TD\_ETL\_IDS\_ANONYMOUS\_DATA

Version: 0.3

Contents

[2 Document Control 5](#_Toc436814238)

[2.1 Product History 5](#_Toc436814239)

[2.2 Reviewers 5](#_Toc436814240)

[2.3 Approvers 5](#_Toc436814241)

[3 Overview 6](#_Toc436814242)

[3.1 Document Background 6](#_Toc436814243)

[3.2 Definitions 6](#_Toc436814244)

[3.3 Document References 6](#_Toc436814245)

[3.4 Assumptions 7](#_Toc436814246)

[3.5 Dependencies and Constraints 7](#_Toc436814247)

[4 Requirements 8](#_Toc436814248)

[4.1 Functional Requirements 8](#_Toc436814249)

[5 Technical Specifications 9](#_Toc436814250)

[5.1 15/15 Rule 9](#_Toc436814251)

[5.2 High-Level Data Flow 10](#_Toc436814252)

[5.3 Frequency and Data Volume 13](#_Toc436814253)

[5.4 Data Model/Mapping 14](#_Toc436814254)

[6 ETL Specification 15](#_Toc436814255)

[6.1 ETL Workflows/Mappings 15](#_Toc436814256)

[6.1.1 **wf\_CDM\_TO\_STG\_ANONYMOUS\_DATA\_PASS** 15](#_Toc436814257)

[6.1.2 **wf\_STG\_ANONYMOUS\_DATA\_PASS1** 16](#_Toc436814258)

[6.1.3 **wf\_STG\_ANONYMOUS\_DATA\_PASS2** 18](#_Toc436814259)

[6.1.4 **wf\_CDM\_ANONYMOUS\_DATA\_EXTRACT** 18](#_Toc436814260)

[6.1.5 **wf\_ANONYMOUS\_DATA\_SCAN\_FAILED** 20](#_Toc436814261)

[6.2 ETL Reusable Objects 21](#_Toc436814262)

[6.3 File Creation Process 22](#_Toc436814263)

[6.4 File Transfer Process 23](#_Toc436814264)

[6.5 Unix Scripts 23](#_Toc436814265)

[7 Job Scheduling 24](#_Toc436814266)

[7.1 Scheduling configuration 24](#_Toc436814267)

[7.2 Performance Considerations 24](#_Toc436814268)

[7.3 Batch Job Frequency & Data Volume 24](#_Toc436814269)

[7.4 Error Handling 25](#_Toc436814270)

[7.5 Archiving & Re-Run Process: 25](#_Toc436814271)

[7.6 Scheduler Process Flow & Job Name 25](#_Toc436814272)

[7.7 Implementation Considerations 25](#_Toc436814273)

[8 Unit Testing Criteria 27](#_Toc436814274)

[8.1 Test Plan 27](#_Toc436814275)

[9 Appendices 28](#_Toc436814276)

[9.1 Appendix A – Key Terms 28](#_Toc436814277)

[9.2 Appendix B – Outstanding Items 28](#_Toc436814278)

# Document Control

## Product History

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Author(s) | Change Description |
| 0.1 | 10/12/2015 | Manoj Moorthy | Initial Draft |
| 0.2 | 11/05/2015 | Manoj Moorthy | Updated based on new requirement to capture Accont level Exclusion reason in the summary table.  Also changed History Load approach to be able to execute from Production Environment. |
| 0.3 | 12/01/2015 | Manoj Moorthy | Updated based on new/changed requirements |

## Reviewers

|  |  |  |  |
| --- | --- | --- | --- |
| Role | Name | Version | Review Date |
| Build Order Buddy | Dambaru Jena |  |  |
| IT Analyst Lead | Deepan Patel |  |  |
| IT Analyst Lead | David Blackmore |  |  |
| IT Analyst | Travon Lockett |  |  |
| EDE Project Manager | Bob Walters |  |  |
| EDE Project Manager | Mike Abts |  |  |
| Functional Designer | Alberto Riego |  |  |
| CDW Maintenance | Brian Paasch |  |  |
| MW App Lead | Svitlana Dukhovna |  |  |

## Approvers

|  |  |  |  |
| --- | --- | --- | --- |
| Role | Name | Version | Approved Date |
| Build Order Owner | Silvia Guevara |  |  |
| Lead Architect | Deepak Kulkarni / Gregg Snapp |  |  |
| IDS Domain Owner | Dambaru Jena |  |  |
| IDS App Lead | Abhishek Sood |  |  |
| Technical Architect Lead | Mike Sacco/Mike Nguyen |  |  |
| AMI Mantainance | Haitao Sam Yu |  |  |

# Overview

## Document Background

Per the ICC docket, anonymized interval usage data must be provided to any external entity that requests it. The EDE Anonymous Data Build Order will enable the functionality to fulfill anonymous interval data usage requests. IDS is creating an automated solution by generating the monthly delta files that will be scheduled to be extracted every month.

This technical design document explains the process by which the Interval Data Services (IDS) will anonymize interval data for all accounts that have at least one AMI meter using the 15/15 rule*.*

IDS will generate anonymous data files per zip code containing all accounts that pass the 15/15 rule in a delivery class at the monthly level. This process will be executed once a month. IDS will extract the file and make it available in an uncompressed form. All activities involving compressing and transferring the files (ftp) will be done be a different team.

