**Oozie Workflow Automation**

**(LGL\_IIR)**

**Prerequisite (Required files to run the Oozie Workflow Automation):**

kinit -kt /etc/security/keytabs/x987731.keytab x987731@NMCORP.NISSAN.BIZ

EWS\_IIR\_Ingest\_Inc\_daily.ini

iir\_sqoop.properties

sqoop\_arch.properties

sqoop\_footer.xml

sqoop\_header.xml

sqoop\_full.xml

sqoop\_inc.xml

data\_extrcn\_rule\_bfr\_run.hql

data\_extrcn\_rule\_aftr\_run.hql

***The below external tables (folders) and Hive tables need to be created by BDE admin (if it doesn’t exists)***

hdfs://bdedev/projects/ews/oozie/iir-ingest-daily

hdfs://bdedev/projects/ews/scripts/spark/search/conf/iir\_sqoop

hdfs://bdedev/data/lgl\_iir/raw/incdnt\_ivstgn\_raw/

hdfs://bdedev/data/lgl\_iir/raw/incdnt\_ivstgn\_trd\_raw/

hdfs://bdedev/data/lgl\_iir/base/incdnt\_ivstgn\_base/

hdfs://bdedev/data/lgl\_iir/base/incdnt\_ivstgn\_trd\_base/

**Hive tables needs to be created:**

LGL\_IIR.incdnt\_ivstgn\_raw

LGL\_IIR.incdnt\_ivstgn\_base

LGL\_IIR.incdnt\_ivstgn\_trd\_raw

LGL\_IIR.incdnt\_ivstgn\_trd\_base

**Sample Files:**

|  |  |
| --- | --- |
| **File Name** | **Location** |
| EWS\_Drive\_Ingest\_Inc\_daily.ini (from HDFS) | hdfs://bdedev/projects/ews/scripts/spark/search/conf/EWS\_Drive\_Ingest\_Inc\_daily.ini |
| drive\_sqoop.properties (from HDFS) | hdfs://bdedev/projects/ews/scripts/spark/search/conf/drive\_sqoop/drive\_sqoop.properties |
| sqoop\_arch.properties(From SVN) | <http://svn.na.nissan.biz/svn/NNA/NNA_TCS_EWS_BDE/trunk/dev/migrate/sqoop_arch.properties> |
| sqoop\_footer.xml | hdfs://bdedev/projects/ews/scripts/spark/search/conf/drive\_sqoop/sqoop\_footer.xml |
| sqoop\_header.xml | hdfs://bdedev/projects/ews/scripts/spark/search/conf/drive\_sqoop/sqoop\_header.xml |
| sqoop\_full.xml | hdfs://bdedev/projects/ews/scripts/spark/search/conf/drive\_sqoop/sqoop\_full.xml |
| sqoop\_inc.xml | hdfs://bdedev/projects/ews/scripts/spark/search/conf/drive\_sqoop/sqoop\_inc.xml |

**File editing:**

Replace the text “**drive**” with “**iir**” in the filenames and also inside the files

**Filename:** EWS\_Drive\_Ingest\_Inc\_daily.ini 🡺 EWS\_IIR\_Ingest\_Inc\_daily.ini

**Changes to do inside a file:**

conf\_location = /projects/ews/scripts/spark/search/conf/drive\_sqoop 🡺 conf\_location = /projects/ews/scripts/spark/search/conf/iir\_sqoop

properties\_file = drive\_sqoop.properties 🡺 properties\_file = iir\_sqoop.properties

wf\_location = ${nameNode}/projects/ews/oozie/drive-ingest-daily 🡺 wf\_location = ${nameNode}/projects/ews/oozie/iir-ingest-daily

convertScriptName = /projects/ews/scripts/hive/convert\_raw\_to\_base\_drive.hql 🡺 convertScriptName = /projects/ews/scripts/hive/convert\_raw\_to\_base\_iir.hql

hiveSchemaName = drive 🡺 hiveSchemaName = lgl\_iir

daysToAdv = 3 🡺 daysToAdv = 0

[cmdta] 🡺 [incdnt\_ivstgn]

rawTableName = CMDTA\_RAW 🡺 rawTableName = incdnt\_ivstgn\_raw

baseTableName = CMDTA\_BASE 🡺 baseTableName = incdnt\_ivstgn\_base

Copy the entire sqoopQuery from “Drive file” and change the following entries accordingly

|  |  |
| --- | --- |
| --connect ${driveodmConnector} | --connect ${bidwConnector} |
| --query "Drive Query” | --query "IIR Query” |
| --target-dir hdfs://bdedev/data/drive/landing/CMDTA\_RAW | --target-dir hdfs://bdedev/data/lgl\_iir/raw/incdnt\_ivstgn\_raw |
| --map-column-java “Drive\_CMDTA Columns” | --map-column-java INCDNT\_DTL\_DS=String |
| --merge-key CMDTA\_KY | --merge-key INCDNT\_IVSTGN\_RPT\_FORM\_NM |

**Filename:** drive\_sqoop.properties 🡺 iir\_sqoop.properties

**Changes to do inside a file:**

queueName=qcs\_dev 🡺 queueName=iir\_dev

driveodmConnector=jdbc:oracle:thin:@usnencvl916.nmcorp.nissan.biz:58533/ODDBDRVE 🡺 bidwConnector=jdbc:oracle:thin:@10.78.78.150:58532/ODDBEFLO

emailList=Xxxx.Yyyy@nissan-usa.com 🡺 [emailList=Aaaa.Bbbb@nissan-usa.com](mailto:emailList=Aaaa.Bbbb@nissan-usa.com)

username=EQUIPBDE\_ING 🡺 username=BDESELECT

passwordfile=${nameNode}/projects/ews/lib/.drivepassword 🡺 passwordfile=${nameNode}/projects/ews/lib/.bidwPassword

