Computer Science 4411b

Winter 2017

Project Description

Proposal due Jan. 25, 2017

Report due April 5, 2017

The purpose of the project is to “discover” something about a database-related topic that is not covered in class. If you choose to cover one of the new database technologies, you can, for example, demonstrate its use and explain the process you went through, or discuss or compare it to more traditional approaches. In demonstrating the use of a system, you should use some “local” data so that we can be sure you did the work yourself.

Your project can be related to new “database” technologies such as:

* key-value stores:
  + - Dynamo (Amazon)
    - Voldemort (LinkedIn)
    - Berkeley DB
* column-oriented databases (wide-column stores):
  + - Google's big table
    - Cassandra
    - SimpleDB
    - Hbase
* document-based data stores:
  + - CouchDB
    - MongoDB
* graph databases:
  + - Neo4j
    - GraphDB
* map-reduce
* hadoop

You should look for information on the following aspects:

* software/data architecture
* data structures/model
* querying capabilities
* how storage is organized
* query optimization
* concurrency control
* recovery
* access control

Other ideas are can be found in the project description for CS9538b.

What is required:

* a proposal, which is a 1 page description of what you plan to do, with some citations to sources you plan to make use of, due Jan. 25.
  + Sources can include books or articles (from OWL site for example).
  + Manual references are also okay to use.
  + Careful about using opinionated/blog sources.
* a presentation on your topic – the time limit will be a maximum of 10 minutes, hopefully less.
  + If you are doing a project on Neo4j, for example, your presentation can be on only one aspect of it, say the data model, but the whole project can cover all aspects, perhaps with an attempt to demonstrate its use.
* a written report, of 10 - 12 pages, with subheadings and proper references to the literature, handed in on-line, due on April 5.
  + This report must be in your own words. These reports will be checked for plagiarism by Turnitin.

End of Set 5:

Check list of things to look for in query language. Good for MongoDB