USER MANUAL IIF

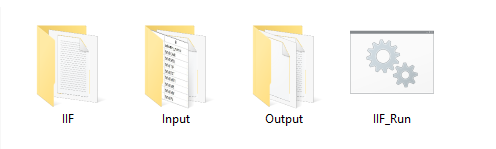
# Introduction:

IIF (Integrated Ingestion Framework) is aimed at automating the entire process of ingestion. It performs the following tasks:

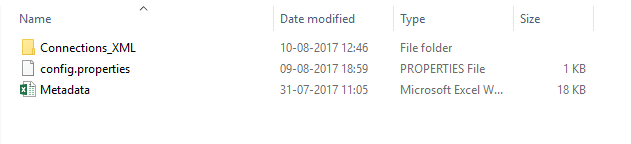
* Validation of metadata provided by user
* Creation of DDLs for the tables at different zones
* Creation of ingestion streams and count check streams
* Generation of release tar (including the Installer script and Tidal script) in the required folder structure format
* Moving and executing the DDLs and streams in the EDL
* Unit Testing: Performs count check and generates reports in .csv files corresponding to success (count is matching), failure (count mismatch) and missing (tables found in metadata but not executed)

# Guidelines for using the IIF:

1. The folder structure of the IIF is as shown below:



1. The INPUT folder structure is as shown below:



**a) Connection XMLs:**

The user has to provide the source and target connection xmls in the Connection xmls folder.

**b) Configuration properties:**

**Project Inputs:**

1. Input\_Metadata\_Name: This holds the name of the metadata sheet. The metadata sheet and the config.properties file are both kept in the “Input” folder.
2. SRC\_SYS: Indicates the type of source system used i.e. flat file, DB2, etc. If it is a flat file the value should be assigned as “files”.
3. Prj\_EPM\_Code: Assign this variable with the epm code for the particular project (Ex: BBCY for CEDl, etc.)
4. SRC\_ID: Holds the value that must be assigned to the source\_id column of the tables in that source system. (Ex: B85R for CR-SP, BC5S for TSYS, etc.)
5. Country\_Code: The ISO country code.

**Target Table Inputs:**

**Tgt Connection Inputs:**

1. Target\_Connection\_XML\_Path: Holds the path where the target connection xml is present-“Input\Connections\_XML\<target\_connection\_xml\_name.xml>”
2. TGT\_Schema\_Name: schema name of target.

**Database Inputs:**

1. Staging\_Zone\_DB: Database in which staging zone scripts should be executed.
2. Technical\_Zone\_DB: Database in which technical zone scripts should be executed.
3. Enterprise\_Zone\_DB: Database in which enterprise zone scripts should be executed.
4. Consumer\_Zone\_DB: Database in which production zone scripts should be executed.

**Naming Convention Inputs:**

1. Staging\_Zone\_Name: Prefix given to staging zone tables (Ex: stg\_)
2. Technical\_Zone\_Name: Prefix given to technical zone tables (Ex: tech\_, tsz\_, etc.)
3. Enterprise\_Zone\_Name: Prefix given to enterprise zone views (Ex: ent\_, edz\_, etc.)
4. Consumer\_Zone\_Name: Prefix given to production zone views/tables (Ex: pz\_)

**Target Tables Location Inputs:**

1. HDFS\_Location: Location for the hive tables.

(Ex: /user/dev\_bbcv\_etl/data/tz/b6xp/cedl/bnspa/pub)