## Definitions

|  |  |
| --- | --- |
| Term | Definition |
| ETL | Extract, Transform, and Load (Data movement process between tables) |
| IDS | A database which consists of centralized data from one or more sources and from which data is taken for reporting purposes. |
| CDW | Customer Data Warehouse |
| Best Available Interval Data | Usage data that has not been validated/estimated/edited for billing, and can therefore not be used for billing. |
| Common Data Mart (CDM) | A schema within IDS that contains a set of physical database tables/views to standardize data exchange from IDS to external and internal customers. |

## Document References

|  |  |
| --- | --- |
| Name | Location |
| FD\_IDS\_ANONYMOUS DATA | <http://teamapps.exeloncorp.com/sites/ITComEdAMI/Release%20Agnostic%20Design%20Repository/IDS/IDS%20Functional%20Designs/FD_IDS_ANONYMOUS%20DATA.docx> |

## Assumptions

|  |  |
| --- | --- |
| No. | Description |
| No. | Description |
|  | There are no more than 600 zip codes in the ComEd territory with AMI Meters |
|  | Anonymization process can be executed off hours |
|  | File will be compressed at the source by Unix Shell Script by a different team and not by IDS |
|  | Interval Data available in the IDS will be provided |
|  | Best available data meter product (e.g. 30 min intervals) |
|  | Data prior to August 2014 is unavailable in the source system and therefore cannot be provided |

## Dependencies and Constraints

The Anonymous Data Extract process will be dependent on the following:

|  |  |
| --- | --- |
| Area | Dependency Description |
| Data Integration | CDM.TF\_BEST\_AVAILABLE\_INTERVAL incremental data loading job should be completed and CDW dimension tables having upto date. IDS will be extracting the monthly delta files once every month. |
| Tech Arch | This interface build, test, and deployment activities are dependent upon the support of the Tech Arch team for setting up and supporting environments required for these efforts. |

# Requirements

## Functional Requirements

| Req # | Requirement Description |
| --- | --- |
| FR\_001 | IDS must provide the ability to execute the job to create anonymous data monthly |
| FR\_002 | IDS must apply the 15/15 rule to interval data per delivery class |
| FR\_003 | IDS must exclude accounts with net meters |
| FR\_004 | IDS must exclude channel 2 data (energy sent back from customer) for all meters |
| FR\_005 | IDS must give a random identifier to each account included in the anonymous data file |
| FR\_006 | Random identifiers must be retained month to month |
| FR\_007 | IDS must use the delivery class linked to an account at the end of the month for which the anonymous data file is being generated |
| FR\_008 | IDS must aggregate usage for accounts with multiple meters |
| FR\_009 | IDS must generate delta files per month per zip code with anonymous data for all accounts within a delivery class that pass the 15/15 rule |
| FR\_010 | IDS must provide a summary table of zip codes that passed 15/15 rule each month |
| FR\_011 | Anonymous data file output must be in CSV format specified in 6.1 Output File Format |
| FR\_012 | The file should be generated within 10 days after the job for anonymous data kicks off |
| FR\_013 | The compressed files will always have the same name and the links will always be static (no archival)  This FR is not in scope for IDS and will be handles by a different team. |
| FR\_014 | Compressed files must be in zip format  This FR is not in scope for IDS and will be handles by a different team. |
| FR\_015 | Source uncompressed files shall remain at the source until the next month’s file generation |
| FR\_016 | IDS must exclude new meters set on the month being analyzed |
| FR\_017 | IDS must capture the exclusion reason for any accounts that fail the 15/15 rule in an account level summary table |
| FR\_018 | IDS must extract summary table with zip codes and delivery classes that passed the 15/15 rule in csv format each month |
| FR\_019 | IDS must extract account level summary table with exclusion reasons in csv format each month |

# Technical Specifications

## 15/15 Rule

The 15/15 rule consists of two parts:

• For a given month, there must be at least 15 accounts per delivery class per zip code

• Out of the 15 accounts, not one account can represent 15% or more of the monthly usage of the delivery class for up to 24 months

Below is a flow diagram with the functional anonymization logic:



## High-Level Data Flow

This section details about the high level data flow of the Anonymous data extract process.

**Initial Stage Load process:** Data from CDM.TF\_BEST\_AVAILABLE\_INTERVAL where CHANNEL\_NBR = 1 (CHANNEL\_NBR = 101 is part of Channel 1 in TF\_BEST\_AVAILABLE\_INTERVAL) is joined to CDW tables to get a initial data set that pass all applicable joins and filters. Data will be aggregated to calculate sum of monthly usage at 2 different levels. One level is Month, Zipcode and Delivery Class and the other is Month, Zipcode, Delivery Class and Account. Here we also Load into Summary table with all distinct list of Zipcode, Delivery Class and Account for the processing month.During this first time load into Summary table, both the DC level Status flag and Account Level Status flag is set to ‘PASSED’



**Looping Process:** A looping process will be built to recalculate the usage and re-apply the 15/15 rule until the final list of account that passed the 15/15 rule is ready.

