

1. Create this table in Hive CLI for sqoop incremental import of card_member table *[if not created as part of Capstone-Script.pdf Task 4]* :

```
CREATE TABLE IF NOT EXISTS card_member_incremental(  
card_id string,  
member_id string,  
member_joining_dt timestamp,  
card_purchase_dt string,  
country string,  
city string  
)ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde' WITH  
SERDEPROPERTIES ("separatorChar" = ",");
```

2. Add the necessary jars for e.g. – sql-connector jar into Oozie – Sqoop sharelib path.
3. Add below jars / xml files into Oozie – Hive sharelib path :
hive-site.xml

hbase-site.xml

```
hadoop fs -put /etc/hbase/conf.cloudera.hbase/hbase-site.xml hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

metrics-core-2.2.0.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/jars/metrics-core-  
2.2.0.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

htrace-core.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/lib/hive/lib/htrace-  
core.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/  
/opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/lib/hbase/lib/htrace-core.jar
```

hive-hbase-handler-1.1.0-cdh5.13.0.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/jars/hive-hbase-  
handler-1.1.0-cdh5.15.0.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

hive-hbase-handler.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/lib/hive/lib/hive-  
hbase-handler.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

hbase-common.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/lib/hive/lib/hbase-  
common.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

hbase-server.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/lib/hive/lib/hbase-  
server.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

hbase-hadoop-compat.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/lib/hive/lib/hbase-  
hadoop-compat.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

hbase-hadoop2-compat.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/lib/hive/lib/hbase-  
hadoop2-compat.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

hbase-client.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/lib/hive/lib/hbase-  
client.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

hbase-protocol.jar

```
hadoop fs -put /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/lib/hive/lib/hbase-  
protocol.jar hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

4. Assign permission to Oozie on sharelib path for all the added jars:

```
hadoop fs -chown -R oozie hdfs://ip-172-31-46-  
84.ec2.internal:8020/user/oozie/share/lib/lib_20180815055730/hive/
```

5. Add the necessary properties to sqoop-site.xml file and put it at the same place where you will have the workflow.xml

```
<property>  
  <name>sqoop.metastore.client.record.password</name>  
  <value>true</value>  
  <description>If true, allow saved passwords in the metastore.  
</description>  
</property>
```

```

<property>
  <name>sqoop.metastore.client.autoconnect.password</name>
  <value>upgraduser</value>
  <description>The password to bind to the metastore.
</description>
</property>

<property>
  <name>sqoop.metastore.client.autoconnect.url</name>
  <value>jdbc:hsqldb:hsqldb://ip-172-31-46-84.internal:16000/sqoop</value>
  <description>The connect string to use when connecting to a
    job-management metastore. If unspecified, uses ~/.sqoop/.
    You can specify a different path here.
  </description>
</property>

<property>
  <name>sqoop.metastore.client.enable.autoconnect</name>
  <value>true</value>
  <description>If true, Sqoop will connect to a local metastore
    for job management when no other metastore arguments are
    provided.
  </description>
</property>

```

6. Have the folder structure ready for Oozie runs: **[This structure should not be altered]**

In HDFS:

```

[root@ip-172-31-46-84 app]# hadoop fs -mkdir /user/root/oozie_sqoop
[root@ip-172-31-46-84 app]# hadoop fs -mkdir /user/root/oozie_sqoop/app/

```

On root:

```

[root@ip-172-31-46-84 ~]# mkdir oozie_sqoop
[root@ip-172-31-46-84 ~]# cd oozie_sqoop/
[root@ip-172-31-46-84 oozie_sqoop]# mkdir app

```

Adding files:

```

[root@ip-172-31-46-84 oozie_sqoop]# aws s3 cp s3://capstone-
deepika/oozie_sqoop/job.properties .

```

```

[root@ip-172-31-46-84 oozie_sqoop]# hadoop fs -put job.properties
/user/root/oozie_sqoop/

```

