

A Problem-based approach to teach physical database design: an experience report

João Marcelo Borovina Josko¹

¹Federal University of ABC (UFABC)
Avenida dos Estados, 5001 – Santo André – SP – Brazil

marcelo.josko@ufabc.edu.br

***Abstract.** This work presents our first experience in applying a Problem-Based Learning approach to improve students' abilities on how to analyze and evaluate database performance problems. The preliminary qualitative results reveal our approach contribution to students learning process. Additionally, we also discuss key lessons learned and indicate some future areas of improvement.*

1. Introduction

The database design process (DDP) comprises the conceptual, logical, and physical steps responsible for transform real-world needs in database structures. Therefore, its knowledge has broadly recognized as crucial to ensuring the proper database support to the organization's operations, even more within contexts of Big Data and multi-model databases.

DDP is an intellectually demanding work because each step requires a different set of skills, analysis abilities, and technical knowledge. Conceptual database design forms abstraction from ill-structured and complex problem domain through consensus among data modelers and users. At the other end, physical database design (PDD) demands deep technical and contextual (e.g., information systems) understandings to decide for the most proper physical structure according to a given performance need [Lightstone et al. 2010]. Such complex nature makes database design learning and teaching a challenge.

This reality has been attracting educational researchers to address these challenges through distinct approaches. Several studies contribute to developing conceptual design knowledge based on learning theories (e.g., cognitive load [Eid 2012]), learning approaches (e.g., collaborative learning), or teaching approaches (e.g., Problem-Based Learning [Marsicano et al. 2016]). Logical design literature also provides several contributions to teaching, including educational software systems [Cvetanovic et al. 2010]. Conversely, physical design teaching has not received similar attention. Indeed, we have found just one study that addresses this subject entirely [Davis 2018].

This work reports our first experience of applying a problem-based learning approach to teach physical database design, contributing to the previous context. Our approach combines visual lectures and cumulative changes-based project to introduce performance analysis knowledge progressively. It differs from the related works [Davis 2018] in teaching method, content coverage, assessment instruments, and practice type.

This work has the following structure: We present our theoretical foundation in Section 2 and discuss our approach characteristics and dynamics in Section 3. In Section 4, we report encouraging preliminary results. We then summarize some challenges faced throughout the intervention in Section 5. Finally, we conclude in Section 6.

2. Background and Motivation

Current database products offer a powerful toolkit of physical structures (e.g., indexes, horizontal partitioning) for increasing their overall performance. However, these features are deeply connected and strongly affected by infrastructure characteristics (e.g., disk transfer rate). Considering all such information to choose the most appropriate solution according to business needs requires to know analysis strategies [Lightstone et al. 2010].

Long-term and authentic contexts projects can provide students the adequate conditions to develop this critical thinking ability. Problem-Based Learning (PBL) is a student-centered teaching method that intends this very same purpose. It uses the project as an instrument to promote deeper learning and to develop soft skills such as collaboration and communication [Marsicano et al. 2016]. In this setting, students are actively involved in their learning process, while teachers act as coaches.

3. Problem-Based Approach

We considered previous pedagogical awareness to create our teaching approach. It intends to develop students' analysis abilities gradually. Following, we present the project characteristics (Section 3.1) and discuss the organization of the classroom (Sections 3.2).

3.1. The Project Structure

We designed our project to simulate the cumulative effects of business and physical changes over a real database to offer an authentic context of analysis (Figure 1). Initially, we provide a script that creates and populates a baseline database model in PostgreSQL 12. Furthermore, we supply the corresponding model metadata and a set of isolated queries that represent the current operations over the baseline database. To mimic a real-world scenario, our baseline model has some data modeling issues (e.g., absence of foreign keys, wrong data type) and ineffectual resources (e.g., composite B-Tree indexes) to solve some of the isolated queries. The baseline model has six relations of moderate size (100K tuples and six columns on average) and their corresponding relationships that represent a subset of a problem domain related to project-oriented organizations.

