CSC443 Winter 2018

Project

Due: Sunday Mar 25, 2018 at 11:59 PM

Objective:

In this project, we evaluate a few of the most popular open-source databases in the market in order to understand their architectures from system point of view.

Logistics:

This is a team project. Teams can have maximum of 3 members. Each team will select 3 preferences from the list of the databases given later in this document. One database will be assigned to each team. We will do our best to allocate one of the preferences to each team but it is not guaranteed.

One of the members of each team requires to email the instructor the name of the students in the team and the list of preferred databases by Feb 26, 2018.

A few teams will be selected, based on the quality of their project report, to present in the class in the last week of the semester (Apr 02, 04, & 05). The selected teams will be given a maximum of 20% bonus (based on the quality of their presentation). If a team prefers not to present, we will give the chance to the second-best team for the same database.

You will be graded based on completeness and correctness of your report.

Databases:

We will evaluate the current version of the following databases:

1. MySQL Community Edition (https://www.mysql.com/)
2. PostgreSQL (https://www.postgresql.org/)
3. SQLlite (https://www.sqlite.org)
4. MongoDB Community Server (https://www.mongodb.com/
5. CouchDB (http://couchdb.apache.org/)
6. Cassandra (http://cassandra.apache.org/)
7. Elasticsearch (https://www.elastic.co/products/elasticsearch)
8. HBase (https://hbase.apache.org/)
9. Parquet (https://parquet.apache.org/)
10. Hive (https://hive.apache.org/)
11. Impala (https://impala.apache.org/)
12. Neo4J Community Edition (https://www.neo4j.com/)
13. VoltDB Community Edition (https://www.voltdb.com/)
14. Redis (https://redis.io/)

Task:

You need to do research about the assigned database system through reading documents, tutorials and perhaps looking at their source code. You also need to install the database system and evaluate it. At the end, you will write a project report. Your report, at least, needs to contain the following sections:

1. Introduction: name, license, a bit of history, how popular it is, key differences with other databases in the market, strong points, supported platforms and so on.

2. Potential applications: what real applications are suitable for this database. Please provide some examples.

3. System architecture: detailed architecture of the database system including all internal and external modules (if any). Please try to represent the architecture in diagrams with description for each part.

4. Query language/interface: you require to explain what query languages (e.g. ANSI SQL) the system supports and what interfaces exist for the system (e.g. REST API). Some of these systems have their own console as well as APIs.

5. Storage details: in this section, you will explain the details of how data and indexes are stored on disk and/or in memory. If applicable, you need to explain the file structure and format in details. You may have to refer to the source code too. In addition, you need to explain any configurable parameter that will affect storage (e.g. page size)

6. Evaluation: you need to install the database system and try them out by creating test databases and query from them. You can use any programming or scripting language you prefer. You need to explain, in details, what you have done and demonstrate the results. Specifically, we are interested to learn how queries optimized by taking advantage of indexes (if applicable). You need to include enough data in your evaluation to observe the effects.

Deliverables:

There is only one deliverable and that is the report in PDF format that includes all diagrams and charts too. The report should not be more than 20 pages.

There is no need to submit any code for this project. You only explain, in the report, what your scripts have done and show the results.

For those teams who selected to present, they need to create a short presentation document. Every presentation will be maximum 15 minutes.