Checkbook_NYC_2

Checkbook NYC 4.0 M/WBE & Sub Vendors ETL Implementation Approach Document

version 0.1

October 24, 2014

**DOCUMENT INFORMATION AND APPROVALS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Author** | **Version** | Reviewer | Position | Version approved |
| 10/24/2011 | TirupatiReddy Manyam | 0.1 |  | Principal Engineer | 0.1 |

**Table of Contents**

[Background 5](#_Toc343697424)

[Scope 5](#_Toc343697425)

[Overview 5](#_Toc343697426)

[ETL Process 8](#_Toc343697428)

M/WBE [ETL Process 8](#_Toc343697428)

[Update Spending (FMS) Data 12](#_Toc343697435)

[Update Contracts (CON & MAG) Data 12](#_Toc343697436)

[Refresh Aggregate Data 12](#_Toc343697436)

Sub Vendor [ETL Process 8](#_Toc343697428)

[Process Sub Contract business type 12](#_Toc343697436)

[Process Sub Contract Details 13](#_Toc343697437)

[Process Sub Contract Spending 13](#_Toc343697438)

[Appendix A](#_Toc343697449) 20

[SV data validations](#_Toc343697451) 21

[SC data validations](#_Toc343697452) 22

[SF data validations](#_Toc343697453) 24

## Background

Checkbook NYC 4.0 is a major enhancement to Checkbook NYC 3.0 by adding M/WBE and Sub Vendors feature dashboards.

## Scope

This document provides an insight into the ETL Implementation Approach for M/WBE and Sub Vendors enhancements in Checkbook NYC 4.0. ETL process starts with the consumption of raw data, validating, processing and posting it to the transaction tables.

## Overview

Kettle, Pentaho data integration tool will be used to integrate the various ETL processes to fetch files from remote server(s), loading data to database from flat files, data cleansing etc.

Master kettle job will be created which has several transformations and invokes several other jobs. This will be started around 11 PM when the files are published by FISA. All the source files from FISA are UTF-8 complaint. Processing of the data begins only when the required files are available.

## ETL Process

ETL process starts with the directory listing of FISA server and once the required details are obtained, each of the data file is processed & changes are posted in the transaction table in the order of COA/ FMSV, Revenue/Budget/MAG, CON, FMS/PMS, SubContract, SubSpending.

Various transformations in the master job are explained in this section. If any error occurs in a transformation a DB function is executed by passing the load file identifier as a parameter to determine if ETL job can continue with the next data file or it has to abort the entire job.

* If any of the COA data file result in an error the entire job is aborted. Processed\_flag is set to E for the corresponding file and all data files which had to be consumed will be set to C for cancelled.
* For any feed other than COA only the non-processed files of the specific feed are cancelled. For example if FMS data file resulted in an error, processed\_flag is set to E for the corresponding file and all other FMS data files which had to be consumed will be set to C for cancelled.

##### 

## M/WBE ETL Process

In this we will categorize the data based on vendor minority category.

##### Update Spending (FMS) data

1. In this the ETL Script will update the spending transactions and payroll summary transactions that are inserted/updated in the current data load.
2. Update payroll summary transactions minority category as “Individuals & Others”.
3. Update spending transactions with the latest minority category of the associated vendor.
4. Update spending transactions which are not categorized into any minority category and the associated contract’s agreement type code is one of ('35', '36', '39', '40', '44', '65', '68', '79', '85') to “Individuals & Others”.
5. Update spending transactions which are not categorized into any minority category and the associated contract’s award method code is one of ('07', '08', '09', '17', '18', '44', '45', '55') to “Individuals & Others”.
6. Update spending transactions which are not categorized into any minority category and vendor organization classification code is one of (1, 5, 6, 7, 8, 9, 12, 13, 14, 15, 19, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36) to “Individuals & Others”.
7. Update spending transactions which are not categorized into any minority category to “Non-Minority”.
8. For one time as Checkbook started getting data from 01/13/2013, update the minority category of the spending transactions whose check issued date is before 01/13/2013 using the M/WBE Application (<http://mwbereportcardnyc.com/>) data.