**Note**: IDS process will round the percentage value to the next higher integer.

Ex: 14.0 will be treated as 14 and 14.01 to 14.99 will be treated as 15.

So, any Accounts with 14.0 percent or less of usage after all the looping are considered passed and all Accounts with 14.01 percent or more of usage after all the looping are considered failed.



**Extraction Process:** Data from CDM.TF\_BEST\_AVAILABLE\_INTERVAL where CHANNEL\_NBR = 1 is joined to CDW tables and the list of Account that passed the 15/15 to extract files by zipcode for a given month. If there is any data updated in TF\_BEST\_AVAILABLE\_INTERVAL during the time the list of Accounts that pass the 15/15 rule is being generated for a given month, the changed data will not be considered for creating the list and hence will not refles in extracting the files.



**History Process:**

The same Informatica objects used in the incremental load will be used to also populate the History data. This will be done by having separate shell scripts that will read a month ID from a Parameter file created for the History load. This Month ID (YYYYMM) parameter will be passed to the script and then the entire process of Anonymous Data including extraction of files will be done for that month.

This Initial History load of data will executed from Production repository by adding JSS jobs to an on-demand schedule. The History data will be populated one month at a time.

## Frequency and Data Volume

Anonymous data monthly delta files will be created every month following go live which will contain the previous month’s anonymous interval data. The History job will start generating files from August 2014, one month at a time.

Currently there are 572 zip codes in the ComEd territory. 378 of those zip codes have at least one account with an AMI meter. Based on the number of zip codes, no more than 600 files will be created per month, each file will contain the previous month’s anonymous interval data.

Projected AMI Meters in ComEd Territory

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year End Projection | AMI Meters Installed | File Size per Meter (KB) | Total File Size (MB) | Total File Size (GB) | Total File Size (TB) | Total Compressed File Size (GB) |
| 2015 | 1,724,361 | 600 | 1034616.6 | 1034.62 | 1.03 | 51.73 |
| 2016 | 2,654,361 | 600 | 1592616.6 | 1592.62 | 1.59 | 79.63 |
| 2017 | 3,584,361 | 600 | 2150616.6 | 2150.62 | 2.15 | 107.53 |
| 2018 | 4,157,000 | 600 | 2494200 | 2494.2 | 2.49 | 124.71 |

## Data Model/Mapping

Table structures:



Data Mapping for the final Extract files:

<http://teamapps.exeloncorp.com/sites/ITComEdAMI/CIDS/02.%20Data%20Architecture/03.%20Design/IDS_ANONYMOUS_Extract%20files%20DATA%20MAPPING.xlsx>.

The Anonymous Data Extracts Files will include the columns below:

* Zip Code
* Delivery Service Class
* Delivery Service Name
* Account Identifier
* Interval Reading Date
* Interval Length
* Total Registered Energy
* Interval HR 0030 to Interval HR 2500

The naming standard for the files will be as follows

ANONYMOUS\_DATA\_<MONTHID in YYYYMM format>\_<ZIPCODE>.csv

Ex: ANONYMOUS\_DATA\_201510\_60606.csv

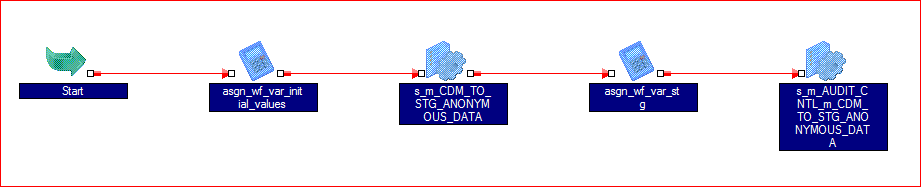
# ETL Specification

## ETL Workflows/Mappings

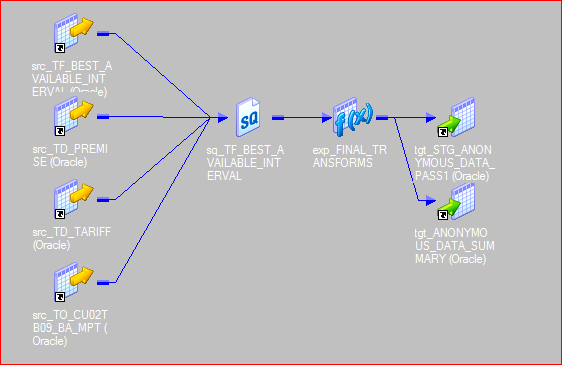
All Informatica objects for this project will be in the project folder ‘**CAMI\_DATA\_HUB**’.

### **wf\_CDM\_TO\_STG\_ANONYMOUS\_DATA\_PASS**

**CDM to Staging workflow** – First workflow that Joins CDM BA table (only Channel1) data with CDW table and stages the data for processing. Using this separate process is to make sure that we only read from the CDM and CDW tables once until we get to the extract part.

