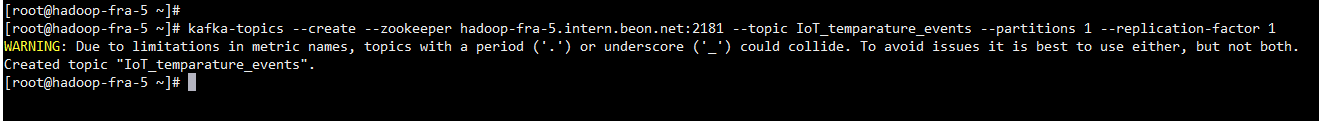
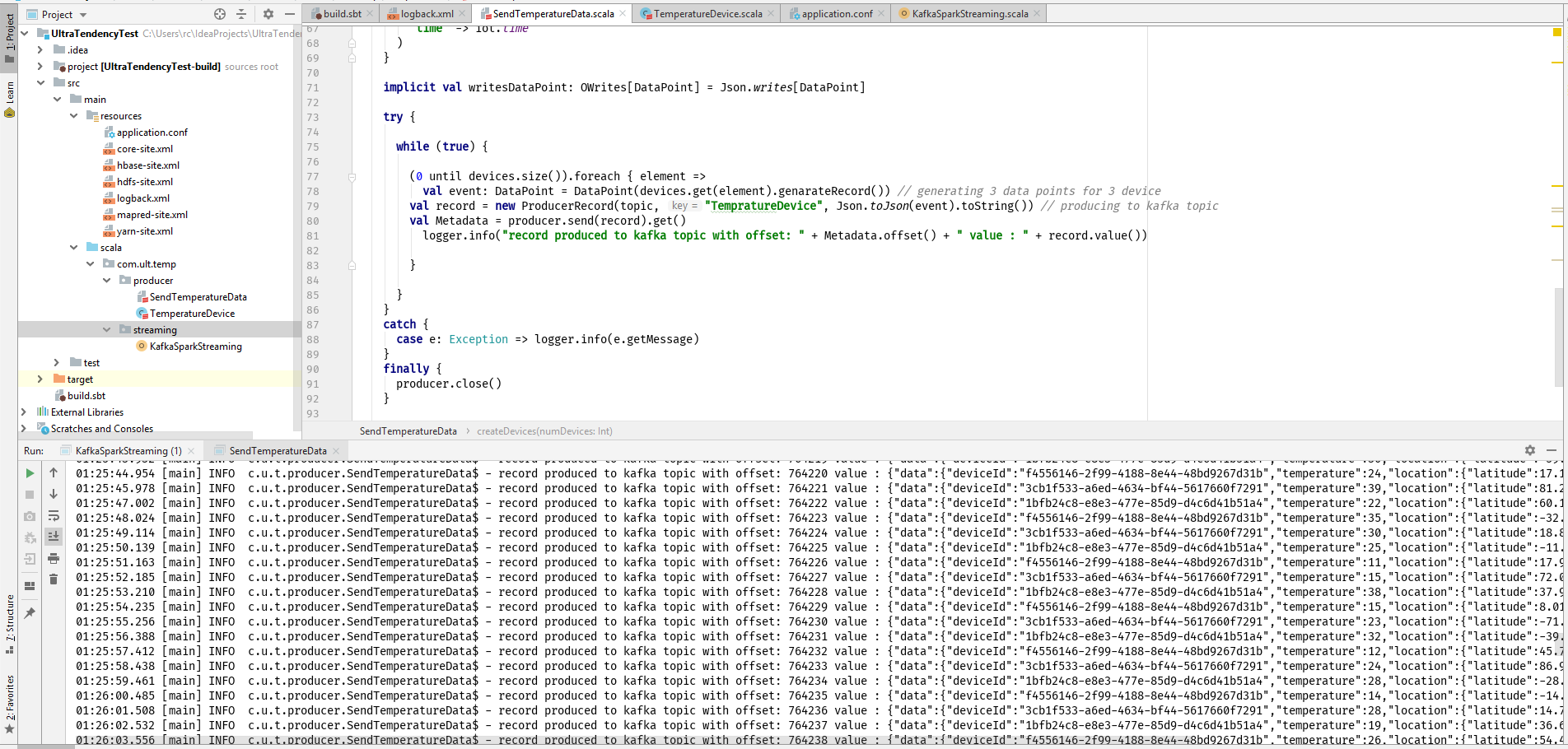
**Ultra Tendency Data Engineering Challenge Practical Assignment**