##### Update Contracts (CON & MAG) data

1. In this the ETL Script will update the contract transactions that are inserted/updated in the current data load.
2. Update contract transactions with the latest minority category of the associated vendor.
3. Update contract transactions which are not categorized into any minority category and the agreement type code is one of ('35', '36', '39', '40', '44', '65', '68', '79', '85') to “Individuals & Others”.
4. Update contract transactions which are not categorized into any minority category and the award method code is one of ('07', '08', '09', '17', '18', '44', '45', '55') to “Individuals & Others”.
5. Update contract transactions which are not categorized into any minority category to “Non-Minority”.
6. Below are the agreement type codes used to categorize into “Individuals & Others”.

|  |  |
| --- | --- |
| Agreement Type  Code | Agreement Type Name |
| 35 | LESSEE |
| 36 | MISCELLANEOUSPROPERTYRENTAL |
| 39 | LESSOR - REVENUE |
| 40 | LESSOR - ACCOUNTING LINES EXIST |
| 44 | PUBLICUTILITY |
| 65 | LOANS |
| 68 | FORCEACCOUNTAGREEMENT |
| 79 | WATERSHEDLANDACQUISITION |
| 85 | INTRA-AGENCYFUNDAGREEMENTS |

1. Below are the award method codes used to categorize into “Individuals & Others”.

|  |  |
| --- | --- |
| Agreement Type  Code | Agreement Type Name |
| 07 | LESSEE NEGOTIATION |
| 08 | LOAN NEGOTIATION |
| 09 | RENTAL SUBSIDY NEGOTIATION |
| 17 | GOVERNMENT TO GOVERNMENT |
| 18 | NON PROCUREMENT TRANSACTION |
| 44 | PUBLIC UTILITY |
| 45 | SMALL PURCHASE-PUBLIC UTILITY |
| 55 | NO LONGER VALID - CNV |

##### Refresh Aggregate Data

Once all the M/WBE category updates are done to spending, payroll summary and contract transactions, all M/WBE contracts and spending aggregate tables should be refreshed. ETL script will refresh table by table, first by truncating the data in the table and then inserting the aggregate data back using the queries on the respective transaction tables.

1. Update contract transactions with the latest minority category of the associated vendor.
2. Update contract transactions which are not categorized into any minority category and the agreement type code is one of ('35', '36', '39', '40', '44', '65', '68', '79', '85') to “Individuals & Others”.