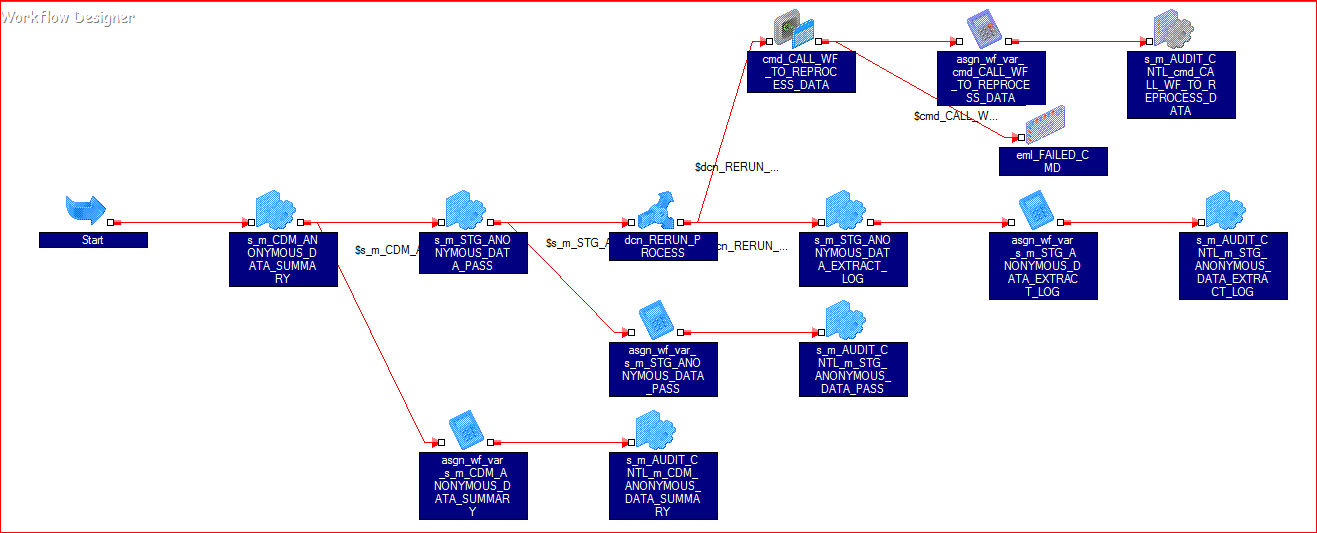


**m\_CDM\_TO\_STG\_ANONYMOUS\_DATA\_PASS**: Mapping to read from CDM and CDW tables and loads into the first Staging table. The Filter to get only Channel 1 records and Aggregation of the usage will be implemented in SQL Override of the mapping. We will load the distinct list of Zipcodes, Delivery Class and Accounts here for the first time with the default status as ‘PASSED’ for the first time. Here we will use the formula “round((BILL\_ACCOUNT\_NBR \* 3)/2)” to generate a random identifier and store it in the Summary table.

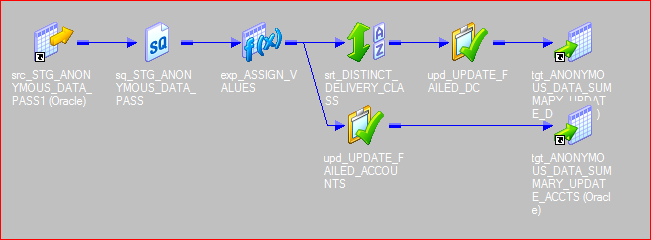


### **wf\_STG\_ANONYMOUS\_DATA\_PASS1**

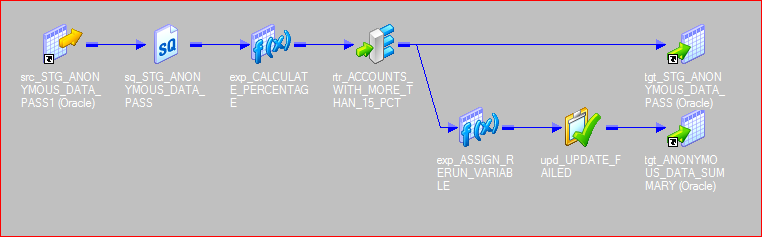
This workflow will read from STG\_ANONYMOUS\_DATA\_PASS1 and load into STG\_ANONYMOUS\_DATA\_PASS2. It will also have a decision task to either reprocess the data using the workflow wf\_STG\_ANONYMOUS\_DATA\_PASS2 (called using command task) to apply 15/15 rule or load the Extract log table and initiate the Extract process. After loading the Extract Log table, a post-session success command is used to create an event file. This is the file that the Extract process is waiting for to start.