**Task 1:**

1. Create Kafka topic for IoT data events

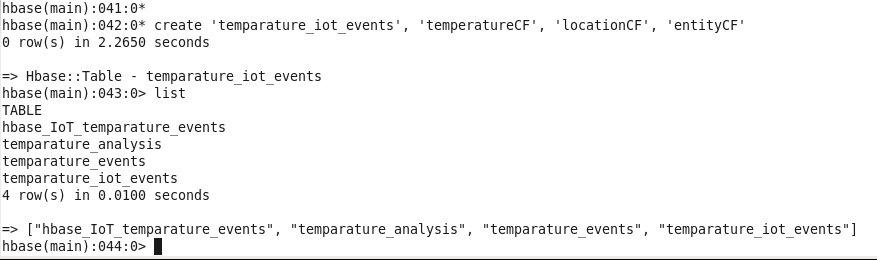


1. Write a IoT data simulator i.e. Kafka producer application which generate random data points (please refer to the SendTemperatureData.scala class in attachment)

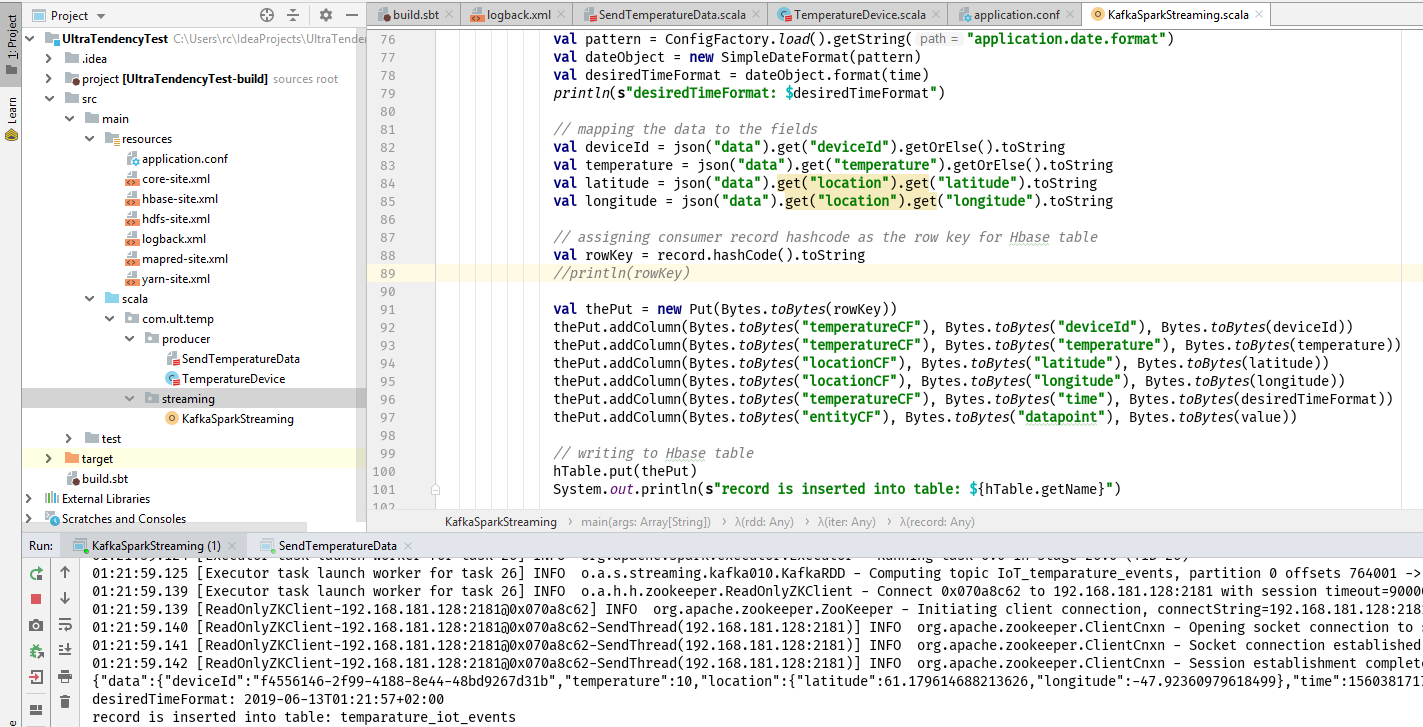


**Task 2:**

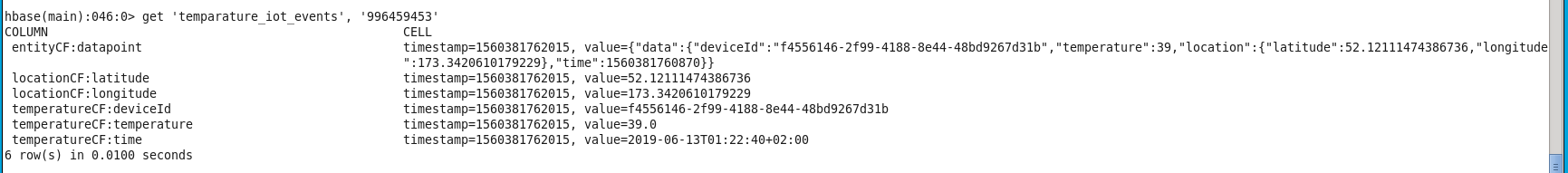
1. Create a Hbase table with two column families as below



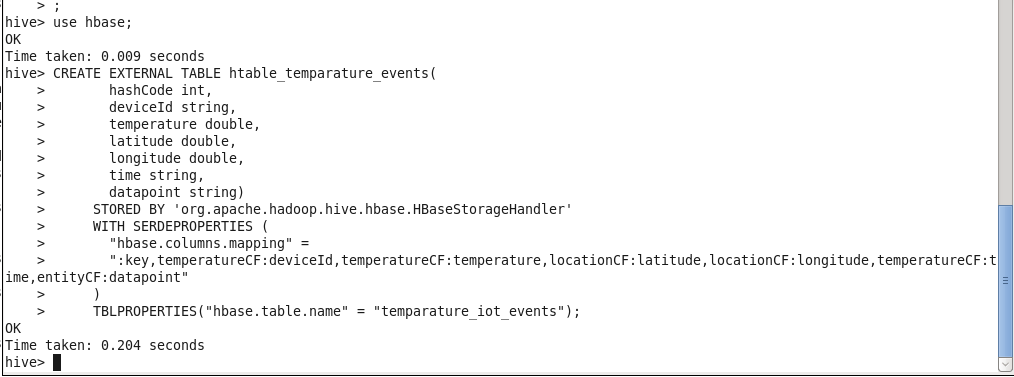
1. Write a spark streaming application which reads events from kafka topic created above and change the time column to human readable format and store it in Hbase table created above. . (please refer to the TempEventsToHbase.scala class in attachment)



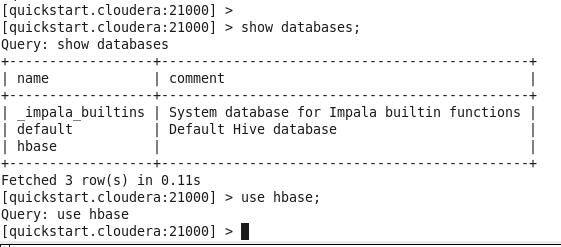
1. Verify that data is produced to Hbase table



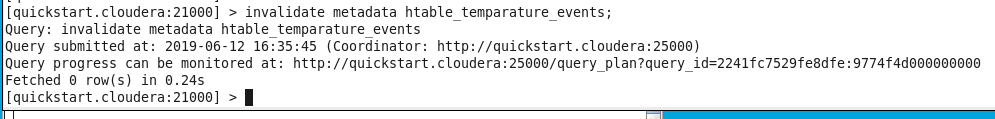
1. Create a hive table on top of hbase table created above and do the column mapping as below