## Appendix A

#### CON data validations

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CON** | | | | | | | | | | | | | |
| **Type of feed** | | Monthly – Incremental  Daily – Incremental | | | | | | | | | | | |
| **File Name Pattern** | | AIDP\_DLY\_PCO\_PO\_CCYYMMDDHHMMSS.asc (Daily) AIMP\_MTHLY\_PCO\_PO\_CCYYMMDDHHMMSS.asc (Monthly) | | | | | | | | | | | |
| **Number of files to be processed** | | Multiple files processed in the order in which it was created in FTP server. | | | | | | | | | | | |
| **Tables** | | | | | | | | | | | | | |
| **Entity Name** | **Document Type** | | **Entity Identifier** | **Unique elements** | | **Staging Table** | | **Transaction Table** | | **History Table** | | | |
| CON header |  | | H | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no | | etl.stg\_con\_ct\_header(CT1, CTA1, CTA2)  etl.stg\_con\_po\_header (POC, PCC1, POD)  etl.stg\_con\_do1\_header (DO1) | | history\_agreement | |  | | | |
| CON Award detail | All except DO1 | | W | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  awd\_meth\_cd | | etl.stg\_con\_ct\_award\_detail(CT1, CTA1)  etl.stg\_con\_po\_award\_detail(POC, PCC1, POD) | | history\_agreement  history\_agreement\_worksite  history\_agreement\_worksite | |  | | | |
| CON Vendor | All | | V | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  doc\_vend\_ln\_no | | etl.stg\_con\_ct\_vendor (CT1, CTA1)  etl.stg\_con\_po\_vendor (POC, PCC1, POD)  etl.stg\_con\_do1\_vendor (DO1) | | history\_agreement | |  | | | |
| CON Commodity | All | | C | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  doc\_vend\_ln\_no  doc\_vend\_ln\_no  doc\_comm\_ln\_no | | etl.stg\_con\_ct\_commodity(CT1, CTA1)  etl.stg\_con\_po\_commodity(POC, PCC1, POD)  etl.stg\_con\_do1\_commodity(DO1) | | history\_agreement\_commodity | |  | | | |
| CON accounting line | All | | A | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  doc\_vend\_ln\_no  doc\_vend\_ln\_no  doc\_comm\_ln\_no  doc\_actg\_ln\_no | | etl.stg\_con\_ct\_accounting\_line(CT1)  etl.stg\_con\_po\_accounting\_line(POC, PCC1, POD) etl.stg\_con\_do1\_accounting\_line(DO1) | | history\_agreement\_accounting\_line | |  | | | |
| **Validation Rules** | | | | | | | | | | | | | |
| **Entity Name** | **Missing Key Values** | | **Duplicate** | | **Invalid/Inconsistent Values** | | **Missing Parent/child Entity** | | **Associated to invalid entities** | **Multiple records** | | **Inter-Load duplicates** | |
| Header | 1.doc\_cd  2.doc\_dept\_cd  3.doc\_id  4.doc\_vers\_no | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no | | 1.cntrct\_strt\_dt > cntrct\_end\_dt  2. max\_cntrc\_am < 0 | | Header without an award detail record (for document types other than DO1) | | 1. Invalid award detail record  2. Invalid vendor  3. Invalid commodity  4.Invalid accounting line |  | |  | |
| Award detail  (All except DO1) | 1.doc\_cd  2.doc\_dept\_cd  3.doc\_id  4.doc\_vers\_no  5.awd\_meth\_cd | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  awd\_meth\_cd | | 1.Invalid awd\_meth\_cd (Not in ref\_award\_method)  2.Invalid cttyp\_cd ( Not in ref\_agreement\_type)  3. Invalid ctcat\_cd\_1 (ref\_award\_category). Same for the other 4 award categories. | | Award detail records with a header | | 1.Invalid header  2.Invalid vendor  3.Invalid commodity  4.Invalid accounting line | Multiple award detail records for a header. | |  | |
| Vendor | 1.doc\_cd  2.doc\_dept\_cd  3.doc\_id  4.doc\_vers\_no  5. doc\_vend\_ln\_no | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no | |  | | Vendor record without a header | | 1.Invalid header  2.Invalid award detail  3.Invalid commodity  4.Invalid accounting line | Multiple vendor records for a header. | |  | |
| Commodity | 1.doc\_cd  2.doc\_dept\_cd  3.doc\_id  4.doc\_vers\_no  5.doc\_vend\_ln\_no  6.doc\_comm\_ln\_no | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  doc\_vend\_ln\_no  doc\_comm\_ln\_no | |  | | commodity record without a header | | 1.Invalid header  2.Invalid award detail  3.Invalid vendor  4.Invalid commodity  5.Invalid accounting line |  |  | |
| Accounting line | 1.doc\_cd  2.doc\_dept\_cd  3.doc\_id  4.doc\_vers\_no  5.doc\_vend\_ln\_no  6.doc\_comm\_ln\_no  7. doc\_actg\_ln\_no | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  doc\_vend\_ln\_no  doc\_comm\_ln\_no  doc\_actg\_ln\_no | |  | | Accounting line without a header | | 1.Invalid header  2.Invalid award detail  3.Invalid vendor  4.Invalid commodity  5.Invalid accounting line |  |  | |