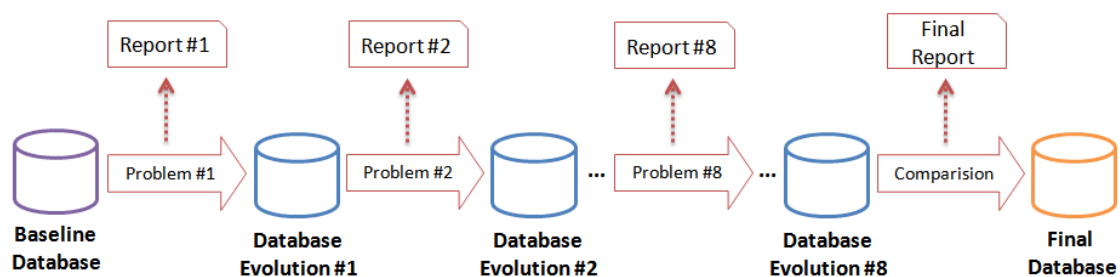


Figure 1. The cumulative steps of the PDD project (Source: The author)

From this baseline, we expose students to a particular business problem-situation every week (Section 3.2). It describes (in business terms) a hypothetical performance need that may involve some of the given queries or be related to new demands or both. These new demands may refer to a set of additional isolated SQL queries, a SQL transaction, or a database growth (e.g., new relations, relationships or attributes).

For each problem-situation, students must consider physical structure characteristics and their current database state until determining the most appropriate physical design solution. Subsequently, students apply their solutions and record all analysis steps and corresponding information in a report. Figure 1 illustrates this sequence of problem resolutions until the final comparative analysis. In this final analysis, students produce a benchmark report by contrasting the performance of their last database model with the baseline one based on a set of predefined queries and transactions.

It is worth mentioning that it is possible to use any relational database product to support the previous PBL idea. However, it involves some changes in the lectures and the SQL artifacts (e.g., baseline model) according to the desired database specificity.

3.2. The organization of the classroom

We interspersed the lectures with the practice sessions to better combine the theory and practice of physical database design in a 12-week course. All teaching and support activities involve just one teacher.

We use the 1st week of the course to establish PDD foundations. The lectures from the 2nd to the 10th week of the course comprise the following sequence of subjects: query processing, indexing, relation and index storage, vertical and horizontal partitioning, materialized views, denormalization, and query optimizations. We adopt rich visual-textual material to introduce and illustrate all the subjects. Further, we use a short analysis to end the study of each of them. The purpose of this analysis is twofold: *i.*) reinforce the subject characteristics in focus, and *ii.*) describe some analysis strategies. In the 11th and 12th weeks of the course, students attend the individual exams in the form of an essay.

Following each lecture, we introduce a new problem-situation (Section 3.1) every lab practice session. This session represents the occasion that pairs (fixed in the 1st week) collaborate to understand the problem at hand, discuss lines of work, perform the first analysis, and interact with the teacher. Pairs have one week to solve each problem-situation and submit their report. It is worth noting that lab and teacher support (outside classroom) are available in a comfortable timetable every weekday.

4. Preliminary Results

We applied the PBL approach in one term of a non-mandatory course. From 30 enrolled students, about 67% of them (20/30) completed the course, and near to 23% of students (7/30) did not appear. The last 10% of students canceled their enrollment. The students' average grade was 6.9, while the pass rate was 95%.

We used a non-mandatory post-course internet survey to gather student perceptions about our pedagogical intervention. This survey asked for information related to gender (closed-ended question), PDD confidence (Likert scale question), the PBL experience (Likert scale questions), and learning reflections (open-ended question). About 60% (12/20) of students (eleven males and one female) filled out the survey correctly, which has an internal consistency of 0.72 (Cronbach's Alpha Test).

4.1. The Students' Comments

The students' comments resulted from a two-cycle analysis process of the 12 answers to the open-ended question. In the first cycle, we applied a descriptive coding method. Subsequently, we highlighted sentences or paragraphs according to the code list.

Half of the students (50% or 6/12) report the overcoming and achievement feeling with their physical design analysis learning. Furthermore, a notable portion of students is pleased ($\approx 75\%$ or 9/12) with the learning experience as well as knowledge acquired, as mentioned by student *S1*:

“Before the course, I only knew how to make selections using SQL. At the end of the course, I have got good ground knowledge to work with databases. It is a differential both in academia and in the market”

In contrast, some learners ($\approx 41\%$ or 5/12) declare difficulties in their learning process. Mostly, they report time restrictions to perform the weekly reports properly. Lastly, participants ($\approx 16\%$ or 2/12) suggest the increase of lab practice time and thank the teacher for the experience and patience (50% or 6/12).

4.2. The Students’ Perceptions

Table 1 lists the questions about our pedagogical intervention that students were requested to reflect according to their individual learning experience. As observed in Figure 2, their answers suggest a very positive perception of our approach.

Table 1. Likert post-survey questions

Question
Q1. Was the plan and material of the lectures adequate to support your learning process?
Q2. Were the project characteristics relevant to help your learning process?
Q3. Were the reports and essays adequate assessment instruments?
Q4. Did the teacher’s feedback and scaffolding support your learning process?
Q5. Do you feel more confident in performing the physical design analysis at your work?

