Report - ID2221 - 2021-10-18

Please note that there is only one respondent to this form: the person that performs the course analysis.

Course analysis carried out by (name, e-mail):

Amir H. Payberah, payberah@kth.se

DESCRIPTION OF THE COURSE EVALUATION PROCESS
Describe the course evaluation process. Describe how all students have been given the possibility to give their opinions on the course. Describe how aspects regarding gender, and disabled students are investigated.

We had two evaluations in the course: one mid-term evaluation after the third week of the course and one at the end. Moreover, all the students could directly comment on the course, either in person or through Canvas.

DESCRIPTION OF MEETINGS WITH STUDENTS
Describe which meetings that has been arranged with students during the course and after its completion. (The outcomes of these meetings should be reported under 7, below.)

In the course, we had two oral presentations for the lab assignments and the final project. In those presentations, we met each group of students (two students in each group) individually.

COURSE DESIGN
Briefly describe the course design (learning activities, examinations) and any changes that have been implemented since the last course offering.

This course covers various advanced topics in data-intensive computing platforms to store and process big data. The main objective of this course is to provide students with a solid foundation for understanding large-scale distributed systems used for storing and processing massive data. After the course, the student should be able to:
1. ILO1: explain different fundamental concepts of data-intensive computing platforms such as the shared-nothing architecture and dataflow programming model and explain how massive data processing platforms work.
2. ILO2: store and retrieve data in distributed stores, either on distributed file systems or NoSQL databases, and implement different queries over them.
3. ILO3: process different types of data, including structured, streaming, and graph using massive processing platforms, such as Spark and Flink, and apply machine learning algorithms on massive data, using applications such as Mllib.
4. ILO4: build advanced applications using data-intensive platforms, make scalable applications on a cluster, and process massive data.

The course consists of several tasks, in which each one assesses different ILOs.
1. Task1 (two reading assignments): in this task, students should choose three papers from the given pool of papers (or their proposed papers, which the teacher should confirm) and review them, and write a report for each one. (P, F)
2. Task2 (six review questions): through these questions, we will ask related questions about each lecture to motivate students to study the lecture notes and papers. (P, F)
3. Task3 (two lab assignments): the lab assignments are designed to direct students to different types of challenges with storing and processing platforms. Each lab has two levels of difficulty with basic and complex challenges. (A-B for the advanced part, and C-D for the basic part, F)
4. Task4 (final project): the purpose of this task is to implement an advanced application on processing massive data (which should be proposed by students and confirmed by the teacher) on a cluster. (A-B for an advanced project, and C-D for a simpler project, F)
5. Task5 (final exam): the final exam consists of a number of questions from different parts of the course that assesses students' theoretical knowledge about covered platforms in the course. (C-E, F)
To pass the course, students should complete all tasks, and the final grade is the average of grades in Task3, Task4, and Task5.

THE STUDENTS' WORKLOAD
Does the students' workload correspond to the expected level (40 hours/1.5 credits)? If these is a significant deviation from the expected, what can be the reason?

There is no significant deviation from the expected level of workload. However, for the next round of the course, we may change some of the tasks, for example, by skipping the reading papers assignment.

THE STUDENTS' RESULTS
How well have the students succeeded on the course? If there are significant differences compared to previous course offerings, what can be the reason?

Less than 10% of the students couldn't pass the course, and more than 80% got grades C or higher.

STUDENTS´ANSWERS TO OPEN QUESTIONS
What does students say in response to the open questions?

Below, I summarize some of the answers students gave to the following questions:

* What was the best aspect of the course?
The students were happy about exploring multiple new technologies in the course as they are important topics to find good positions as data engineers in the industry. Moreover, they liked the course structure, the teaching style, and the lab assignments. Furthermore, the continuous homework (review questions) helped them refresh and solidify what was taught in class.

* What would you suggest to improve?
There are two main issues that some of the students mentioned: (1) the course workload is high, considering different home assignments, and (2) the course covered many frameworks, and some of the students suggested covering fewer frameworks but present them in more details.

SUMMARY OF STUDENTS' OPINIONS
Summarize the outcome of the questionnaire, as well as opinions emerging at meetings with students.

As mentioned in the box above, the students were happy about the course structure, topics, and assignments. However, they would prefer to have less workload and focus on a fewer number of frameworks.

OVERALL IMPRESSION
Summarize the teachers' overall impressions of the course offering in relation to students' results and their evaluation of the course, as well as in relation to the changes implemented since last course offering.

Overall the students expressed positive evaluation of the course, with all LEQ statements scoring more than 5. There was no significant difference in evaluation from different groups of students, except the students from "international exchange", which gave lower scores. This might have been since they did not have sufficient background to take the course.

ANALYSIS
Is it possible to identify stronger and weaker areas in the learning environment based on the information you have gathered during the evaluation and analysis process? What can the reason for these be? Are there significant difference in experience between:
- students identifying as female and male?
- international and national students?
- students with or without disabilities?

The highest scores were on questions 1 (6.1) "I worked with interesting issues", 6 (6.1) "The atmosphere on the course was open and inclusive", 22 (6.0) "I was able to get support if I needed it", and the lowest scores were on the questions 5 (5.0) "I felt togetherness with others on the course", 14 (5.2) "I received regular feedback that helped me to see my progress", and 15 (5.1) "I could practice and receive feedback without being graded" These are mainly due to the fact that the course had a relatively large number of students expected for a master course (127 students) and the team had very limited resources for giving feedback to the students.

PRIORITIZED COURSE DEVELOPMENT
What aspects of the course should be developed primaily? How can these aspects be developed in short and long term?

We are restructuring some of the lectures by focusing on fewer platforms, reducing the course assignments by skipping the reading assignment task, and focusing mainly on the review questions and labs.