact that the scenario was not fully specified it generated a vivid discussion with the students and resulted in interesting business models and the respective data models.

III.3.3 WS2 at Reutlingen University on Module DB Design Methodology (18 January 2005)

Setup

For the second pilot run I prepared a more limited scenario to fit better in the available time frame. I added descriptions for the most commonly asked question hoping to make the discussion sessions shorter.

Presentation Material: see above.

Running the lab

We streamlined the workshop by modifying the requirements and assignments so that It fitted better in our course frame.
The student's solutions converged better due to the more precise description and more limited scenario.

III.4 Pilots of Workshop 3

III.4.1 WS3 at Helia on Module Data Warehousing (24 September 2004)

Participants

6 international students of Helia, 9 local students of Helia, 7 students of HAMK, 4 teachers from other Finnish polytechnics, 2 industry representatives, 1 teacher at Helia + representatives DBTech Pro partners

Setup

The lab exercises took place at laboratory class of 20 workstations + teacher's workstation connected to a projector. We had installed Microsoft SQL Server 2000 Analysis Services in every workstation

Presentation material

Introduction to OLAP and Data Warehousing by Georgios Evangelidis and Leonidas Karamitopoulos was presented by Georgios Evangelidis in a theory classroom during the morning hours.

Running the lab

Introduction to exercises was presented by Georgios Evangelidis and students run the exercises working 2-3 persons on workstation, due to small number of workstations.

III.4.2 WS3 at Helia on Module Data Mining (25 September 2004)

Participants

6 international students of Helia, 9 local students of Helia, 7 students of HAMK, 4 teachers from other Finnish polytechnics, 2 industry representatives, 1 teacher from Helia, and representatives of DBTech Pro partners

Setup

The lab exercises took place at laboratory class of 20 workstations + teacher's workstation connected to a projector. We had installed IBM's Intelligent Miner for Data version 8.1 in every workstation.

Presentation material

Introduction to Knowledge Discovery from Databases and Data Mining by Dimitris Dervos was presented by the author in a theory classroom during the morning hours. Students got printed handouts.

Running the lab

Introduction to exercises was presented by Dimitris Dervos and students run the exercises working 2-3 persons on workstation, due to small number of workstations.

Summary for both workshop days

For collecting the student reports on the lab works we used the Blackboard learning environment at T.E.I. of Thessaloniki, Greece. The laboratory was available for the Helia students 2 weeks afterwards for finalizing the exercises. The international workshop introducing new technologies for the students and using the latest software technology was found highly motivating by the students and teachers.

III.5 Pilots of Workshop 4

III.5.1 WS4 at Reutlingen University on Module XQuery (28 April 2005)

Setup

This WS module was set up using the XML database Tamino from Software AG. This was done because our department had already installed and used Tamino before. As result, it was not necessary to install a separate tool on each lab PC. The documents to query were provided by Matti Kurki (Helia), the author of the XQuery module.

Presentation Material

We used own presentation material on Tamino and Matti Kurki's presentation on XQuery.

Running the lab

The students found the WS material easy to use. So, only on a few occasions the Tamino reference manuals were needed to solve the XQuery assignments. The provided results enabled the students to check their solutions. Sample solutions were provided for the instructor as a convenience.

III.5.2 WS4 at University of Paisley on Module XML (04 May 2005)

Setup
This WS module was set up using the XML tool from Altova. The documents to query were provided by Matti Kurki (Helia), the author of the XQuery module.

Presentation Material
We used the same material as provided for the workshop.

Running the lab
For the WS4 (XML), the Altova tool was used and the sample questions that were provided. Again, the students had no problems understanding what they had to do.
I didn't use the standard feedback form but I used my University's own feedback form and everything was extremely positive (there were no negative comments provided).

III.6 Results on pilot runs

The student's evaluation forms have been sent to Markku Kuivalahti for pedagogical evaluation. (See WP7 report)

If we intend to execute the pilots at EACH partner location, it must be pointed out that we have to consider that a certain "pilot course" might not be compatible with the local study program or infrastructure. This may prevent a local implementation. So, the recommendation for future projects is to implement smaller units as workshop in the future. Another point of concern is the separation of conceptual and technical matters. This could point out stable portions of a workshop with respect to technological change and separate them from more changing parts.

As we have already prepared formal descriptions of existing and reference courses (see WP4) we suggest to extend these by a comment and code indicating the readiness for pilot use. For the coding scheme the classification developed by Jaakko Rantanen has been extended. For example a regular lecture course with student hand outs in English would be coded like this: [l0D] (lecture, English student's material).
A second example: A guided lab which is now not any more used with setup instructions, teachers material (solutions), and written assignments in

native language will code as: [g4BCD]. The comments could give an indication of how much effort (what action and how many hours work) would be needed to make it ready to use as pilot course.

III.7 Conclusions & Future Challenges

The results from WP5 proved the concept of modular, repeatable workshops to gain a set of reusable learning units. The most benefit can be achieved from thoroughly written instructions which can be used as a "cook book" to set up a lab environment which is known to consume valuable time. Furthermore, the setup has to be repeated when new software releases are available or the hardware is replaced.

> *To make workshops more insensitive for technological change is still a challenge which should be investigated in a continuation project.*

Another approach to this technological challenge is to provide the environment at only one partner's site and make it available for the others. This "service provider" concept was tested successfully with the hosting of databases in the case of WS2 "Database Index Design" and the WS4 "XQuery" part. The written lecture and lab material was hosted mainly on the blackboard e-learning platform made available by the T.E.I.

> *We need more experience to find out if it is feasible from personal and technical resources to act as "service provider" for other partners. This could turn out as a business model on a larger scale.*