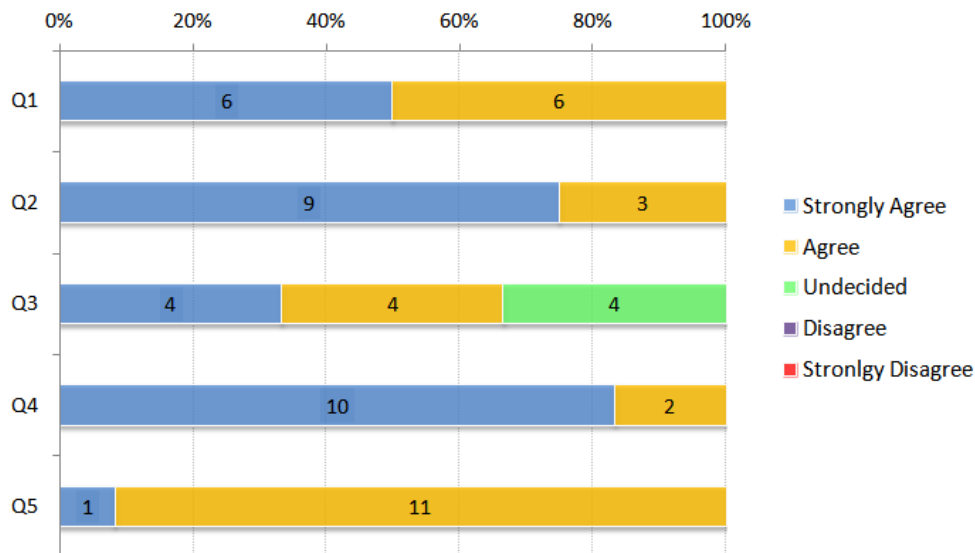


Figure 2. Students’ reflections per Likert question (Source: The author)

Numerous students (75% or 9/12) strongly agree that the project arrangement is valuable for their learning process (question *Q2*), despite the time restrictions previously mentioned (Section 4.1). Moreover, they also express the teacher-coach support as equally

relevant (question Q4). Interesting to remark that all learners report confidence in their PDD knowledge and feel comfortable applying it in real-world contexts (question Q5).

Students are less emphatic about the lectures' format and the assessment instruments (questions Q1 and Q3, respectively). Anecdotal feedback exposes that specific subjects (e.g., query optimization) require more detailed and visual materials to improve their understanding. Lastly, some learners ($\simeq 33\%$ or 4/12) are in doubt as to the adequacy of the report-essay as an assessment instrument (question Q3). Such perception is likely to be related to the unusual assessment format, as suggests some student's comments (Section 4.1).

5. Challenges and Lessons Learned

In this section, we present our reflection on two main issues that we experienced during the pedagogical intervention based on the PBL approach. *Conceiving problems for the project* has proved to be the hardest part of it. We had to consider numerous information (e.g., time constraints, physical structures, and data model characteristics) to create compelling and feasible business problem-situations. Regrettably, some of them did not provide a valuable scenario of analysis.

An additional challenge refers to *the reports' evaluation process*. Students' reports contained diverse mistakes of different criticality levels, including lack of accuracy, calculation errors, and naive analyzes. However, our primary criteria were too high-level to permit a proper effort-based evaluation of such heterogeneous reports.

6. Conclusion

In this paper, we discussed the characteristics of a PBL approach to teaching physical database design. It introduces an authentic database context to promote learning and place students at the center of their learning process. Preliminary results and anecdotal feedback show promising benefits of our educational intervention in students' PDD capability. In future works, we expect to refine our PBL approach to mitigate lessons learned and perform a longitudinal investigation to examine its effect on students' learning. Moreover, we plan to include some multi-model performance subjects and produce animations to make technical discussions easier (e.g., query processing).

References

- Cvetanovic, M., Radivojevic, Z., Blagojevic, V., and Bojovic, M. (2010). Advice—educational system for teaching database courses. *IEEE Trans. Educ.*, 54(3):398–409.
- Davis, K. C. (2018). Teaching physical database design. In *International Conference on Conceptual Modeling*, pages 165–175. Springer.
- Eid, M. (2012). A learning system for entity relationship modeling. In *PACIS*, page 152.
- Lightstone, S. S., Teorey, T. J., and Nadeau, T. (2010). *Physical Database Design: the database professional's guide to exploiting indexes, views, storage, and more*. Morgan Kaufmann.
- Marsicano, G., Mendes, F. F., Fernandes, M. V., and de Freitas, S. A. A. (2016). An integrated approach to the requirements engineering and process modelling teaching. In *2016 IEEE 29th CSEET*, pages 166–174. IEEE.