```

[root@ip-172-31-46-84 oozie_sqoop]# cd app

```

```

[root@ip-172-31-46-84 app]# aws s3 cp s3://capstone-deepika/oozie_sqoop/workflow.xml .

```

```
[root@ip-172-31-46-84 app]# hadoop fs -put workflow.xml /user/root/oozie_sqoop/app/
```

After changes to sqoop-site.xml, get the global sqoop-site.xml to 'root' user and then copy it to below folders:

```
[root@ip-172-31-46-84 ~]# cp sqoop-site.xml oozie_sqoop/app/
[root@ip-172-31-46-84 ~]# hadoop fs -put sqoop-site.xml /user/root/oozie_sqoop/app/
```

Add the coordinator.xml for Oozie run as follows:

```
[root@ip-172-31-46-84 oozie_sqoop]# aws s3 cp s3://capstone-deepika/oozie_sqoop/coordinator.xml .
```

```
[root@ip-172-31-46-84 oozie_sqoop]# hadoop fs -put coordinator.xml
/user/root/oozie_sqoop/
```

Add the Hive script for the workflow as follows:

```
[root@ip-172-31-46-84 app]# aws s3 cp s3://capstone-deepika/oozie_sqoop/hiveScript.sql .
```

```
[root@ip-172-31-46-84 app]# hadoop fs -put hiveScript.sql /user/root/oozie_sqoop/app/
```

7. Now start the sqoop metastore for Sqoop Jobs and incremental imports:

```
[root@ip-172-31-46-84 oozie_sqoop]# sudo -u sqoop sqoop-metastore
```

8. Create Sqoop Import Jobs as below on different putty instance *[if not created as part of Capstone-Script.pdf Task 4]* :

Creating Sqoop Job 1 for incremental import of card_member table data

```
[root@ip-172-31-46-84 ~]# sqoop job --create inc_upsert_card_member --meta-connect
jdbc:hsqldb:hsqldb://ip-172-31-46-84.ec2.internal:16000/sqoop -- import --connect
jdbc:mysql://upgradawsrds.cpclxrkdvwzmz.us-east-1.rds.amazonaws.com/cred_financials_data --
username upgraduser --password upgraduser --table card_member --null-string 'NA' -m 1 --
incremental lastmodified --check-column member_joining_dt --last-value 0 --merge-key card_id --
target-dir /capstone/card_member_incremental
```

Check the created sqoop job

```
[root@ip-172-31-46-84 ~]# sqoop job --list --meta-connect jdbc:hsqldb:hsqldb://ip-172-31-46-
84.ec2.internal:16000/sqoop
```

Warning: /opt/cloudera/parcels/CDH-5.15.0-1.cdh5.15.0.p0.21/bin/../lib/sqoop/./accumulo does not exist! Accumulo imports will fail.

Please set \$ACCUMULO_HOME to the root of your Accumulo installation.

18/12/06 14:30:05 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.15.0

Available jobs:

inc_upsert_card_member

update_member_score

Checking if created job executes fine without the Oozie workflow:

```
[root@ip-172-31-46-84 ~]# sqoop job --exec inc_upsert_card_member --meta-connect  
jdbc:hsqldb:hsqldb://ip-172-31-46-84.ec2.internal:16000/sqoop
```

Creating Sqoop Job 2 for member_score full import:

```
[root@ip-172-31-46-84 ~]# sqoop job --create update_member_score --meta-connect  
jdbc:hsqldb:hsqldb://ip-172-31-46-84.ec2.internal:16000/sqoop -- import --connect  
jdbc:mysql://upgradawsrds.cpclxrkdvwvmz.us-east-1.rds.amazonaws.com/cred_financials_data --  
username upgraduser --password upgraduser --table member_score --null-string 'NA' -m 1 --  
delete-target-dir --target-dir /capstone/member_score
```

Check the created sqoop job

```
sqoop job --list --meta-connect jdbc:hsqldb:hsqldb://ip-172-31-46-84.ec2.internal:16000/sqoop
```

Checking the details of created job:

```
[root@ip-172-31-46-84 ~]# sqoop job --show update_member_score --meta-connect  
jdbc:hsqldb:hsqldb://ip-172-31-46-84.ec2.internal:16000/sqoop
```

Checking if created job executes fine without the Oozie workflow:

```
sqoop job --exec update_member_score --meta-connect jdbc:hsqldb:hsqldb://ip-172-31-46-  
84.ec2.internal:16000/sqoop
```

9. Oozie execution for the workflow *without Oozie Coordinator*:

```
[root@ip-172-31-46-84 ~]# cd oozie_sqoop/
```

```
[root@ip-172-31-46-84 oozie_sqoop]# oozie job -oozie http://ip-172-31-46-84.ec2.  
internal:11000/oozie -config job.properties -run
```

```
job: 00000000-181206165718503-oozie-oozi-W
```

```
[root@ip-172-31-46-84 oozie_sqoop]# oozie job -oozie http://ip-172-31-46-84.ec2.  
internal:11000/oozie -info 00000000-181206165718503-oozie-oozi-W
```

```
Job ID : 00000000-181206165718503-oozie-oozi-W
```

```

[root@ip-172-31-46-84 oozie_sqoop]# oozie job -oozie http://ip-172-31-46-84.ec2.internal:11000/oozie -info 0000005-181208053443289-oozie-oozi-W
Job ID : 0000005-181208053443289-oozie-oozi-W
-----
Workflow Name : capstone
App Path      : hdfs://ip-172-31-46-84.ec2.internal:8020/user/root/oozie_sqoop/app
Status        : SUCCEEDED
Run           : 0
User          : root
Group         : -
Created       : 2018-12-08 08:10 GMT
Started       : 2018-12-08 08:10 GMT
Last Modified : 2018-12-08 08:25 GMT
Ended        : 2018-12-08 08:25 GMT
CoordAction ID: -

Actions
-----
ID                               Status  Ext ID  Ext Status  Err Code
-----
0000005-181208053443289-oozie-oozi-W@start:  OK      -      OK          -
0000005-181208053443289-oozie-oozi-W@card_member_import  OK      job_1544247269872_0035  SUCCEEDED  -
0000005-181208053443289-oozie-oozi-W@member_score_import  OK      job_1544247269872_0037  SUCCEEDED  -
0000005-181208053443289-oozie-oozi-W@load_hbase_lookup  OK      job_1544247269872_0039  SUCCEEDED  -
0000005-181208053443289-oozie-oozi-W@finish  OK      -      OK          -
-----
[root@ip-172-31-46-84 oozie_sqoop]#

```

10. Oozie execution for the workflow with Oozie Coordinator:

```

[root@ip-172-31-46-84 oozie_sqoop]# oozie job -oozie http://ip-172-31-46-84.ec2.internal:11000/oozie -config job.properties -run

```

```

job: 0000000-181208144337430-oozie-oozi-C

```

```

[root@ip-172-31-46-84 oozie_sqoop]# oozie job -oozie http://ip-172-31-46-84.ec2.internal:11000/oozie -info 0000000-181208144337430-oozie-oozi-C

```

```

[root@ip-172-31-46-84 oozie_sqoop]# oozie job -oozie http://ip-172-31-46-84.ec2.internal:11000/oozie -info 0000000-181208144337430-oozie-oozi-C
Job ID : 0000000-181208144337430-oozie-oozi-C
-----
Job Name      : capstone-coord
App Path      : hdfs://ip-172-31-46-84.ec2.internal:8020/user/root/oozie_sqoop/coordinator.xml
Status        : RUNNING
Start Time    : 2018-12-08 15:02 GMT
End Time      : 2019-01-31 01:00 GMT
Pause Time    : -
Concurrency   : 2

ID                               Status  Ext ID  Err Code  Created          Nominal Time
-----
0000000-181208144337430-oozie-oozi-C@1  WAITING  -      -          2018-12-08 14:59 GMT  2018-12-08 16:00 GMT
-----

```