**Operation Manual**

**Data Ingestion to EDL using aDIP**

*Ascena Brands*



Table of Contents

[1 Document Revision History 3](#_Toc20479608)

[2 Program Overview 4](#_Toc20479609)

[2.1 Overview 4](#_Toc20479610)

[2.2 Overview of Functionality and Key Processes 4](#_Toc20479611)

[2.2.1 Transmission of Files to Edge Node (Predecessor process) 4](#_Toc20479612)

[2.2.2 File Decryption Process 4](#_Toc20479613)

[2.2.3 Validation and Load in Staging layer 4](#_Toc20479614)

[2.3 Scheduled Events 5](#_Toc20479615)

[2.4 Infrastructure and Design 6](#_Toc20479616)

[2.4.1 Process Flow Diagram 6](#_Toc20479617)

[2.4.2 Code and aDIP Script Details 7](#_Toc20479618)

[2.4.3 Source Files 8](#_Toc20479619)

[3 Security and Access Control 8](#_Toc20479620)

[4 Maintenance and Configuration Management 9](#_Toc20479621)

[4.1 Code Repository 9](#_Toc20479622)

[4.2 Maintenance of Source files 9](#_Toc20479623)

[5 Monitoring and Alerting 10](#_Toc20479624)

[5.1 Daily Operational Monitoring and Alerting 10](#_Toc20479625)

[6 Operational Tasks 10](#_Toc20479626)

[6.1 Deployment 10](#_Toc20479627)

[7 Contact Details 10](#_Toc20479628)

[8 Additional Document References 11](#_Toc20479629)

[9 Appendix 11](#_Toc20479630)

# Document Revision History

| Date | Author | **Revision Description** |
| --- | --- | --- |
| 2019/09/23 | Bhaskar Ghosh | Initial Version |
|  |  |  |

# Program Overview

## Overview

|  |  |
| --- | --- |
| **Purpose** | All History Incremental daily delta feeds are sourced from Teradata CRM for Plus (LB and CA)  All such feeds land on edge node encrypted. These files are decrypted and subsequently ingested in the Enterprise Data Lake (EDL)  Most of these delta files (10 out of 11 files) are already available in Edge Node when aDIP cron jobs run to decrypt them and push them to Staging tables.  Once delta files are pushed to EDL Staging, then next steps of data flow process will take over and ingest into EDL Conform Layer (those downstream processes are NOT covered in this document). |
| **Document Scope** | This operation manual will contain documentation of all cron jobs that decrypts and pushes History Incremental delta files coming from CRM Teradata database to EDL staging via process called aDIP for downstream processing. |

## Overview of Functionality and Key Processes

### Transmission of Files to Edge Node (Predecessor process)

From CRM Teradata, History Incremental files are extracted via daily Control M batches and sent to EFT server.

Files are encrypted in EFT server and its own batches push these files to EDL cem/cdp/lbca/incremental input directory.

So, EFT job encrypts all History Incremental Daily files with the EDL key, and then transmits to the edge node.

### File Decryption Process

All History Incremental daily delta files from CRM are sent to edge node in decrypted .gpg format (via EFT server) in the cdp incremental directory for lbca. All such files are then decrypted to text (.txt) format directly by a cron shell script run around 1:40 AM as part of aDIP process.

### Validation and Load in Staging layer

Decrypted text daily delta files are then pushed to EDL Staging. As part of this process, it is checked, if all History Incremental files are present for that batch id date. Only if all files are present, then they are posted in hdfs.