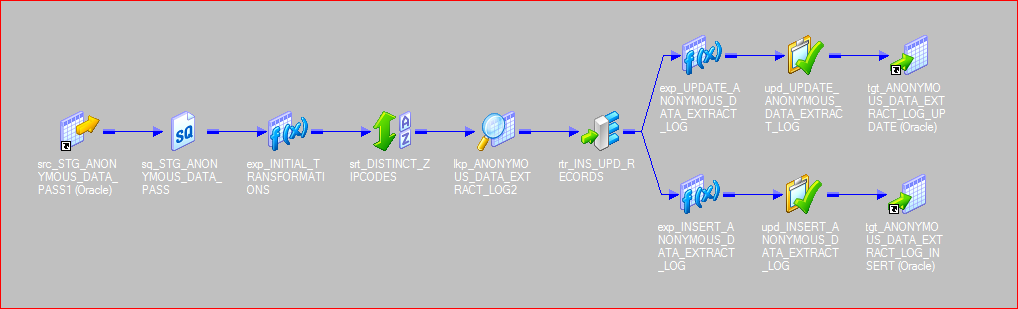
**m\_CDM\_ANONYMOUS\_DATA\_SUMMARY**: Mapping to update the delivery class status and account status to 'FAILED' in the ANONYMOUS\_DATA\_SUMMARY table when first 15-15 rule is not satisfied.



**m\_STG\_ANONYMOUS\_DATA\_PASS**: This will be reused in different workflows to apply a loop to get the final list of Accounts that passed the 15/15 rules. The Aggregation of the usage and part of the 15/15 rule will be implemented in SQL Override of the mapping. The other part of the 15/15 rule will be implemented within the Informatica code in the mapping. This mapping will also be used to update the account status to 'FAILED' in the ANONYMOUS\_DATA\_SUMMARY table when second 15-15 rule is not satisfied.



**m\_STG\_ANONYMOUS\_DATA\_EXTRACT\_LOG**: This mapping is run the final iteration of the 15/15 process is complete. This is used to load into the CDM.ANONYMOUS\_DATA\_EXTRACT\_LOG table. Here the table CDM.ANONYMOUS\_DATA\_EXTRACT\_LOG will be Update else Insert to tract the file extract history for all files for each month and Zipcode combination.



### **wf\_STG\_ANONYMOUS\_DATA\_PASS2**

This workflow is a copy of the PASS1 workflow. Here the process will read from STG\_ANONYMOUS\_DATA\_PASS2 and load into STG\_ANONYMOUS\_DATA\_PASS1. It will also have a decision task to either reprocess the data using the workflow wf\_STG\_ANONYMOUS\_DATA\_PASS1 to apply 15/15 rule or load the Extract log table and initiate the Extract process.

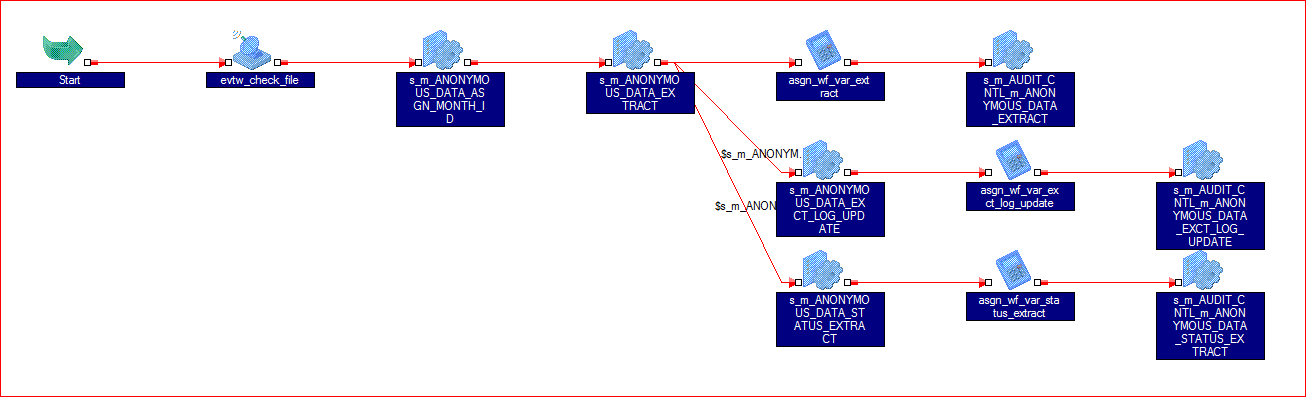
**m\_CDM\_ANONYMOUS\_DATA\_SUMMARY**: The same mapping used in PASS1 will be reused by changing the source and target names at the workflow level.

**m\_STG\_ANONYMOUS\_DATA\_PASS**: The same mapping used in PASS1 will be reused by changing the source and target names at the workflow level.

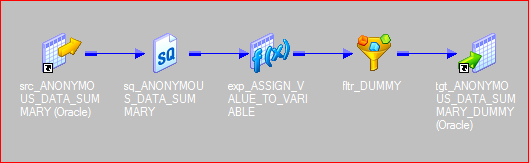
**m\_STG\_ANONYMOUS\_DATA\_EXTRACT\_LOG**: The same mapping used in PASS1 will be reused by changing the source and target names at the workflow level.

