**Adhoc File to Table Load**

**Problem Statement:**

There has been a lot of requests from BI Team to load data into tables on an Adhoc basis for doing analysis and other purposes. To achieve this, we had to spend a lot of time in creation of mapping, workflows, creating metadata entries migrating the code dependent on the nature of request. These kinds of requests used to take a lot of time and used be very frequently changed.

**Solution to our problem:**

To overcome this, we have created an idea to skip informatica for some dirty-quick loads request from Business’ end. The load will be trigged by calling a python script which in turn will read a STTM document to find the mappings between the source file and the table to be loaded. We also pass source file name, target table name and STTM document name as arguments while calling the script.

**What script does:**

The script reads the Source data file, and STTM document, relates which column from Source file needs to be loaded to which columns in Target table. This is done by the mapping provided by the user in STTM Document. If no name is provided for a column we will insert NULL for these types of columns.

Then the script truncates the Target table and generate an insert statement for the data available in Source data file. Then the script executes the insert statement on the database. At last it commit’s the data.

**Example to call the script:**

python Adhoc\_file\_to\_table\_v1.py ‘Source\_file.csv’ ‘schema\_name.table\_name’ ‘STTM\_Document\_name.csv’

**How to use it:**

1. Get your Source file (in CSV format), Python code, STTM Document in same location.
2. Make sure you have Anaconda 3.5 and above installed.
3. Make sure you have 2 columns in STTM Document **Table\_Column\_Name** and **Source\_Column\_Name.**
4. Make sure that you mention the column name from table you wish to insert data.
5. Make sure mentioned column names are correct and there are no naming issues.
6. Make sure column name mentioned in Source data.csv file is same that you mention in STTM document.

**Assumptions for current version:**

* Source file data and STTM Document should be available in CSV format (comma seperated)
* The table name mentioned while calling the script is mentioned with Schema name.
* All Column names mentioned in the STTM Document should be correct.
* All columns mentioned under columns **Source\_Column\_Name** should be available in Source data file.
* Table to have Fiscal\_period,Row\_update\_timestamp,Row\_Create\_timestamp, Extract\_ID, Source\_row\_seq\_number column.
* Data will only be inserted to Stage table.

**Future Scope:**

* Validation to be done on Column name mentioned in STTM Document.
* Validation to be done on number of rows to be inserted from Source data file and number of rows actually inserted in table.
* Validation if table exists in given schema.
* More separators for source file, (Delimiter definition).
* Data type definition to be given by user, will help in data insertion in Daily, Weekly, etc. tables.
* User defined values for specific columns.
* Metadata Entries to be maintained.
* Database’s configuration to be parameterized and on same security level is achieved as in other PDW scripts.