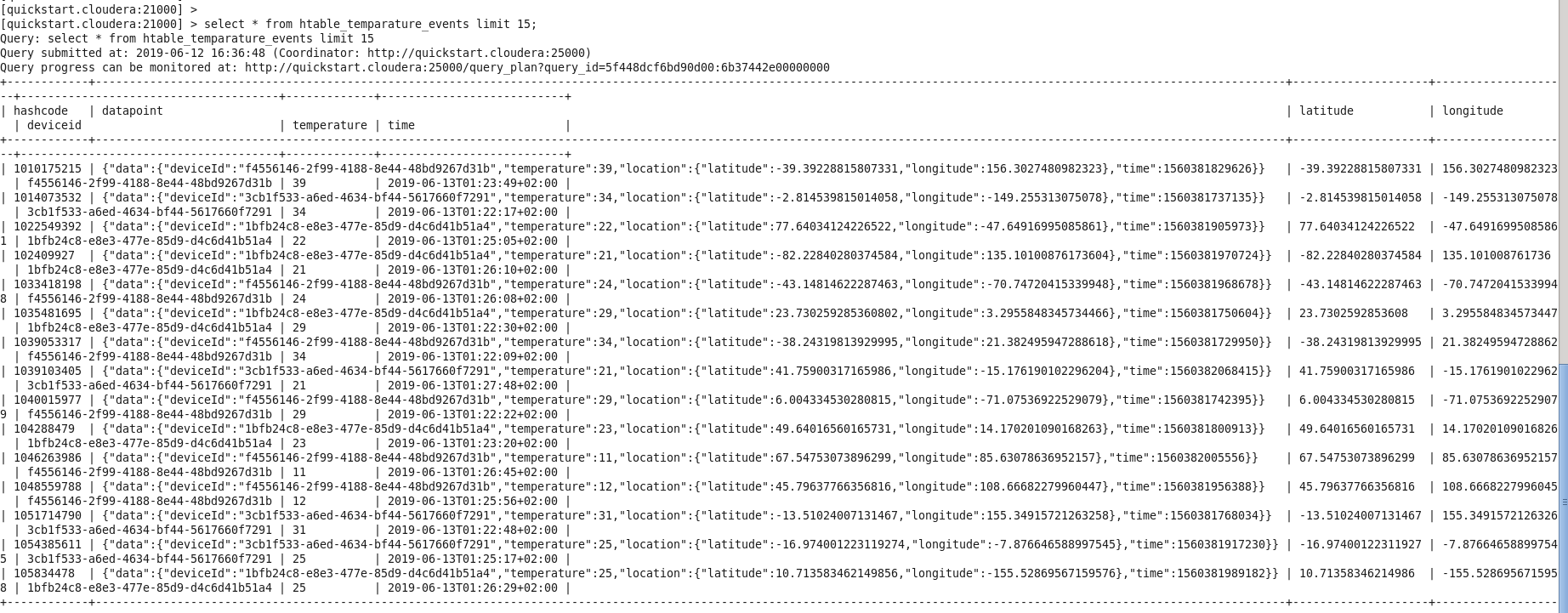
1. Open Impala shell and create a database named **hbase**

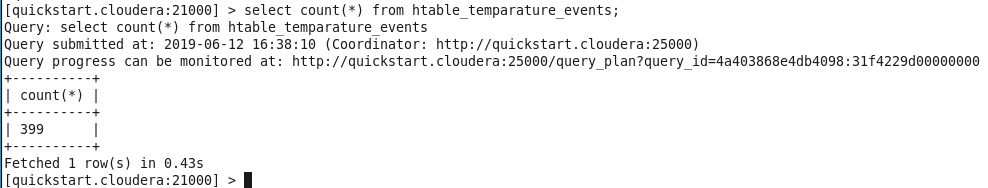


1. Run the below command so that impala get to know about the table created in hive shell above.



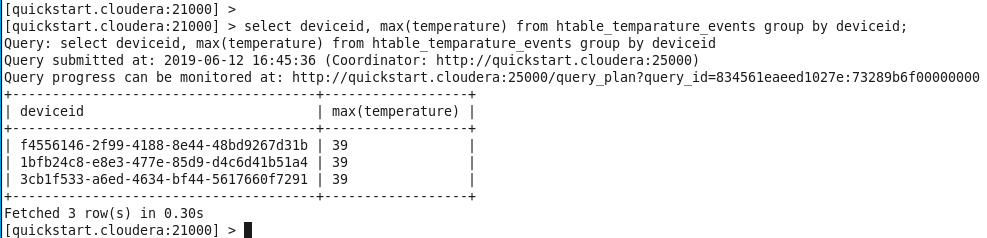
1. List few records to check whether impala can read data from Hbase table.