### **wf\_CDM\_ANONYMOUS\_DATA\_EXTRACT**

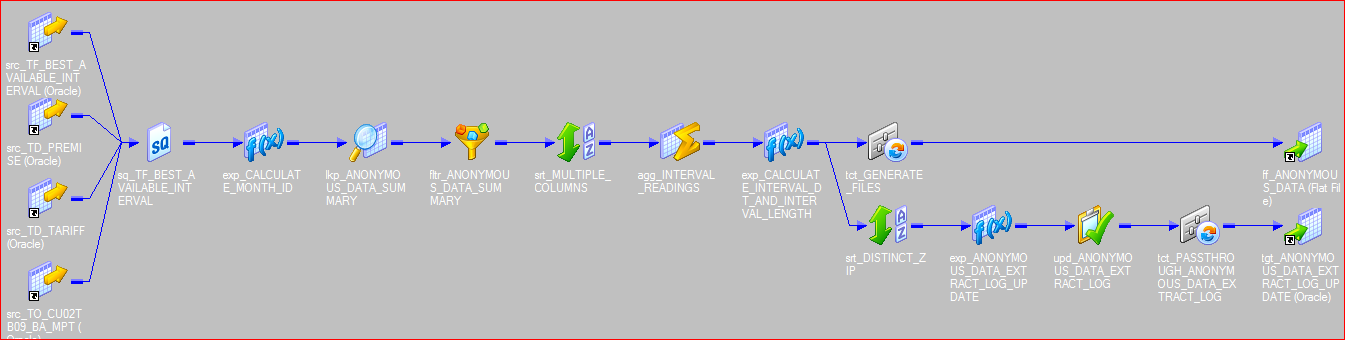
This workflow will be used to extract the variout .csv files for Anonymous data. Here the first task is the event wait task that waits for the looping process to be completed.



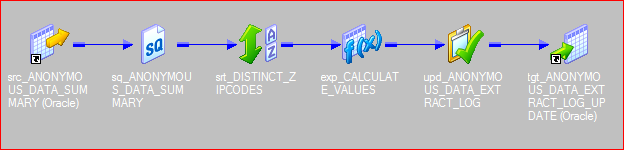
**m\_ANONYMOUS\_DATA\_ASGN\_MONTH\_ID**: Mapping to get the EXTRACT\_MONTH\_ID from the ANONYMOUS\_DATA\_SUMMARY table and assign it to Mapping Variable which is then copied over to the workflow variable. This doesn’t load any data it is used only to assign a value to the MonthID variable.



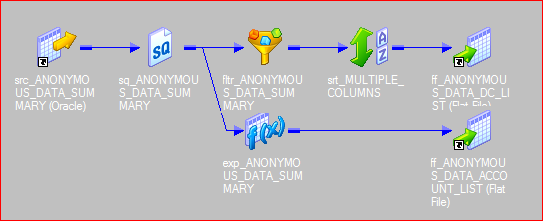
**m\_ANONYMOUS\_DATA\_EXTRACT**: This mapping is used to create the extract files after the final list of valid accounts that passed the 15/15 rule have been identified. Here we will make use of the Transaction Control transformation to dynamically split the target extract files based on Zipcodes. This mapping will also update the Extract status column to 'COMPLETED' in the CDM.ANONYMOUS\_DATA\_EXTRACT\_LOG table.



**m\_ANONYMOUS\_DATA\_EXCT\_LOG\_UPDATE**: This mapping is used to update the CDM.ANONYMOUS\_DATA\_EXTRACT\_LOG table with the 'FAILED' status flag in case the extract file session fails. This is a all or nothing approach where we will update all the records for the given processing month, Zipcode and Delivery Class combination.



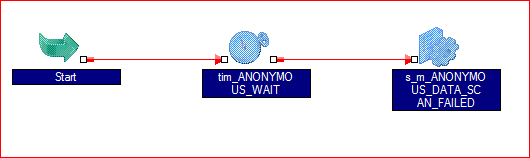
**m\_ANONYMOUS\_DATA\_STATUS\_EXTRACT**: This mapping is used to extract a csv file with the list of accounts for the processing month. Here the list is for both Passed and Failed account along with the reason for failure. This mapping is also used to extract a csv file with the list of delivery class for the processing month that passed the 15/15 rule.



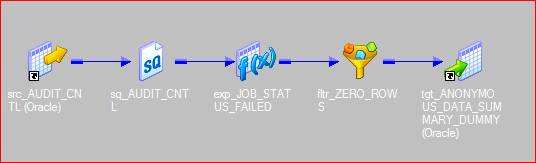
### **wf\_ANONYMOUS\_DATA\_SCAN\_FAILED**

This workflow is used to scan the AUDIT\_CNTL table to check if the last run of the workflows wf\_STG\_ANONYMOUS\_DATA\_PASS1 or wf\_STG\_ANONYMOUS\_DATA\_PASS2 failed during execution. If any one of those workflows are in failed state, then this workflow is forcedto fail which inturn fails JSS job. This way the maintainance team is notified that something is wrong.

The first task is a timer task setup to wait for 3 hours before proceeding.



**m\_ANONYMOUS\_DATA\_SCAN\_FAILED**: This Mapping is used to scan the IHUB.AUDIT\_CNTL to check if the last run of the workflows wf\_STG\_ANONYMOUS\_DATA\_PASS1 or wf\_STG\_ANONYMOUS\_DATA\_PASS2 failed during execution. If any one of those workflows are in failed state, then this workflow is forcedto fail which inturn fails JSS job. This way the maintainance team is notified that something is wrong.