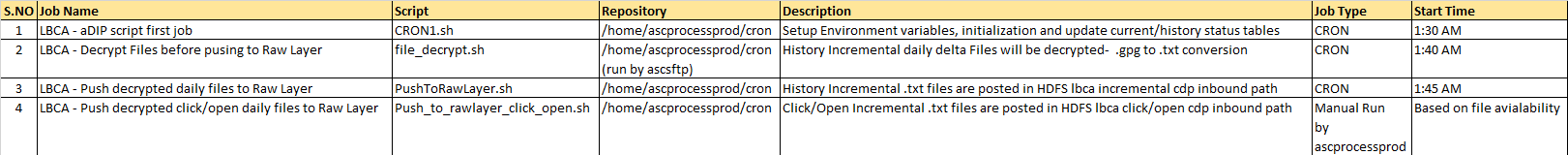
Then hql scripts for each History Incremental feed loads data into staging Hive tables. These staging tables are used for downstream consumption by conform layer Hive tables.

After successful staging of data, original encrypted .gpg files are then archived.

## Scheduled Events

The “aDIP\_Cron Jobs\_LBCA.xlsx” spreadsheet will provide schedules for all cron jobs part of aDIP process.

Unless otherwise noted, all times are in EST and all frequencies are daily.



aDIP cron job details

## Infrastructure and Design

### Process Flow Diagram

**CRM (Teradata)**

**History Incremental**

lbca\_bc\_address

lbca\_address\_customer

lbca\_ecom\_address

lbca\_email\_customer

lbca\_emp\_address

lbca\_missing\_phone\_customer

lbca\_missing\_email\_customer

lbca\_merge

lbca\_pos\_address

lbca\_phone\_customer

PushToRawLayer.sh - Checks status of aDIP batches. Then it will push all decrypted text files to HDFS. Next .hql scripts consume these files to Staging Layer Hive tables. Run by id: ascprocessprod. Execution time: 1:45 AM

file\_decrypt.sh - Decrypts all History Incremental Delta encrypted .gpg files to .txt text format. Run by id: ascftp. Execution time: 1:40 AM

aDIP cron scripts

CRON1-Checks status of aDIP batches current/previous batch ids. Run by id: ascprocessprod. Execution time: 1:30 AM

Daily delta files land

Host Name: edllbphed02.prod.ascena.com

Enterprise Data Lake (EDL)

EFT Pushes

Encryption of delta files

EFT Server

**CONTROL M Batches**

### Code and aDIP Script Details

Following cron aDIP scripts are run in prod Edge Node host edllbphed02.prod.ascena.com

**Code Path:** /home/ascprocessprod/cron

**Script Name:** CRON1.sh

**Run by id:** ascprocessprod

**Cron execution time:** 1:30 AM

LB and CA aDIP batch progress status are maintained in hive tables lbca\_incremental\_batch\_details and lbca\_incremental\_batch\_details\_history

This script checks status of aDIP batches for current and previous batch ids.

For Current batch\_id, if status is OPEN/INPROGRESS/COMPLETED then it will abort

For previous batch\_id, if status is OPEN/INPROGRESS, script will abort.

In case previous batch is COMPLETED, then it will overwrite status tables with status as OPEN

If status tables are empty, this script will initializes status/progress of the current and previous batch ids in hive tables with status as ‘OPEN’.

The multi-threading/parallel processing of this script is controlled by go file cron1.go

**Code Path:** /home/ascprocessprod/cron

**Script Name:** file\_decrypt.sh

**Run by id:** ascsftp

**Cron execution time:** 1:40 AM

This script checks all History Incremental files are present in the input cem/cdp incremental directory for lbca

(Checks by file count as well as presence of .ctl files)

Then script will decrypt all such History Incremental Delta encrypted .gpg files to .txt text format

Upon successful decryption, this script will rename all input .ctl files with extension .flag

The multi-threading/parallel processing of this script is controlled by go file decrypt.go

**Code Path:** /home/ascprocessprod/cron

**Script Name:** PushToRawLayer.sh

**Run by id:** ascprocessprod

**Cron execution time:** 1:45 AM

This script checks status of aDIP batches for current and previous batch ids.

