Week 7 FAQs

W7:1 In hbase we can achieve random access and by using orc, parquet also we can achieve random access right the both statements are valid right. please clarify

Ans: Right searching/scanning in HBase is quick as hFile can be of max 64KB.

W7:2 How is the logical grouping performed and how the column families will be made?

Ans: Although question is not very clear to me, but we can specify the column families while creating the table in Hbase
e.g.
create 'students', 'personal_details', 'contact_details', 'marks' this creates table students with 3 column families.

W7:3 What happens when a zookeeper fails in Hbase?

Ans: Zookeeper also runs in a zookeeper cluster(Zookeeper quorum). It has a concept of zookeeper leader and followers , so even if the leader zookeeper fails, a new leader is elected among the followers.

W7:4 How to execute Hbase commands on itversity? I do not have sudo access hence can't restart hbase-master and hbase-regionserver. Due to this I am unable to create tables. Plz help soon

Ans: if services are not running contact itversity support

W7:5 Since Metatable(which holds the information of region,region server along with row key) is present in one of the Region servers. And what if that region server goes down? Will the Zookeeper do something?

Ans: When a RegionServer crashes the Master allocates the regions handled by that server to another regionServer. Hope it's clear

W7:6 during 'Hive-hbase integration' when I try to create a Hbase table in hive the following error occur, and checked that 'HBase master daemon is running' & 'hbase-regionserver is running'

```
hive> CREATE TABLE hbase_table_1(key INT,value STRING) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERD
EPROPERTIES ("hbase.columns.mapping" = ":key,cf1:val") TBLPROPERTIES ("hbase.table.name" ="xyz");
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. MetaException(message:org.apache.hadoop.h
base.client.RetriesExhaustedException: Failed after attempts=36, exceptions:
Tue Oct 27 12:03:52 PDT 2020, null, java.net.SocketTimeoutException: callTimeout=60000, callDuration=67862: row 'xyz,,' on ta
ble 'hbase:meta' at region=hbase:meta,,1.1588230740, hostname=quickstart.cloudera,60020,1603742381319; seqNum=0

        at org.apache.hadoop.hbase.client.RpcRetryingCallerWithReadReplicas.throwEnrichedException(RpcRetryingCallerWithReadR
eplicas.java:320)
        at org.apache.hadoop.hbase.client.ScannerCallableWithReplicas.call(ScannerCallableWithReplicas.java:247)
        at org.apache.hadoop.hbase.client.ScannerCallableWithReplicas.call(ScannerCallableWithReplicas.java:62)
        at org.apache.hadoop.hbase.client.RpcRetryingCaller.callWithoutRetries(RpcRetryingCaller.java:210)
        at org.apache.hadoop.hbase.client.ClientScanner.call(ClientScanner.java:327)
        at org.apache.hadoop.hbase.client.ClientScanner.nextScanner(ClientScanner.java:302)
        at org.apache.hadoop.hbase.client.ClientScanner.initializeScannerInConstruction(ClientScanner.java:167)
        at org.apache.hadoop.hbase.client.ClientScanner.<init>(ClientScanner.java:162)
        at org.apache.hadoop.hbase.client.HTable.getScanner(HTable.java:862)
        at org.apache.hadoop.hbase.MetaTableAccessor.fullScan(MetaTableAccessor.java:602)
        at org.apache.hadoop.hbase.MetaTableAccessor.tableExists(MetaTableAccessor.java:366)
        at org.apache.hadoop.hbase.client.HBaseAdmin.tableExists(HBaseAdmin.java:421)
        at org.apache.hadoop.hbase.client.HBaseAdmin.tableExists(HBaseAdmin.java:431)
        at org.apache.hadoop.hive.hbase.HBaseStorageHandler.preCreateTable(HBaseStorageHandler.java:195)
        at org.apache.hadoop.hive.metastore.HiveMetaStoreClient.createTable(HiveMetaStoreClient.java:727)
        at org.apache.hadoop.hive.metastore.HiveMetaStoreClient.createTable(HiveMetaStoreClient.java:728)
```

Ans
https://stackoverflow.com/questions/49379434/cant-connect-to-hbase-via-hive-with-jdbc
Same issue for me. When I checked the service both Master and region-server are not running. I did restart the service and created the Hive-Hbase table; it worked fine.

W7:7 In "Hadoop Integration project' , if I use sqoop create-hive-table , then the hive tables is created in default database but what is the sqoop syntax if I want that hive table should create in other database

Ans:-In --hive-table argument you can mention database_name.table_name to create hive table in the specific database

W7:8 when running the command sudo service --status-all getting error help me resolving this

Ans:-write it on normal terminal not hbase shell,hope it will work

W7:9 Error:-

Caused by: java.net.SocketTimeoutException: callTimeout=60000, callDuration=68432: row 'xyz,,' on table 'hbase:meta' at region=hbase:meta,,1.1588230740, hostname=quickstart.cloudera,60020,1603863156728, seqNum=0
        at
org.apache.hadoop.hbase.client.RpcRetryingCaller.callWithRetries(RpcRetryingCaller.java:169)

Ph:9108179578

Ans:- Try to restart HBase master and region server. And give a try

W7:10 In hive-hbase table, when we insert data and if we have duplicate keys based on which factor it will discard the duplicate row? Which row will be deleted and which will be kept? How can we make sure the latest data is loaded and previous data is discarded?

Ans:- It won't discard. Instead it will update the TS, so you will see the latest data. To see the previous version use below

scan '<table name>', {TIMERANGE => [0, 1416083300000] }

Just check the latest timestamp value alone.

W7:11 In the scenario where we are loading the hive-hbase table from a hive table which has duplicate records , then in the hive-hbase table we will just see one record. On what bases that record will be selected?

Ans:- Consider column id (int) is a key in table 'x'.
First time we inserting like below
put 'x', 'cf:id', 1
And then we inserting 2
put 'x', 'cf:id', 2
scan 'x'
If you scan, you will get only id 2. Id 1 is still there in the table, but you will get only the latest row. To see the previous row use below
scan '<table name>', {TIMERANGE => [0, 1416083300000] }
Note: 1416083300000 take this timestamp value from scan 'x' after inserting id 2.

W7:12 Is there any criteria based on which column family should be decided? Like the number of Column families in a table, which column should go to which column family?

Ans:- No.. but you want to have 2 to 3 column families not just one. because you get benefits.. data related to one column family is stored in one file.
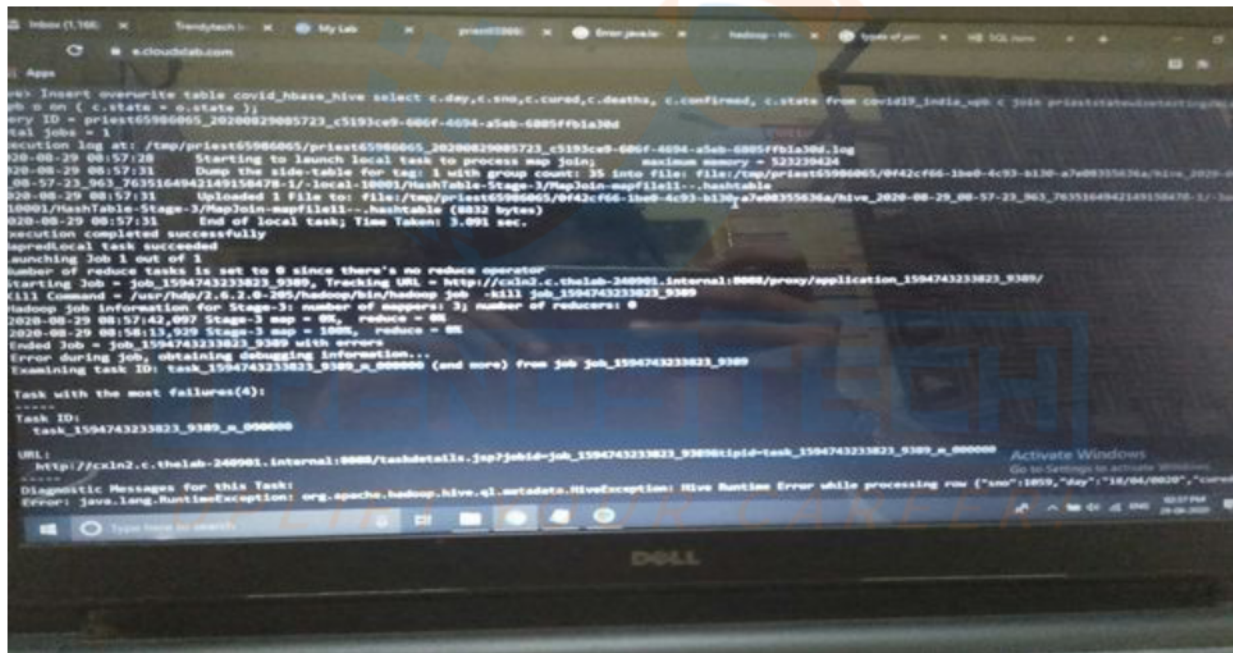this is a logical grouping and all is in our control..

Ph:9108179578

W7:13 how to ensure that HBase master and RegionServer services are running all the time even when you save the machine state in Cloudera VM?

Ans : there is an issue with VM. The services get turned off. so don't worry. It's an issue which most of you will face.

W7:14 In Hbase, can I skip including column families in a table structure? I mean let's say I want only columns in my hbase table and not column families. Is it allowed?

Ans : No it's not. At least one column family is required

W7:15 When I'm inserting data on hbase_hive table from doing the join operation of 2 hive tables it is showing runtime error ?

Ans : No.but you want to have 2 to 3 column families not just one. because you get benefits.. data related to one column family is stored in one file.this is a logical grouping and all is in our control.

W7:16 What happens when a zookeeper fails in Hbase?

Ans: Zookeeper also runs in a zookeeper cluster(Zookeeper quorum). It has a concept of zookeeper leader and followers , so even if the leader zookeeper fails, a new leader is elected among the followers.

W7:17 Getting the below error when trying to create the Hive-Hbase table. Did anyone else face this issue? I made sure my Hbase master and region-server are running.

CREATE TABLE hbase_table_1(key int, value string) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,cf1:val") TBLPROPERTIES ("hbase.table.name" = "xyz");

FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask.

MetaException(message:org.apache.hadoop.hbase.client.RetriesExhaustedException: Failed after attempts=36, exceptions:

Wed Dec 30 13:32:19 PST 2020, null, java.net.SocketTimeoutException: callTimeout=60000, callDuration=68281: row 'xyz,,' on table 'hbase:meta' at region=hbase:meta,,1.1588230740, hostname=quickstart.cloudera,60020,1609259744867, seqNum=0

at

org.apache.hadoop.hbase.client.RpcRetryingCallerWithReadReplicas.throwEnrichedExcepti on(RpcRetryingCallerWithReadReplicas.java:320)...etc

Ans: check sudu Services:-

Sudo service hbase-master restart

Sudo service hbase-Regionserver restart

Then create a table

Dated Till:-3rd Jan 2021