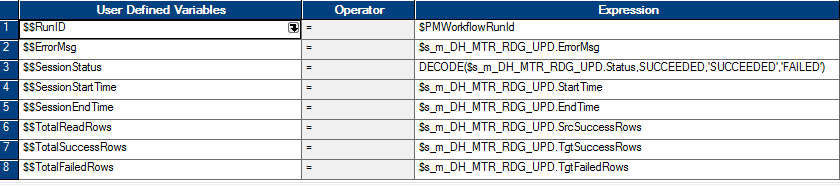


## ETL Reusable Objects

All ETL jobs execution statistics such as number of records processed/ rejected and duration will capture in IHUB.AUDIT\_CNTL table for audit purpose using below Informatica re-usable objects/ tasks.

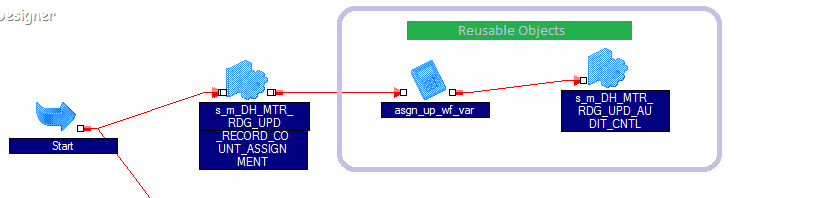
**Assignment Task:** ASGN\_WF\_VAR

This assignment task will capture below job stats and assign it to corresponding workflow variables after completion of every main business logic session in a workflow.



**s\_XXXXX\_AUDIT\_CNTL**

This session will be very last job in every workflow to process above workflow variables into the audit transcation table IHUB.AUDIT\_CNTL.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Column Name** | **Data type** | **Stage Column Name** | **Transform / Rules** | **Maps To Target Column** | **Data type** |
|
|  |  | NONE | RULE: Generated by INFA | BTCH\_ID | INTEGER |
|  | TEXT | NONE | RULE: Passed in via INFA variables | JOB\_NM | VARCHAR2(50) |
|  |  | NONE | RULE: Generated by INFA | REC\_PROC\_IN\_CNT | INTEGER |
| SYS\_TYPES.SRC\_TYPES | VARCHAR2(50) | NONE | RULE: Decode from SRC values | SRCE\_SYS\_NM | VARCHAR2(50) |
| SYS\_TYPES.TGT\_TYPES | VARCHAR2(50) | NONE | RULE: Decode from TGT values | TGT\_SYS\_NM | VARCHAR2(50) |
|  |  | NONE | RULE: Generated by INFA | RE\_PROC\_OUT\_CNT | INTEGER |
|  |  | NONE | RULE: Current System Date | BTCH\_START\_DTTM | TIMESTAMP |
|  |  | NONE | RULE: Current System Date | BTCH\_END\_DTTM | TIMESTAMP |

## File Creation Process

The following folder structure will be created in the Informatica server(s) to store the Anonymous Data extract files

<Environment>/TgtFiles/EDE/Anonymous/

IDS will generate delta files in .csv format, per month, per zip code with anonymous data for all accounts that pass the 15/15 rule within a delivery class. This results in one extract file per zipcode for a processing month.

The naming standard for the files will be as follows

ANONYMOUS\_DATA\_<MONTHID in YYYYMM format>\_<ZIPCODE>.csv

Ex: ANONYMOUS\_DATA\_201510\_60606.csv

## File Transfer Process

N/A - File transfers will not be part of the IDS process and will be handled by a different team.

## Unix Scripts

**CDM\_TO\_STG\_ANONYMOUS\_DATA.ksh:** Used to run the workflow wf\_CDM\_TO\_STG\_ANONYMOUS\_DATA\_PASS. This will be invoked by JSS.

**STG\_ANONYMOUS\_DATA\_PASS1.ksh:** Used to run the wf\_STG\_ANONYMOUS\_DATA\_PASS1 workflow with a no wait pmcmd option. The no wait pmcmd option is called using the newly built ihub\_pmcmd\_anonymous.ksh. This script will be called by JSS or through command task in an Informatica workflow.

**STG\_ANONYMOUS\_DATA\_PASS2.ksh:** Used to run the wf\_STG\_ANONYMOUS\_DATA\_PASS2 workflow with a no wait pmcmd option. The no wait pmcmd option is called using the newly built ihub\_pmcmd\_anonymous.ksh. This script will be called by command task in an Informatica workflow.

**CDM\_ANONYMOUS\_DATA\_EXTRACT.ksh:** Used to run the workflow wf\_CDM\_ANONYMOUS\_DATA\_EXTRACT. This will be invoked by JSS.

**ANONYMOUS\_DATA\_SCAN\_FAILED.ksh:** Used to run the workflow wf\_ANONYMOUS\_DATA\_SCAN\_FAILED. This will be invoked by JSS.

The following scripts will be used for **History data load** -

**CDM\_TO\_STG\_ANONYMOUS\_DATA\_HIST.ksh:** Used to run the workflow wf\_CDM\_TO\_STG\_ANONYMOUS\_DATA\_PASS. This will be invoked by JSS. Here the month ID is passed as a parameter with the value stored in the control file ANONYMOUS\_DATA\_HIST\_MONTH\_ID.txt.

