INF-3200

Assignment 2

Dato

**To do:**

* Implement a key-value store in the form of a distributed hash table
* Investigate the characteristics of my design
* The implementation must support PUT and GET operations. The system should maintain the key&value of each insert so that you can get the correct value on GET operations
* Data that is put into the system should be placed into its corresponding node using consistent hashing
* The implementation should should include finger table implementation so that the complexity of look-ups are better than a linear search
* The code does not need to support dynamic leaving and joining of nodes. It can be launched with a fixed set of nodes
* Fully connected networks are not allowed. Nodes should only be aware of other nodes if they are their neighbours or entries in their finger table. The only exception to this is immediately following the initialization of the system
* The service should run at least 16 nodes
* Any node in the system should be able to serve incoming GET requests from a client. Meaning you need to forward PUT and GET requests to the correct node that is responsible for storing the data according to the hashing algorithm.
* The implementation does NOT have to handling multiple requests in parallel, but you can implement this if you want.
* Store the key & value data in memory only, and do not persist? It on disk
* The nodes must only communicate through network protocols.
* The code must run on the cluster
* The code must be run via the “run.sh” script
* Test the implementation where you measure the throughput of the system in PUTs and GETs per second. The experiment must be performed on these network sizes: 1,2,4,8 and 16 nodes. Do at least 3 runs of each trial and plot the results in a graph WITH errors bars showing the standard deviation.
* Implement the client test to test the correctness of the Chord protocol.

**What is:**

* A distributed hash table
  + Is a decentralized storage system that stores data in nodes based on the hash value from the data.
* Associative array
  + A data structure that stores key-value pairs
* Deterministic hash function
  + A hash function that ALWAYS returns the same output for the same key.
* An overlay network
  + A smaller network where nodes only know about each other
* Identifier circle
  + Is a circular number line which decides which nodes should store specific data
* Chord protocol
  + Is a way for a bunch of computers to work together as if they were one big dictionary
* Finger table
  + A simple way for nodes to find other nodes in the circle
* Measure throughput of the system?
  + How many operations(PUT/GET) per second the system can handle
  + (See how the performance is when you add more nodes)
  + For example, make 100 PUT and 100 GET requests, measure the time and see how long it took for each operation
  + Throughput = number of operations/time taken
  + Run with different number of nodes at least 3 times for each setup to test error ranges
  + plot the results
* When I run the tests, should I have the average in the middle, and then plot the range of all tests as error bars in the plot?
  + yes

**How to do it**

**Report**

* The report should be approximately 1400 words long.
* The report should explain the protocol that has been implemented, its advantages and disadvantages, and explain how it was implemented.
* If something in the implementation is not working as expected or intended, it should be explained in the report.
* The report should be handed in as a PDF (and preferably typeset with LaTeX). See "Writing your report" below.
* The report should contain at least one experiment, measuring the through put of your system. See the experiment section below.
* The results of your experiment should be presented with a plot in vector graphics (.pdf or .svg being the most common formats), with error bars showing standard deviation.
* The report should provide citations for work that you reference. For this assignment, it would for example be natural to cite the Chord paper [1], in addition to other sources that you might find useful. References should be in the IEEE citation style.
* Your report should be structured like a scientific paper. Use the Chord paper as an example.
* The report should preferably be typeset in LATEX, but this is not a requirement. To get started with LaTeX, if you do not have a local installation, you can try Overleaf1, which provides an online environment.