CSE 6331 Cloud Computing (1)
Spring 2018, © DL, UTA, 2018

Programming Assignment 3
Data Repository Performance
Due: In Blackboard

Cloud Assignment 3 – Amazon Web Services (AWS)

Description:
Various mechanisms for storing data (files, RDBs-SQL, "no-SQL", in-memory, hybrid schemes) are critical to selecting data repositories and guarantees and properties including functional, structure and performance.

Your assignment is to measure performance on SQL tables: creating, querying, modifying data (tuples).

Starting with (fairly) large, well structured data at:
https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php
(all earthquakes for the last 30 days)
If earthquakes make you nervous, equally large (or larger) data exists on:
https://www2.census.gov/programs-surveys/popest/datasets/
descriptions on: https://www2.census.gov/
Or weather data at:
https://www.ncdc.noaa.gov/data-access/quick-links#loc-clim

Create a SQL table, calculate time to create the table (and indexes).
Allow a user to specify on a web interface:
1. A number of random queries (up to 1000 queries of random tuples in the dataset)
2. A restricted set of queries, similar to previous (1.) but where selection is restricted (ie only occurring in CA, or within N<100 km of a specified lat,long location.
   Or: a time range, or a magnitude range.
3. Measure time expended to perform these queries.
4. Show results.

Users of this service will interact with your performance service through web page interfaces, all processing and web service hosting is (of course) cloud based.

Please, submit through Blackboard.
All work must be your own, or from a group.

You must submit this lab, working (or partially) by the due date. Your submission subject should clearly state your name and the lab number.
You may (optionally) demonstrate this lab, working (or partially) to the GTA before the due date.
Your program should be well commented and documented, make sure the first few lines of your program contain your name, this course number, and the lab number.
Your comments should reflect your design and issues in your implementation.
Your design and implementation should address error conditions.