**CDM\_ANONYMOUS\_DATA\_EXTRACT\_HIST.ksh:** Used to run the workflow wf\_CDM\_ANONYMOUS\_DATA\_EXTRACT. This will be invoked by JSS. This script will also increment the month Id by one month and update the ANONYMOUS\_DATA\_HIST\_MONTH\_ID.txt file with the new value. This way the next run will automatically be using the next month ID to process.

There is no need to create a separate copy of the other scripts for history load. All other regular scripts will be used as is to process the data and it will automatically process it for that history month ID used in the first script (CDM\_TO\_STG\_ANONYMOUS\_DATA\_HIST.ksh).

# Job Scheduling

## Scheduling configuration

After this process successfully completes, JSS will be used to submit a batch process that will launch Informatica ETL processing , please note that is only an example:

|  |  |
| --- | --- |
| JSS Job Stream Execution Configuration | Definitions |
|  | Name of the JSS job stream |
|  | Ksh directory path |
|  | Name of the script |
| **Agent: t42-ccc-01.ceco.com** | Name of the agent |
| **USER: infcami** | Name of the user |
| **Args: CAMI\_DATA\_HUB** | Name and folder of the file containing the passed arguments |
|  | Name of the workflow being processed |
| **exitcode 0 success** | Return code sent back to JSS |
| **Endjob** | End of the JSS job stream |
| **Dependency** |  |

* See page 54 of the PowerPath-Developers-Handbook.doc for further explanation of how JSS works.

## Performance Considerations

Database tables & Informatica jobs are highly optimized for processing and handling data (millions of records in every run) in less time with high efficiency to reduce the over all runtime using below techniques.

## Batch Job Frequency & Data Volume

Anonymous data delta files will be created on a monthly basis following go live. The History job will generate data for up to 24 files with anonymous data from accounts that pass the 15/15 rule per zip code.

Currently there are 572 zip codes in the ComEd territory. 378 of those zip codes have at least one account with an AMI meter. Based on the number of zip codes, no more than 600 files will be created per month, each file will contain the previous month’s anonymous interval data.

Projected AMI Meters in ComEd Territory

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year End Projection | AMI Meters Installed | File Size per Meter (KB) | Total File Size (MB) | Total File Size (GB) | Total File Size (TB) | Total Compressed File Size (GB) |
| 2015 | 1,724,361 | 600 | 1034616.6 | 1034.62 | 1.03 | 51.73 |
| 2016 | 2,654,361 | 600 | 1592616.6 | 1592.62 | 1.59 | 79.63 |
| 2017 | 3,584,361 | 600 | 2150616.6 | 2150.62 | 2.15 | 107.53 |
| 2018 | 4,157,000 | 600 | 2494200 | 2494.2 | 2.49 | 124.71 |

## Error Handling

**Log files**: Unix Path

**Bad files**: Unix Path

## Archiving & Re-Run Process:

Archiving: The IDS team is not building any process to compress or archive the files generated for Anonymous Data. This will be handled by a different team.

Re-extracting data for any month: If there is a need to re-extract data for any month, simply update the control file $ENVR/TgtFiles/ANONYMOUS\_DATA\_HIST\_MONTH\_ID.txt with the MONTHID in ‘YYYYMM’ format (Ex: 201512 for December 2015). Once that is completed, utilize the the on-demand schedule for Anonymous data (used for History load) to process the data.

## Scheduler Process Flow & Job Name

JSS scheduling will submit the Informatica ETL. No external parameters other than the normal Informatica and environmental parameters will be required to pass into the mapping. This ETL is designed to be standalone.

JSS Scheduler will submit Unix batch processing script:

* Set environment variables.
* Launch Informatica.
* Batch processing will provide email: “wf\_ has SUCCESSFULLY COMPLETED!”
* Return code is passed back to JSS

## Implementation Considerations

**Mapping Performance:** Proper optimization techniques will apply for faster and efficient exectution of this job.

**Mapping Properties:** All mapping should have at minimum the following properties set for the target file.

* The owner and group users should have read, write and executable privileges.

# Unit Testing Criteria

## Test Plan

Below is the information about the test scenario for which the code has been tested. This is only an example. The full list will be provided as part of Unit Testing.

**Unit Test:**

1. Verify if each target column is mapped to the correct source column.
2. Verify if the 15/15 rule is applied correctly.
3. Verify if the files are populated with the correct data.
4. For a processing month, verify that Zipcode and Delivery Class combinations marked as ‘Passed’ in the summary table should be part of the Extract files and the ones marked as ‘Failed’ should not be part of the extract files.

# Appendices

## Appendix A – Key Terms

A full glossary of terms can be found at the SharePoint location below:

<http://teamapps.exeloncorp.com/sites/ITComEdAMI/MDM%20Implementation/02.%20Analyze/00-Project%20Glossary/ComEd%20AMI%20Project%20-%20Glossary%20of%20Terms.xlsx>

## Appendix B – Outstanding Items

|  |  |  |  |
| --- | --- | --- | --- |
| **Issue /Item** | **Related QC Item** | **Notes** | **Status** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |