**Lab 8: Hive**

1. Work the “max-temperature” example in Hive. Print out a screenshot for the result of the example in CSUEB Hadoop.

DROP TABLE IF EXISTS records22;

CREATE TABLE records22 (year STRING, temperature INT, quality INT)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '\t';

LOAD DATA LOCAL INPATH 'pig/sample.txt'

OVERWRITE INTO TABLE records22;

SELECT year, MAX(temperature)

FROM records22

WHERE temperature != 9999 AND quality IN (0, 1, 4, 5, 9)

GROUP BY year;

Text

Description automatically generated

2. Create the “partitions” in Hive. Print out the screenshot of all the partitions from the example in CSUEB Hadoop.

CREATE TABLE logs22 (ts BIGINT, line STRING)

PARTITIONED BY (dt STRING, country STRING);

LOAD DATA LOCAL INPATH 'file1'

INTO TABLE logs22

PARTITION (dt='2001-01-01', country='GB');

LOAD DATA LOCAL INPATH 'file2'

INTO TABLE logs22

PARTITION (dt='2001-01-01', country='GB');

LOAD DATA LOCAL INPATH 'file3'

INTO TABLE logs22

PARTITION (dt='2001-01-01', country='US');

LOAD DATA LOCAL INPATH 'file4'

INTO TABLE logs22

PARTITION (dt='2001-01-02', country='GB');

LOAD DATA LOCAL INPATH 'file5'

INTO TABLE logs22

PARTITION (dt='2001-01-02', country='US');

LOAD DATA LOCAL INPATH 'file6'

INTO TABLE logs22

PARTITION (dt='2001-01-02', country='US');

! echo; # vv partitions;

SHOW PARTITIONS logs22;

! echo; # ^^ partitions;

SELECT ts, dt, line

FROM logs22

WHERE country='GB';

Text

Description automatically generated with low confidence

3. Create the “buckets” in Hive. Print out the screenshot of the first bucket out of the 4 buckets from the example in CSUEB Hadoop.

DROP TABLE IF EXISTS users22;

CREATE TABLE users22 (id INT, name STRING);

LOAD DATA LOCAL INPATH 'users.txt'

OVERWRITE INTO TABLE users22;

DROP TABLE IF EXISTS bucketed\_users22;

CREATE TABLE bucketed\_users22 (id INT, name STRING)

CLUSTERED BY (id) INTO 4 BUCKETS;

dfs -ls /user/hive/warehouse/bucketed\_users22;

dfs -cat /user/hive/warehouse/bucketed\_users22/000000\_0;

SELECT \* FROM bucketed\_users22

TABLESAMPLE(BUCKET 1 OUT OF 4 ON id);

SELECT \* FROM bucketed\_users22

TABLESAMPLE(BUCKET 1 OUT OF 2 ON id);

SELECT \* FROM users22

TABLESAMPLE(BUCKET 1 OUT OF 4 ON rand());

Text

Description automatically generated