**INCREMENTAL DATA INGESTION**

The following steps illustrate how we can ingest data incrementally from SQL Server into HIVE Table in Hadoop Cluster.

Step1:

**Sqoop-script for importing the latest data into HIVE table.**

sqoop import --connect "jdbc:sqlserver://10.10.26.25:1433;username=test;password=test;database=testcr" --table trans --hive-table trans --hive-import -m 5

Step 2:

**Create incremental sqoop job for appending daily data.**

sqoop job --create trans\_increment -- import --table trans --connect "jdbc:sqlserver://10.10.26.25:1433;username=test;password=test;database=testcr" --hive-import --hive-table trans --split-by atminternalid --incremental append --check-column atminternalid --last-value 32931 --verbose -m 5

Step 3:

**Create oozie-worflow script to execute Sqoop Action.**

<!--\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*-->

<!-- workflow.xml for Sqoop Action -->

<!--\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*-->

<workflow-app name="WorkflowWithSqoopAction" xmlns="uri:oozie:workflow:0.2">

<start to="sqoopAction"/>

<action name="sqoopAction">

<sqoop xmlns="uri:oozie:sqoop-action:0.2">

<job-tracker>${jobTracker}</job-tracker>

<name-node>${nameNode}</name-node>

<arg>job</arg>

<arg>--exec</arg>

<arg>trans\_increment</arg><arg>--verbose</arg>

</sqoop>

<ok to="end"/>

<error to="killJob"/>

</action>

<kill name="killJob">

<message>"Killed job due to error: ${wf:errorMessage(wf:lastErrorNode())}"</message>

</kill>

<end name="end" />

</workflow-app>

Step4:

**Create oozie-job property file to execute oozie-worflow.**

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# job.properties for Sqoop Action

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

nameNode=hdfs://agent2.poc.com:8020

jobTracker=agent2.poc.com:8050

queueName=cardtronics

oozie.libpath=${nameNode}/user/oozie/share/lib

oozie.use.system.libpath=true

oozieProjectRoot=${nameNode}/user/${user.name}/cardtronics

appPath=${oozieProjectRoot}/apps/sqoop

oozie.wf.application.path=${appPath}

#\*\*\*\*\*\*\*\*\*\*end\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Step5:

**Running oozie job to execute oozie-workflow.**

oozie job -oozie http://localhost:11000/oozie -config /opt/cardtronics/apps/sqoop/job.properties -run -verbose