**Base Experiment (MQ)**

This experiment is an extension of the [Base experiment](https://github.com/CS-6381/FinalProject/blob/main/DesignOfExperiments/Base.md) that is specific to DS technologies.

**Qualitative Data**

TODO: Change this These metrics are inspired by [Yahoo! Cloud Serving Benchmark (YCSB)](https://github.com/brianfrankcooper/YCSB/wiki). Fill in with information you discover about your chosen DS.

| **Metric** | **Value** |
| --- | --- |
| Centricity (data-centric or message-centric) | <https://cassandra.apache.org/>: Data-centric (stores data) |
| Connection (machine-to-machine or point-to-point) | <https://cassandra.apache.org/>: Point-to-point. Data replicated to nodes.  <https://www.geeksforgeeks.org/application-connectivity-with-cassandra/>: Database can be used on the same computer. |
| Underlying Architecture (decentralized or hub-and-spoke) | <https://cassandra.apache.org/>: Decentralized, no single place of failure |
| Protocol | <https://stackoverflow.com/questions/28127182/data-transmission-between-nodes-and-client-cassandra>: Gossip  <https://en.wikipedia.org/wiki/Gossip_protocol>: Gossip allows data to spread to all group members. |
| Transport(s) | <https://cassandra.apache.org/doc/latest/faq/>: Ports are TCP. |
| Data Serialization | <http://www.diva-portal.org/smash/get/diva2:839521/FULLTEXT02>: SQL is limited by table format and the need to change table rules when changing data structure. Data serialization alleviates this issue by treating the database as a file storage system that stores data in binary form in columns with unique ID’s. This allows queries to read fewer columns, thus dropping overhead. Apache Avro, a serialization library, allows the user to define JSON-based data structures in serialization. |
| Supports Queuing | <https://medium.com/@anirudhkanabar/clustering-order-in-cassandra-how-to-achieve-ordering-of-data-in-cassandra-f52e8a73d5d5>: Clustering order feature allows queries to be sorted according to their given parameter |
| Data Type Representation | <https://cassandra.apache.org/doc/latest/cql/types.html>: CQL (Cassandra Query Language) supports many native types (e. g. BLOB, BOOLEAN, DATE, FLOAT, INT, TEXT, TIMESTAMP, UUID, VARCHAR, etc.). |
| QoS Parameters | <https://docs.datastax.com/en/dse/5.1/cql/cql/cql_using/useQueryColumnsSort.html>: You can set the LIMIT or PARTITION LIMIT parameter to control how much data you get back with a SELECT query. |
| Supports Dynamic Discovery | <https://stackoverflow.com/questions/46643437/cassandra-read-performance>: Cassandra’s writing performance is greater than its reading performance. |
| Communication Patterns | <https://cassandra.apache.org/doc/latest/faq/>: Cassandra uses 7000 for cluster communication, 9042 for native protocol clients, and 7199 for JMX. Configure native procol and internode communication ports in Cassandra Configure File (<https://cassandra.apache.org/doc/latest/configuration/cassandra_config_file.html#cassandra-yaml>) and JMX port in cassandra-env.sh. Ports are TCP. |
| Abstraction Layer | <https://www.geeksforgeeks.org/features-of-cassandra/>: Cassandra uses CQL, which provides an abstraction layer to hide structure implementation.  <https://www.datastax.com/resources/video/datastax-accelerate-2019-casquatch-open-source-java-abstraction-layer-cassandra>: Casquatch is a database abstraction layer with code for streamlining Cassandra works that includes load balancing, connection pooling, and geo-redundancy and uses the concept of POJO instead of CQL. |
| Up-front Complexity | <https://cassandra.apache.org/>: Decentralized. Data replicated to nodes.  <https://www.quora.com/Is-Cassandra-hard-to-learn>: Setup not too difficult. The focus is on your project database design.  <https://en.wikipedia.org/wiki/Apache_Cassandra>: Free, open source  <https://www.youtube.com/watch?v=z0fPRyR12KA>: Setup can be not too difficult (e. g. on Ubuntu). |
| Large Implementations | <https://www.quora.com/How-much-data-can-a-single-node-in-a-Cassandra-cluster-handle-Are-there-configuration-settings-to-set-the-data-size-limit-the-data-size-on-the-node/https://stackoverflow.com/questions/4775388/how-much-data-per-node-in-cassandra-cluster>: Under 1 TB per node  <https://stackoverflow.com/questions/7190573/cassandra-node-limitations>: A row must fit 1 node.  <https://docs.datastax.com/en/cql-oss/3.x/cql/cql_reference/refLimits.html#:~:text=Table%20%2F%20CF%20name%20length%3A%2048,65535%20(2%2016%2D1)>: CQL has some limits: 2 GB max per column or blob, 48 characters max per table name, 65535 parameters max per query, etc. |