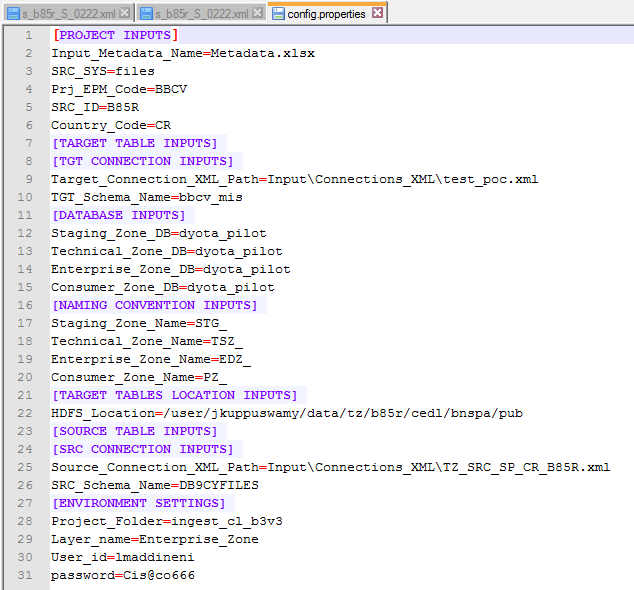
**Source Table Inputs:**

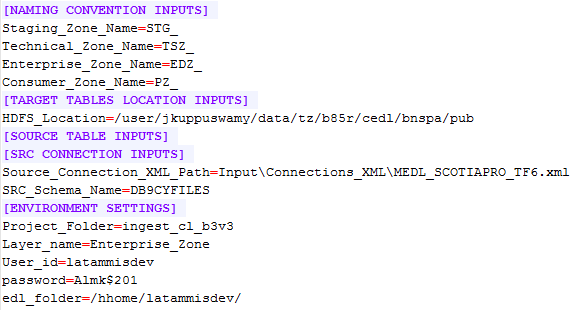
**Src Connection Inputs:**

1. Source\_Connection\_XML\_Path: Holds the path where the source connection xml is present-“Input\Connections\_XML\<source\_connection\_xml\_name.xml>”
2. SRC\_Schema\_Name: Holds the source schema name (Ex: TC9CYFILES, DBO, etc.)

**Environment Settings:**

1. Project\_Folder: The folder where you want your stream xmls to be imported.
2. Layer\_name: The layer under the project folder where you want your streams to be imported.
3. User\_id: Give any Diyotta id, which has access to your required folder.
4. password: Password of the Diyotta id given by you.
5. edl\_folder: The server folder path, where you want to move your local outputs.

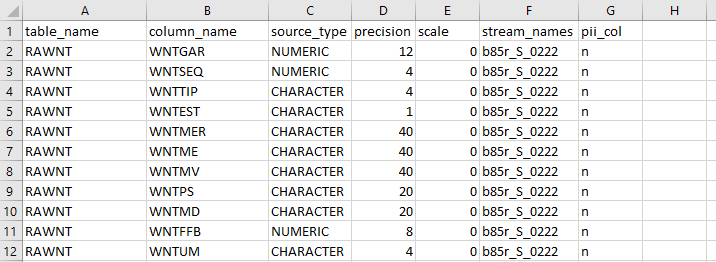
****Following is a snippet of the file:



**c) Metadata sheet:**

The format of metadata sheet should be as shown below:

**Kindly do not change the values of pii\_col to “y” as the feature is not enabled and might result into garbage outputs or errors**

****



***Note:*** *The sample query for extracting the metadata from DB2 system is given below:*

select \* from sysibm.columns where table\_schema in ('DC9CYFILES') and table\_name in ('ACMST,'RAGAR');

**Table\_name:** Holds the source table name

**Column\_name:** The column names of the corresponding tables are given against them

**Type:** The source datatype of the corresponding columns given against them

**Precision:** The precision for a column as given in source

**Scale:** The scale for a column as given in source

**Stream\_names:** The name of the stream in which a particular table has to be present is given against them in this column in the following format:

<epm\_code>\_<table\_type>\_<index>

table\_type: S if small, M if medium, L if large, XL if extra large

index: a number

**pii\_col:** This column is kept as a placeholder for the future implementation of pii logic. As of now, this column’s value should be ‘n’.

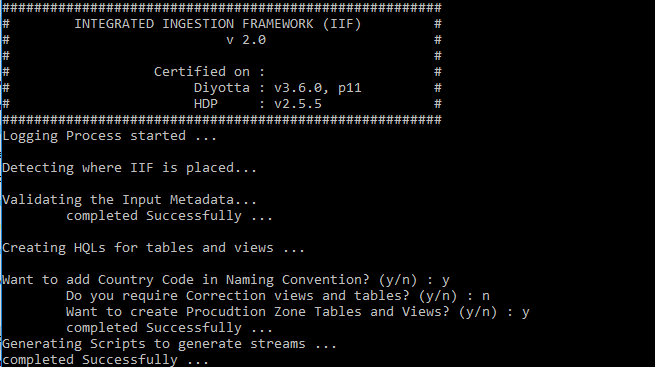
1. **Execution:** In the IIF folder there is a .batch file **IIF\_RUN.** After providing the necessary inputs as explained in the previous section, double click this .batch file.

**🡪**The framework is triggered and it proceeds to **metadata validation.**

In the following cases, the metadata sheet is invalid:

1. blank spaces in the data area
2. if there are duplicate columns for a table

**🡪**After metadata validation, it proceeds to **DDL generation**. Give necessary inputs(y/n) according to the requirement.



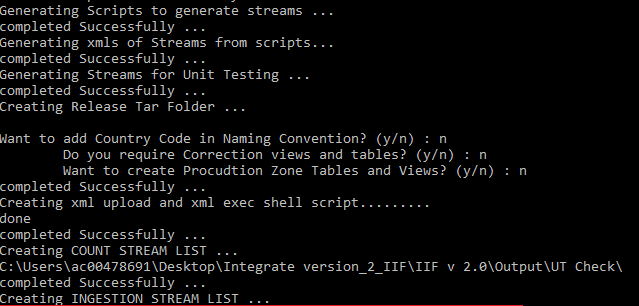
Ex: If you want country code in your table names then give ‘y’ for the question “*Want to add Country Code in Naming Convention?*”

Similarly give input for correction tables and production table/views according to the requirement. The generated DDLs will be present in the path “Output\HQLs\”.

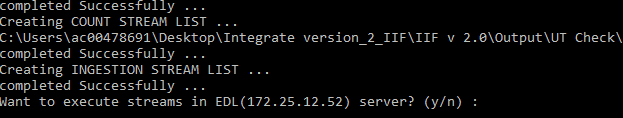
**🡪** Next, the **data ingestion stream xmls** will be generated. The user can find the xmls in the “Output\STREAM\_XMLs\” folder.

After this, the **count check stream xmls** and the **Release Tar folder** are created.

The **xml upload** and **xml execution shell scripts**, **count stream list** and **ingestion stream list** are created next.

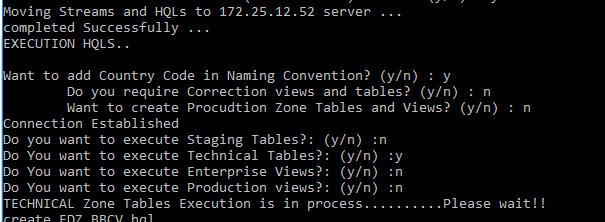


**🡪**After this, the user is asked, “*Want to execute streams in EDL (172.25.12.52) server? (y/n):*”



Once you give ‘y’ the DDLs, xmls, release tar and the shell scripts are moved to the server location, which is given as ‘/home/gsingh9/Demo\_Purpose\_Folder’.

**🡪**Next, the user is asked if he wants to execute the created DDLs in hive. Give appropriate y/n answers.



**🡪**After the HQL execution, the stream xmls are uploaded to Diyotta using the generated xml\_upload script.

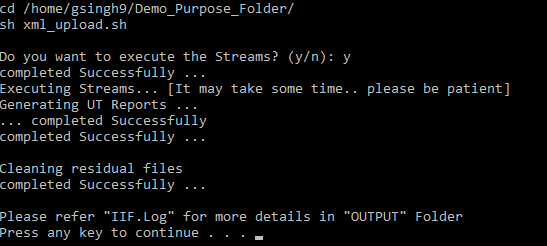


***DISCLAIMER:***

*Under following circumstances, the tool is not responsible for Connection Lock and any other errors or Faults:*

1. IF the connections are not properly given in Input files:
2. If connection names are different, and the streams are executed.
3. If the source connection given is not tested after importing it into Diyotta.

**[Kindly change the password and test the connection, or else it may result into locking the connection]**

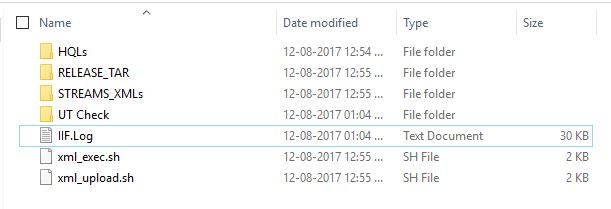


**🡪** The user will be asked, “*Do you want to execute the Streams? (y/n):*”

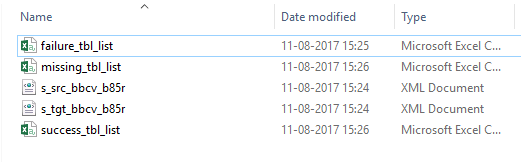
Give ‘y’ if you want to run the streams.

**🡪** After the execution of streams, reports are generated in .csv files.

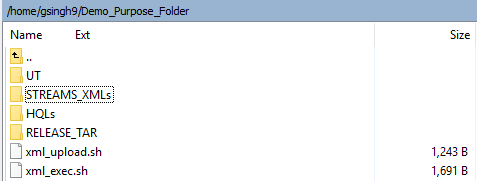
**Output:**

The output folder looks as given below:

1. HQLs folder has the DDLs of tables at different zones.
2. It also has the Release Tar in the required format. This folder has the DDls, the ingestion stream xmls, the installer script and the Tidal script moved to their required folders.
3. STREAMS\_XMLs has the data ingestion stream xmls, the stream list and the env\_setting.proprties file.
4. UT Check has the source and target count check stream xmls and the Unit Testing reports. Success\_tbl\_list has details regarding tables where the source and target counts are matching. Failure\_tbl\_list has details of tables which have count mismatch and Missing\_tbl\_list has details of those tables that were in metadata sheet but were not executed.



1. For details of execution, the user can refer the log file IIF.log.
2. The Output will be moved to EDL and it will be present in the path in the screenshot given below:



**Sample Metadata Sheet:**



# Related Documents

### Release Document v1.0:

