Research Project

Group Work: 20%   **26.10.2020**

1 Introduction

This course covers a number of post-relational data models and systems. Some systems are designed to support one logic data model while others may extend their domain to include other data models. The emergence of various logic data models highlight the reality that a logic data model might be supported by various systems with very different underlying architecture. In this project, you are asked to focus on a particular data model by investigating how it is supported in two database systems. You need to produce a report based on the investigation and give a presentation to summarize your findings.

2 Project Scope

You should start the project by identifying a data model you want to work on. The data model could be relational, document, graph, key-value, time-series and spatial model. If you want to work on another model, you need to confirm with your tutor.

Next, choose two systems that support the data model. One of the systems should be a NoSQL system. The system could use the data model as its primary model or as secondary model.

The main project work involves in-depth comparison of three features of the two systems with respect to that particular model. For each feature, you are expected to present a couple of findings and enough evidence to support your findings. The evidence could come from reviewing credible literature or from your own experiments. You can compare features related with query language, examples include query expressiveness, support for certain types of query and so on. You can also compare features on query execution, examples include execution performance, execution profiling and so on. In additional, you may also consider features related with database architecture that may affect scalability, replication, consistency and others

3   Group Collaboration

This is a group assignment. Each group has three members. Each member is expected to focus on a particular feature and there will be individual marking component.

The groups are expected to decide on a data model with two supporting systems, features to compare and workload distribution in the first week. These need to be confirmed with the tutor during week 10.

4   Deliverable

There are two deliverable: **project report** and 10 minutes **Project Presentation**. The report is due on **Friday 20th of November 23:59 (Week 12)**. It should be submitted as a pdf file in Canvas. The presentation will be scheduled on week 12, before the report submission.

5   Report Structure

The report should contain six sections: **introduction**, one section for each feature, **conclusion** and **appendix**. The **introduction** section should briefly describe the data model, the two systems you investigated and the features you have explored. The **conclusion** section should summarize the findings and also indicate any limitation on your findings. The **appendix** section should briefly describe the contribution of each group member.

The section for each feature should start by describing the feature you are investigating. It should be followed by each finding and evidence to support the finding. You can include sample data and queries, performance statistics and other data in various formats in this section.

There is no required minimum or maximum length for the report. It depends on the features your want to investigate and the evidences you use to support the findings. You should give enough details to show the readers that your findings are not biased and the evidences are produced following scientific principles. On the other hand, you should not copy and paste large amount of text from the documentation of any system.

6   Marking Guide

- Report: 15 points

    - Individual features (4 points each): The points are awarded based on the quality of the feature comparison section.

        * 4 points: a professional written section with unbiased findings supported by evidences produced by independent investigation and/or experiments. The

experiment may involve hypothetical data set, query workload and cluster set up. The findings should be described with proper context without over generalization. The evidence presented, for instance queries written for different systems, or data collected, for instance execution statistics, should not be affected by the experiment environment or student's own familiarity with individual system. References are properly cited.

  * 3.5 points: a well written section with all findings properly stated and supported by evidences through independent investigation; minor issue related with presentation.
  * 3 points: a well written section with properly stated findings but most of the evidences are from literature of credible sources instead of own investigation.
  * 2 points: the section contains some findings and evidences; the findings may not be put in proper context; there is not enough evidence to support the findings;
  * $1 - 0$ point: the section contains one of more of the following major flaw: findings are biased, or irrelevant; the evidences do not support the findings, or are not from reliable resources; or are not produced in unbiased setting; the writing is incoherent and hard to follow;

- Overall quality (3 points)

  * 3 points: a professional written report with all required sections.
  * 2 point: report with minor language issue.
  * $1 - 0$ point: incomplete report that missed one or more sections.

• Presentation: 5 points

  - 5 points: professional presentation with good structure; presentation covers all findings; Slides well prepared with consistent style; slide not too crowded; presentation finish on time; questions answered well;
  - 4 points: well delivered presentation with minor issues on related with slide design or delivery.
  - 3 points: Presentation include all important content but are disorganized or hard to follow
  - 2 point: presentation misses some important content. For instance, no clear statement of findings or no evidence.
  - $1 - 0$ point: Presentation contain only limited information or irrelevant information.