For Current batch\_id, if status is INPROGRESS/COMPLETED then it will abort

If status of current batch\_id is OPEN, then it will overwrite status as INPROGRESS

Then it will push all decrypted text files (10 files in total) to HDFS

Next .hql files in directory /home/ascprocessprod/cron/ctl are used to consume all these files to refresh (truncate and load) the Staging Layer Hive tables for each of these feeds.

**Code Path:** /home/ascprocessprod/cron

**Script Name:** Push\_to\_rawlayer\_click\_open.sh

**Run by id:** ascprocessprod

**Execution time:** Manual Run by ascprocessprod

This script will do same tasks as mentioned for above script PushToRawLayer.sh, but for click/open delta files.

This script is run manually and support associate waits for source file to be extracted from CRM

### Source Files

Following is list of files that are extracted from CRM Teradata as part of History Incremental files to be consumed in downstream EDL

Host directory in Edge Node: /ascena\_prod/data/secure/cem/cdp/lbca/incremental

Sub-Directories with daily delta files:

lbca\_bc\_address

lbca\_address\_customer

lbca\_ecom\_address

lbca\_email\_customer

lbca\_emp\_address

lbca\_missing\_phone\_customer

lbca\_missing\_email\_customer

lbca\_merge

lbca\_pos\_address

lbca\_phone\_customer

Click/Open Host Directory: /ascena\_prod/data/secure/cem/cdp/lbca/click\_open\_396

Sub-Directories with daily delta files:

email\_opt\_in\_out

lbca\_email\_last\_click\_open

phone\_opt\_in\_out

We are archiving all History Incremental files by batch\_id in following location:

/ascena\_prod/data/archive/cem/cdp/lbca/incremental>

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:11 lbca\_bc\_address

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:12 lbca\_ecom\_address

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:12 lbca\_emp\_address

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:13 lbca\_pos\_address

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:13 lbca\_email\_customer

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:14 lbca\_phone\_customer

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:14 lbca\_address\_customer

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:15 lbca\_missing\_phone\_customer

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:15 lbca\_missing\_email\_customer

drwxrwxrw-+ 2 ascprocessprod ascprocessprod 4096 Sep 25 02:16 lbca\_merge

# Security and Access Control

* aDIP cron jobs – These scripts are run within EDL by ids ascprocessprod and ascsftp.

Read/Write permissions to Production host edllbphed02.prod.ascena.com directories should be maintained

/home/ascprocessprod

drwxrwxr-x. 7 ascprocessprod ascprocessprod 4096 Sep 10 15:22 cron

/home/ascprocessprod/cron>

drwxrwxr-x. 2 ascprocessprod ascprocessprod 100 Apr 18 15:28 BKUP

drwxrwxr-x. 8 ascprocessprod ascprocessprod 8192 Sep 6 08:39 ctl

drwxrwxr-x. 5 ascprocessprod ascprocessprod 135168 Sep 25 08:59 log

drwxrwxr-x. 2 ascprocessprod ascprocessprod 65536 Sep 25 08:59 work

(Timestamps shown as example)

Downstream process to merge data from these History Incremental Delta files to Hive tables should also be ensured

# Maintenance and Configuration Management

## Code Repository

* All aDIP cron jobs executed within EDL to consume History incremental files are responsibility of the EDL support team. Edge Node Production host: edllbphed02.prod.ascena.com

All cron scripts are running from directory: /home/ascprocessprod/cron

## Maintenance of Source files

In case, any new incremental feed to be added for aDIP processing or remove any existing feed, we need to include/remove concerned source file names from following file:

/home/ascprocessprod/cron/ctl/lbca\_history\_incremental\_file\_list

Accordingly, need to add/remove source feed directories under:

/ascena\_prod/data/secure/cem/cdp/lbca/incremental

Daily delta files landing should be present in previously mentioned directories with .ctl and .gpg extensions

Similarly, in HDFS, we need to add/remove directories for these delta feeds

hdfs dfs -ls /ascena\_prod/data/inbound/cem/cdp/lbca/incremental

Found 10 items

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:10 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_address\_customer

