**Assignment 8**

Wanyi Su

Q1.

1. What happened when you inserted another row with the same primary key as an existing row?

The value of existing row with that primary key would be replaced and overwrote with the value of this newly inserted row.

Q2.

1. What happened when you query a keyspace with replication factor 1 and one node down? How did it behave with replication factor 2 and a node down?

When querying a keyspace with replication factor 1 and one node down, for some query (e.g.: record with id=2 and “SELECT \* FROM test”), we may get error message showing “NoHostAvailable: (‘Unable to complete the operation against any hosts)” and can’t get any valid result. For some query (e.g.: record with id=4), we can get valid result.

When with replication factor of 2 and only a node down, we can still valid results by querying a keyspace.

Q3.

1. How did the consistency level affect the results with the node up/down?

Suppose the replication factor is 2 and the number of nodes is 3:

When all nodes are up, we can get valid results with node up or down.

When consistency level is “one” and a node is down, we can get valid results with node up or down.

When consistency level is “all”, we may or may not get valid results when a node is down. This depends on if any copy of the record is stored on the dead node. If so, we can’t get valid result and would get error. If not (which means all 2 copies are stored on live nodes), we can get valid results even if a node is down.

When replication factor is 1:

If the consistency level is “one” and a node is down, we still have 2 nodes alive. So we may get valid results if the copy of record is on the live node. If the copy is on the dead node, we can’t get valid result and can get error. When all nodes are up, we can get valid result.

If the consistency level is “all” and a node is down, we may or may not get valid results, similarly to the case of “one” with replication factor of 1.

If all nodes are up, we can get valid results.

Q4.

1. Which of the WHERE id=? values returned successfully with CONSISTENCY ALL when one of the nodes was down? Why do you think some could be returned but not others?

With CONSISTENCY ALL and one node down:

We can’t insert record with id=10, and error is “NoHostAvailable: ('Unable to complete the operation against any hosts'”.

During the inserted records, WHERE id = 1 or 4 or 5 or 7 values return successfully when one of the nodes was down.

(WHERE id = 2, 3, 6 failed)

WHY:

We have 3 nodes in total. The values with id=? that are returned successfully are store their all two replication copies in the live nodes. Those not returned successfully store one of their copies on the node which is down. Therefore, when CONSISTENCY ALL applies, those storing one of their value copies on the node which is down can’t return successfully.

Q5.

1. What was the CREATE TABLE statement you used for the nasalogs table? What was the primary key you choose, and why?

CREATE TABLE nasalogs (

host TEXT,

id UUID,

bytes INT,

date\_time TIMESTAMP,

r\_path TEXT,

PRIMARY KEY (host, id)

);

We choose “id” and “host” as the primary keys.

WHY:

In this assignment 7, we do some aggregation calculation based on different host names. Therefore, we need to partition records using “host”. Therefore, we set “host” as the partition key to indicate any records with same host name stored on the same nodes. However, there are many records within the same host names, we need to set unique “id”s to tell them apart. Hence, we get primary key (host, id),

Q6.

SELECT SUM(bytes) FROM nasalogs;