**Task 3:**

1. **The maximum temperatures measured for every device.**



[quickstart.cloudera:21000] > select deviceid, max(temperature) from htable\_temparature\_events group by deviceid;

Query: select deviceid, max(temperature) from htable\_temparature\_events group by deviceid

Query submitted at: 2019-06-12 16:45:36 (Coordinator: http://quickstart.cloudera:25000)

Query progress can be monitored at: http://quickstart.cloudera:25000/query\_plan?query\_id=834561eaeed1027e:73289b6f00000000

+--------------------------------------+------------------+

| deviceid | max(temperature) |

+--------------------------------------+------------------+

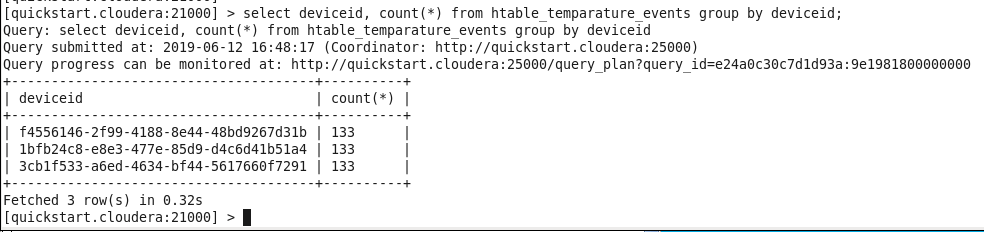
| f4556146-2f99-4188-8e44-48bd9267d31b | 39 |

| 1bfb24c8-e8e3-477e-85d9-d4c6d41b51a4 | 39 |

| 3cb1f533-a6ed-4634-bf44-5617660f7291 | 39 |

+--------------------------------------+------------------+

1. **The amount of data points aggregated for every device.**



[quickstart.cloudera:21000] > select deviceid, count(\*) from htable\_temparature\_events group by deviceid;

Query: select deviceid, count(\*) from htable\_temparature\_events group by deviceid

Query submitted at: 2019-06-12 16:48:17 (Coordinator: http://quickstart.cloudera:25000)

Query progress can be monitored at: http://quickstart.cloudera:25000/query\_plan?query\_id=e24a0c30c7d1d93a:9e1981800000000

+--------------------------------------+----------+

| deviceid | count(\*) |

+--------------------------------------+----------+

| f4556146-2f99-4188-8e44-48bd9267d31b | 133 |

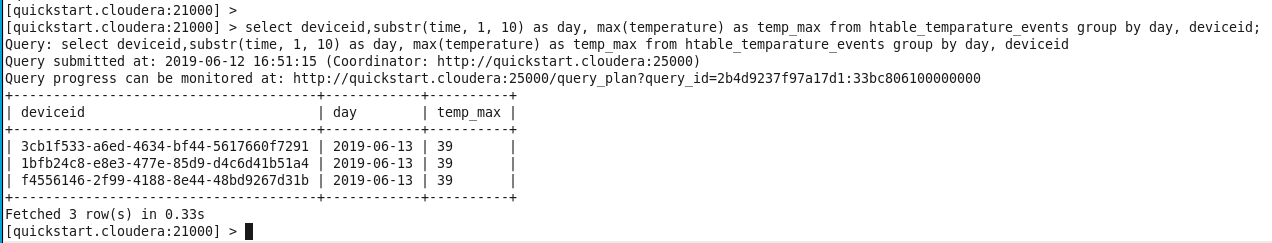
| 1bfb24c8-e8e3-477e-85d9-d4c6d41b51a4 | 133 |

| 3cb1f533-a6ed-4634-bf44-5617660f7291 | 133 |

+--------------------------------------+----------+

Fetched 3 row(s) in 0.32s

1. **The highest temperature measured on a given day for every device.**



[quickstart.cloudera:21000] > select deviceid,substr(time, 1, 10) as day, max(temperature) as temp\_max from htable\_temparature\_events group by day, deviceid;

Query: select deviceid,substr(time, 1, 10) as day, max(temperature) as temp\_max from htable\_temparature\_events group by day, deviceid

Query submitted at: 2019-06-12 16:51:15 (Coordinator: http://quickstart.cloudera:25000)

Query progress can be monitored at: http://quickstart.cloudera:25000/query\_plan?query\_id=2b4d9237f97a17d1:33bc806100000000

+--------------------------------------+------------+----------+

| deviceid | day | temp\_max |

+--------------------------------------+------------+----------+

| 3cb1f533-a6ed-4634-bf44-5617660f7291 | 2019-06-13 | 39 |

| 1bfb24c8-e8e3-477e-85d9-d4c6d41b51a4 | 2019-06-13 | 39 |

| f4556146-2f99-4188-8e44-48bd9267d31b | 2019-06-13 | 39 |

+--------------------------------------+------------+----------+

Fetched 3 row(s) in 0.33s