**Filename:** sqoop\_arch.properties (Keep the filename as it is)

**Changes to do inside a file:**

queueName=qcs\_dev 🡺 queueName=iir\_dev

wfName=EWS-Drive-Ingest-Daily 🡺 wfName=ews-iir-ingest-daily

coord\_Job\_Name = EWS-Drive-Ingest-Daily 🡺 coord\_Job\_Name=ews-iir-ingest-daily

emailList=Xxxx.Yyyy@nissan-usa.com 🡺 emailList=Aaaa.Bbbb@nissan-usa.com

WrkflwNames=EWS-Drive-Ingest-Inc-daily 🡺 WrkflwNames=EWS-IIR-Ingest-Inc-daily

Domain=gcars 🡺 Domain=iir

Copy sqoop\_\*.xml files in to the IIR HDFS location.

hdfs dfs -cp /projects/ews/scripts/spark/search/conf/drive\_sqoop/sqoop\_header.xml /projects/ews/scripts/spark/search/conf/iir\_sqoop/

hdfs dfs -cp /projects/ews/scripts/spark/search/conf/drive\_sqoop/sqoop\_footer.xml /projects/ews/scripts/spark/search/conf/iir\_sqoop/

hdfs dfs -cp /projects/ews/scripts/spark/search/conf/drive\_sqoop/sqoop\_full.xml /projects/ews/scripts/spark/search/conf/iir\_sqoop/

hdfs dfs -cp /projects/ews/scripts/spark/search/conf/drive\_sqoop/sqoop\_inc.xml /projects/ews/scripts/spark/search/conf/iir\_sqoop/

**Files to be copied prior to oozie run**

*For automation(TimeDep=Y):*

hdfs://bdedev/projects/ews/scripts/spark/search/conf/iir\_sqoop/iir\_sqoop.properties

hdfs://bdedev/projects/ews/scripts/spark/search/conf/EWS\_IIR\_Ingest\_Inc\_daily.ini

*For manual run(TimeDep=N):*

/home/x\*\*\*\*\*\*@NMCorp.Nissan.Biz/sqoop\_arch.properties

hdfs://bdedev/projects/ews/scripts/spark/search/conf/iir\_sqoop/iir\_sqoop.properties

hdfs://bdedev/projects/ews/oozie/iir-ingest-daily/EWS-IIR-Ingest-Inc-daily.xml

**#Raw Table (external table) hdfs location**

hdfs://bdedev/data/lgl\_iir/raw/incdnt\_ivstgn\_raw/\*

**Note:** The files inside the “incdnt\_ivstgn\_raw” and “incdnt\_ivstgn\_trd\_raw” folder need to be deleted before a new/fresh run, to confirm the data population

**Execute an insert query in “bde\_stats.data\_extrcn\_rule” table.**

|  |
| --- |
| INSERT OVERWRITE table bde\_stats.DATA\_EXTRCN\_RULE VALUES('EWS-Drive-Ingest-Inc-daily','gcars',1,'sqoop-dflt',8,'EWS\_Drive\_Ingest\_Inc\_daily.ini','/projects/ews/scripts/spark/search/conf','/etc/security/keytabs/x987731.keytab','x987731@NMCORP.NISSAN.BIZ','N','drive\_sqoop','2019-12-04 00:00:00','drive\_sqoop','2019-12-04 00:00:00'), ('EWS-IIR-Ingest-Inc-daily','iir',1,'sqoop-dflt',9,'EWS\_IIR\_Ingest\_Inc\_daily.ini','/projects/ews/scripts/spark/search/conf','','/etc/security/keytabs/x987731.keytab x987731@NMCORP.NISSAN.BIZ','N','drive\_sqoop','2020-06-01','drive\_sqoop','2020-06-01'); |

**#To run a workflow**

oozie job -oozie http://usnencpl079.nmcorp.nissan.biz:11000/oozie/ -config sqoop\_arch.properties -run

**#To see the status of oozie job (workflow status)**

oozie job -oozie http://usnencpl079.nmcorp.nissan.biz:11000/oozie/ -info 0070011-200311173650044-oozie-oozi-W

**#To kill the oozie job**

oozie job -oozie http://usnencpl079.nmcorp.nissan.biz:11000/oozie/ -kill 0070011-200311173650044-oozie-oozi-W

Once the job is succeeded in oozie, we can confirm it by checking the

1. MapReduce Logs (<http://usnencpl079.nmcorp.nissan.biz:19888/jobhistory/logs/usnencpl125.nmcorp.nissan.biz:45454/container_e96_1583969462830_225169_01_000002/attempt_1583969462830_225169_m_000000_0/x987731/stdout/?start=0>)
2. Success entries of part file in
   1. hdfs://bdedev/data/lgl\_iir/raw/incdnt\_ivstgn\_raw/
   2. hdfs://bdedev/data/lgl\_iir/raw/incdnt\_ivstgn\_trd\_raw/
3. Data’s populated in to the below tables
   1. lgl\_iir.incdnt\_ivstgn\_raw
   2. lgl\_iir.incdnt\_ivstgn\_trd\_raw
   3. lgl\_iir.incdnt\_ivstgn\_base
   4. lgl\_iir.incdnt\_ivstgn\_trd\_base