**HQL script creating the data warehouse**

CREATE TABLE route (

    route\_id INT,

    route\_name STRING,

    metrics MAP<STRING, INT>

)

CLUSTERED BY (route\_id) INTO 10 buckets

STORED AS ORC;

CREATE TABLE date\_dim (

    date\_id INT,

    date\_format DATE,

    year INT,

    month STRING

)

PARTITIONED BY (month\_no INT, day\_type STRING)

CLUSTERED BY (date\_id) INTO 30 BUCKETS

STORED AS PARQUET;

CREATE TABLE time\_dim (

    time\_id INT,

    hour INT,

    minutes INT

)

PARTITIONED BY (time\_of\_day STRING)

STORED AS PARQUET;

CREATE TABLE travel (

    bus\_id INT,

    route\_id INT,

   departure\_time INT,

    arrival\_time INT,

   tickets\_validated INT,

   bus\_capacity INT,

   avg\_satisfaction\_level\_received INT,

   satisfaction\_surveys\_number INT,

   junk\_id INT

)

PARTITIONED BY (travel\_date INT)

CLUSTERED BY (route\_id, bus\_id) INTO 50 BUCKETS

STORED AS ORC;

CREATE TABLE junk (

    junk\_id INT,

    satisfaction\_level\_category STRING,

    occupation\_level\_category STRING

)

STORED AS PARQUET;

CREATE TABLE bus (

   bus\_id INT,

    bus\_registration STRING,

    bus\_office\_id INT,

   additional\_equipment ARRAY<STRING>

)

PARTITIONED BY (bus\_type STRING)

CLUSTERED BY (bus\_id) INTO 10 BUCKETS

STORED AS ORC;

CREATE EXTERNAL TABLE bus\_office (

    bus\_office\_id INT,

   bus\_office\_name STRING,

   city STRING,

   size\_category STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE

LOCATION 'hdfs:///user/mmajewska/database';