**Data Integration Tool Run-Book**

1. Create a working directory in the gateway node under your CAD functional account directory ::

Eg: /home/pc081089/di\_tool

1. Goto the below SVN link and unzip the “hdi-tool.zip” in the above location.

(<https://collaborate.bt.com/svn/dss-bigdata/trunk/Data_Integration_Tool/di_tool_runnable>) ::

1. Provide inputs in the configuration.properties file. Keep rest of the file as it is ::

Use the sample\_configuration.properties file in the “hdi-tool.zip” packaged file.

1. Run the command ::

For Importing data from RDBMS :

**java -cp skool.jar:ojdbc6-11.2.0.3.jar:/opt/cloudera/parcels/CDH/jars/\* com.bt.dataintegration.property.config.DIConfigService**

For file imports :

**java -cp skool.jar:/opt/cloudera/parcels/CDH/jars/\* com.bt.dataintegration.property.config.DIFileSystemService**

1. A workspace will be created in HDFS in the following directory:

/user/<instance-name>/workspace/HDI\_<source\_name>\_<database\_name>\_<table\_name>

This directory contains the following file ::

1. HDI\_CREATE\_AUDIT\_TABLE.hql

Hive script to create the AUDIT table. Contents of the table can be viewed by querying over the table in Hive.

1. HDI\_<TABLE\_NAME>.avsc

Avro schema file which will be referred by HDI\_<CREATE\_AVRO\_TABLE>.hql to create the table.

1. HDI\_<CREATE\_AVRO\_TABLE>.hql

Hive script to create AVRO table.

1. HDI\_<TABLE\_NAME>\_ADD\_PARTITION.hql

Hive script to add partition to the AVRO table.

Partitions (year,month,day,hour,minute)

1. HDI\_<TABLE\_NAME>\_COMPRESS\_DATA.pig

Pig script to compress AVRO file.

1. audit\_logs.sh

Shell script to capture audit details.

1. housekeep.sh

Shell script to delete those files which fall over the retention period (retention period taken from configuration.properties)

1. job.properties

This will be a sample for the user. The original one will be same but will be invoked from Gateway-Node

1. ojdbc6-11.2.0.3.jar

Will be used by Oozie to run sqoop action.

1. refresh\_last\_col\_value.sh

Shell script to capture date which will be used by sqoop in next incremental import.

1. unix\_date.txt

Will contain captured date from refresh\_last\_col\_value.sh

1. update\_last\_col\_value.sh

Shell script to finally append the values in unix\_date.txt after pig action finishes.

1. workflow.xml
2. coordinator.xml (if coordinator\_required is true in configuration.properties)

All of the above files will be used by Oozie.

Note: In case of file import the directory structure will be as –

/user/<instance-name>/workspace/HDI\_FILE\_<source\_name>\_<table\_name>

1. A directory with the user specified table name will be created in the current directory inside which the job.properties along with other necessary files will be present.
2. Next you can submit the oozie workflow by running the following command(1A cluster)

from the above created directory::

**/> cd <table\_name>**

**oozie job -oozie http://tplhc01c001.iuser.iroot.adidom.com:11000/oozie -config job.properties -run**

NOTE: If encountered any errors while running the tool Ex: Failed on local exception: java.io.IOException: javax.security.sasl.SaslException: GSS initiate failed [Caused by GSSException: No valid credentials provided (Mechanism level: Failed to find any Kerberos tgt)]; Host Details : local host is: "tplhc01g001.iuser.iroot.adidom.com/10.9.65.4"; destination host is: "tplhc01c001.iuser.iroot.adidom.com":8020;

Run the following –

**kinit**

and provide your CAD functional account password. Then proceed from Step-4