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:07 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_bc\_address

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:08 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_ecom\_address

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:10 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_email\_customer

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:09 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_emp\_address

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:12 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_merge

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:11 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_missing\_email\_customer

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:11 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_missing\_phone\_customer

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:10 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_phone\_customer

drwxr-xr-x - ascprocessprod hdfs 0 2019-09-25 02:09 /ascena\_prod/data/inbound/cem/cdp/lbca/incremental/lbca\_pos\_address

Need to create/remove the .hql Hive data ingestion script within directories of new/removed feed in following location:

/home/ascprocessprod/cron/ctl

Need to create/remove archive directories for new/removed feed under:

/ascena\_prod/data/archive/cem/cdp/lbca/incremental

# Monitoring and Alerting

## Daily Operational Monitoring and Alerting

* aDIP cron Jobs – Execution logs of all such scripts are posted in /home/ascprocessprod/cron/log directory. History Incremental scripts along with other batches (Including ones run manually) are reviewed and monitored by EDL support team

# Operational Tasks

## Deployment

* aDIP cron jobs – Any changes to these scripts are approved via iVanti change request and deployed in Edge Node production path by EDL support team. Currently there is no version control system or repository for these scripts, but plan is to move these scripts to GitHub repository in near future

# Contact Details

| Contact | Role | Email | Phone |
| --- | --- | --- | --- |
| EDL Support Team | Monitors and Maintains all Hygiene processes | [aBS-IT-EDL-Support@AscenaRetail.com](mailto:aBS-IT-EDL-Support@AscenaRetail.com) | N/A |

# Additional Document References

| # | Document Name | Document Link | Notes |
| --- | --- | --- | --- |
| 1. | aDIP\_Cron Jobs\_LBCA.xlsx | <http://epm01/sites/IT_Sites/EnterprisData_Lake_Operations/_layouts/xlviewer.aspx?id=/sites/IT_Sites/EnterprisData_Lake_Operations/Shared%20Documents/Operation%20Manuals/aDIP_Cron%20Jobs_LBCA.xlsx&Source=http%3A%2F%2Fepm01%2Fsites%2FIT%5FSites%2FEnterprisData%5FLake%5FOperations%2FShared%2520Documents%2FForms%2FAllItems%2Easpx%3FRootFolder%3D%252Fsites%252FIT%255FSites%252FEnterprisData%255FLake%255FOperations%252FShared%2520Documents%252FOperation%2520Manuals%26FolderCTID%3D0x012000BF3C75A6CF6ED6428B0F177BD4F62951%26View%3D%7BE035007A%2DA28A%2D42C1%2D86BC%2D4C1E4E5547EE%7D&DefaultItemOpen=1&DefaultItemOpen=1> | This spreadsheet contains list of all aDIP cron jobs that pushes CRM Teradata History Incremental daily files to Edge Node. **It is primary external documentation for this operation manual.** |
| 2. | EFT Transmission to Edge Node - Process Flow Diagram.vsdx | [http://epm01/sites/IT\_Sites/EnterprisData\_Lake\_Operations/Shared%20Documents/Operation%20Manuals/EFT%20Transmission%20to%20Edge%20Node%20-%20Process%20Flow%20Diagram.vsdx](http://epm01/sites/IT_Sites/EnterprisData_Lake_Operations/Shared%20Documents/Operation%20Manuals/EFT%20Transmission%20to%20Edge%20Node%20Documents/EFT%20Transmission%20to%20Edge%20Node%20-%20Process%20Flow%20Diagram.vsdx) | This diagram outlines the steps that a file takes in order to land on the EDL edge node in order to prepare it for ingestion to the staging database in Hive. This diagram is a visual companion to the textual spreadsheet “EFT Transmission to Edge Node - File and Data Orchestration.xlsx” |

# Appendix