#### FMS data validations

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **FMS** | | | | | | | | | | | |
| **Type of feed** | | Daily – Incremental | | | | | | | | | |
| **File Name Pattern** | | AIDM\_DLY\_MMDSBP\_DC\_20100601120000.asc (Partially displayed)  AIDM\_DLY\_MMDSBX\_DC\_20100601120000.asc (Not displayed)  AIDM\_DLY\_MMDSBP\_DC\_20100601120000.asc (displayed fully) | | | | | | | | | |
| **Number of files to be processed** | | Multiple files processed in the order in which it was created in FTP server. | | | | | | | | | |
| **Tables** | | | | | | | | | | | |
| **Entity Name** | **Entity Identifier** | | **Unique elements** | **Staging Table** | | **Transaction Table** | | **History Table** | | | |
| Disbursement | H | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no | etl.stg\_fms\_header | | expenditure  all\_expenditure | | NA | | | |
| Vendor | V | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  doc\_vend\_ln\_no | etl.stg\_fms\_vendor | | expenditure  all\_expenditure | | NA | | | |
| Disbursement Accounting line | A | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  doc\_vend\_ln\_no  doc\_actg\_ln\_no | etl.stg\_fms\_accounting\_line | | expenditure\_line\_item  all\_expenditure\_line\_item | | NA | | | |
| **Validation Rules** | | | | | | | | | | | |
| **Entity Name** | **Missing Key Values** | | **Duplicate** | | **Invalid/Inconsistent Values** | | **Missing Parent/Child Entity** | | **Associated to invalid entities** | **Multiple records** | **Inter-Load duplicates** |
| Disbursement | 1.doc\_cd  2.doc\_dept\_cd  3.doc\_id  4.doc\_vers\_no  5.doc\_rec\_dt\_dc  6. doc\_bfy  7. doc\_fy\_dc  8. doc\_per\_dc  9. chk\_eft\_am  10.chk\_eft\_iss\_dt | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no | |  | | 1.Disbursement without vendor  2. Disbursement without line items | | 1.Invalid vendor  2. Invalid disbursement line item |  | Invalidate the disbursement if there exists a disbursement in the transaction table for the combination of given unique elements. |
| Vendor | 1.doc\_cd  2.doc\_dept\_cd  3.doc\_id  4.vend\_cust\_cd | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no | |  | |  | |  | Multiple vendors for a disbursement. |  |
| Disbursement Accounting line | 1.doc\_cd  2.doc\_dept\_cd  3.doc\_id  4.doc\_vers\_no  5.doc\_vend\_ln\_no  6.doc\_actg\_ln\_no  7. dept\_cd  8. bfy  9. fy\_dc  10. per\_dc  11. chk\_amt  12.rfed\_doc\_cd  13.rfed\_doc\_dept\_cd  14.rfed\_doc\_id  15.rfed\_actg\_ln\_no | | doc\_cd  doc\_dept\_cd  doc\_id  doc\_vers\_no  doc\_actg\_ln\_no | |  | | Disbursement line item without disbursement | | 1.Invalid disbursement  2.Invalid disbursement line item |  |  |

#### Revenue data validations

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Revenue** | | | | | | | | | | |
| **Type of feed** | | Daily – Incremental  Monthly – Incremental | | | | | | | | |
| **File Name Pattern** | | AIDJ\_DLY\_JL\_JL\_032311015417.ASC  AIDJ\_MTHLY\_JL\_JL\_040111015417.ASC | | | | | | | | |
| **Number of files to be processed** | | Multiple | | | | | | | | |
| **Tables** | | | | | | | | | | |
| **Entity Name** | **Entity Identifier** | | **Unique elements** | **Staging Table** | | **Transaction Table** | | **History Table** | | |
| Revenue transaction | NA | | None | Etl.stg\_revenue | | Revenue | |  | | |
| **Validation Rules** | | | | | | | | | | |
| **Entity Name** | **Missing Key Values** | | **Duplicate** | **Invalid/Inconsistent Values** | **Missing Referenced Parent Entity** | | **Associated to invalid entities** | | **Multiple Records** | **Inter-Load duplicate** |
|  |  | |  |  |  | |  | |  | . |