

CRM data upload to DMP (AAM)

Implementation Guide

Last Updated: **Mar-2019**

Akshay Kashyap (DMP Developer)

Gaurav Dembla (DMP Lead)

Fabricio Monge (Data Analyst)

Table of Contents

[**Introduction** 3](#_Toc2803518)

[High-Level Process 3](#_Toc2803519)

[Pre-requisites 3](#_Toc2803520)

[**CRM Data** 4](#_Toc2803521)

[Conditions 4](#_Toc2803522)

[Attributes 4](#_Toc2803523)

[Extraction 6](#_Toc2803524)

[**AAM Upload** 7](#_Toc2803525)

[File Format 7](#_Toc2803526)

[File Name 8](#_Toc2803527)

[File Size 8](#_Toc2803528)

[S3 Details 9](#_Toc2803529)

[**Appendix** 10](#_Toc2803530)

# **Introduction**

The document is intended to guide the generation and onboarding of CRM data feed onto DMP - Adobe Audience Manager (AAM), as outlined below.

## High-Level Process

1. Extract required offline data from database into a csv/xlsx file
2. Create Onboarded Traits in AAM corresponding to attributes required for segmentation and targeting
3. Convert csv/xlsx file into AAM ingestible feed
4. Upload the feed in given S3 location
5. Confirm successful feed upload from Adobe Support team
6. Wait for 48 hours before traits are populated

## Pre-requisites

Prior to uploading required data into AAM, following items should be taken care of.

* Data feed columns/attributes must be finalized by business (DM Ops).
* Data Source in AAM must be present. For CRM data upload, we shall use already existing data source in AAM – “DMP - De-anonymization” (Source Id: 180427), which also serves the purpose of de-anonymization (ID synchronization).
* DM Operations team needs to create Onboarded Traits (to receive data feed Signals).
* For DM Ops team to create Onboarded traits (for each unique key-value pair), EDA needs to share unique values of each attribute with Ops team.

# **CRM Data**

CRM data feed comprises of offline information on Contacts assimilated from various sources in and outside of VMware. We shall use marketing data hub ATLAS (to be replaced by Galileo in near future) for generation of this feed.

Grain of the data shall be at individual contact level (i.e. hashed email id). The hashed email id should be generated using SHA256 algorithm.

## Conditions

There are no known conditions for limiting Contacts data for DMP.

As a result, CRM data shall have Contacts from all the geographies, and NOT be limited to just AMER. In contrast, data upload to Liveramp is limited to AMER contacts only.

## Attributes

CRM feed to AAM shall have the following attributes, the list of which is very similar to that of CRM feed to Liveramp except for PII information (which is not allowed in DMP) and GEO (which wasn’t part of Liveramp feed in Feb 2019, as the data was limited to AMER only).

All the attributes shall be categorical in nature and shall have handful of values only (definitely less than 100 distinct values), so that the values could be conveniently used for segmentation in DMP.

|  |  |
| --- | --- |
| **S. No.** | **Attribute** |
| 1 | Hash\_Email |
| 2 | ISO Country Name |
| 3 | GEO |
| 4 | City Name |
| 5 | State or Province |
| 6 | Company |
| 7 | Department |
| 8 | VM Industry Derived |
| 9 | Populations |
| 10 | Population Solutions |
| 11 | ETM Coverage Type |
| 12 | Hierarchy Level |
| 13 | Specialist |
| 14 | Site Num Of Employees In Country |
| 15 | VMWorld Attendee\_2018 |
| 16 | bookingLast2Years\_Unclassified |
| 17 | bookingLast2Years\_APPS |
| 18 | bookingLast2Years\_AVAILABILITY |
| 19 | bookingLast2Years\_AVAILABILITY CLOUD |
| 20 | bookingLast2Years\_AVAILABILITY SERVICES |
| 21 | bookingLast2Years\_BCS-MCS |
| 22 | bookingLast2Years\_BETA SERVICES |
| 23 | bookingLast2Years\_CISCO N1K |
| 24 | bookingLast2Years\_CLOUD FOUNDATION |
| 25 | bookingLast2Years\_COMPUTE SERVICES |
| 26 | bookingLast2Years\_DATA REPLICATION |
| 27 | bookingLast2Years\_DATA SERVICES |
| 28 | bookingLast2Years\_DESKTOP WO |
| 29 | bookingLast2Years\_DEVOPS SERVICES |
| 30 | bookingLast2Years\_EMERGING SOLUTIONS |
| 31 | bookingLast2Years\_EMM |
| 32 | bookingLast2Years\_EPP |
| 33 | bookingLast2Years\_GO |
| 34 | bookingLast2Years\_HCI APPLIANCE SA |
| 35 | bookingLast2Years\_HCI APPLIANCE SDDC |
| 36 | bookingLast2Years\_HORIZON |
| 37 | bookingLast2Years\_HORIZON APP MANAGER |
| 38 | bookingLast2Years\_HORIZON CLOUD |
| 39 | bookingLast2Years\_HPP |
| 40 | bookingLast2Years\_MOBILE CLOUD |
| 41 | bookingLast2Years\_MOBILE WO |
| 42 | bookingLast2Years\_MOBILE WO CLOUD |
| 43 | bookingLast2Years\_N&S CLOUD |
| 44 | bookingLast2Years\_NETWORK SERVICES |
| 45 | bookingLast2Years\_NSX |
| 46 | bookingLast2Years\_NSX CLOUD |
| 47 | bookingLast2Years\_OTS |
| 48 | bookingLast2Years\_PERSONAL DESKTOP |
| 49 | bookingLast2Years\_PIVOTAL CF |
| 50 | bookingLast2Years\_PKS |
| 51 | bookingLast2Years\_SAP SOLUTION |
| 52 | bookingLast2Years\_SOURCE UNKNOWN |
| 53 | bookingLast2Years\_SPP |
| 54 | bookingLast2Years\_STORAGE |
| 55 | bookingLast2Years\_STORAGE CLOUD |
| 56 | bookingLast2Years\_STORAGE SERVICES |
| 57 | bookingLast2Years\_TRUST POINT |
| 58 | bookingLast2Years\_VCENTER |
| 59 | bookingLast2Years\_VCENTER PROTECT |
| 60 | bookingLast2Years\_VCLOUD AIR MSP |
| 61 | bookingLast2Years\_VCLOUD AIR NETWORK |
| 62 | bookingLast2Years\_VCLOUD DIRECTOR |
| 63 | bookingLast2Years\_VCLOUD/VRS |
| 64 | bookingLast2Years\_VCNS |
| 65 | bookingLast2Years\_VCPP |
| 66 | bookingLast2Years\_VFABRIC & OTHER |
| 67 | bookingLast2Years\_VIRTUAL INFRASTRUCTURE MANAGEMENT |
| 68 | bookingLast2Years\_VKE |
| 69 | bookingLast2Years\_VMC ON AWS |
| 70 | bookingLast2Years\_VMW CLOUD ON AWS |
| 71 | bookingLast2Years\_VREALIZE AIR |
| 72 | bookingLast2Years\_VREALIZE AUTOMATION |
| 73 | bookingLast2Years\_VREALIZE BUSINESS |
| 74 | bookingLast2Years\_VREALIZE OPERATIONS |
| 75 | bookingLast2Years\_VRM CLOUD |
| 76 | bookingLast2Years\_VSOM |
| 77 | bookingLast2Years\_VSPHERE |
| 78 | bookingLast2Years\_WANDERING WIFI |
| 79 | bookingLast2Years\_ZIMBRA |
| 80 | propensity\_high\_veryhigh\_vcf\_\_\_propensity |
| 81 | propensity\_high\_veryhigh\_vmc\_\_\_propensity |
| 82 | propensity\_high\_veryhigh\_vsom\_\_\_propensity |
| 83 | propensity\_high\_veryhigh\_vrealize\_suite\_\_\_propensity |
| 84 | propensity\_high\_veryhigh\_horizon\_\_\_propensity |
| 85 | propensity\_high\_veryhigh\_nsx\_\_\_propensity |
| 86 | propensity\_high\_veryhigh\_aww1\_\_\_propensity |
| 87 | propensity\_high\_veryhigh\_vsan\_\_\_propensity |
| 88 | propensity\_high\_veryhigh\_none |
| 89 | nsx\_target\_accounts\_19q3\_nsx\_target\_accounts |
| 90 | nsx\_target\_accounts\_19q4\_nsx\_target\_accounts |

## Extraction

***<EDA to fill in data extraction details, such as database name, table name, SQL, method, etc., for future reference until the manual uploads are replaced by automated runs>***

# **AAM Upload**

## File Format

Once the data has been extracted and generated in xlsx/csv format, it must be converted into AAM ingestible format (.sync/.overwrite).

 

The data must be structured in single column format with no headers.

* While the first attribute of every record is hashed email id, the second attribute is concatenated list of key-value pairs
* Hashed email id is separated from concatenated key-value pairs in the same record by a *tab*
* In a key-value pair, *key* is the column name (from extracted data) and *value* is the corresponding value in a row under that column
* Each record takes a *new line* and corresponds to each row in the xlsx data feed; hence, number of records in DMP upload feed should be same as that in xlsx data file minus the column header
* In a key-value pair, key is paired with its value using ‘=’ (equal). E.g. “ISO\_Country\_Name”=“United States”
* A key-value pair is concatenated with another key-value pair using ‘,’ (comma). E.g. “Company”=“WEST MICHIGAN HEART”,“Department”=“IT – ALL”
* All the keys and all the values are enclosed within “ “ (inverted commas)

Format of record: **Hash\_Email**< separator>**[Attributes]** (“tab” being used as a separator).

Format of **[Attributes]:** Attributes are concatenation of key-value pairs, using ‘=’ as pairing delimiter and ‘,’ as concatenation delimiter between two key-value pairs in the same record.

## File Name

**Naming Convention**: ftp\_dpm\_<DPID>[<\_DPID\_TARGET\_DATA\_OWNER>]\_<TIMESTAMP>(.sync|.overwrite)[.<SPLIT\_NUMBER>][.gz]

* **ftp\_dpm\_** prefix is mandatory for file uploads.
* **DPID** is Data Source ID from AAM. In this case of CRM data upload, we use 180427 (Id for data source “DMP - De-anonymization”).
* \_**DPID\_TARGET\_DATA\_OWNER**\_ is an optional value.
* **TIMESTAMP**, which is a 10-digit UTC UNIX timestamp in seconds, makes each file name unique.
* **.sync/.overwrite** are file extensions. “.overwrite” file is used in case of Full Feed Refresh. It will overwrite existing data in the AAM and should be used when we do full data refresh.

We shall use “.sync” format in case of dealing with incremental Feed.

For E.g. Below are the 3 feeds, 2 of them being .sync and 1 being .overwrite, are being uploaded on three successive days. Visitor123 is the user identifier and {a,b,c,d} are data associated to each ID in upload.

Day 1- .sync file contents: visitor123 = a,b,c

Day 2- .sync file contents: visitor123 = c,d,e

Day 3- file contains: visitor123 = a,b,c,d,e

Day 4- .overwrite file contents: visitor123 = d,e,f,g

Day 5- file contains: visitor123 = d,e,f,g

* **SPLIT\_NUMBER** is an optional field, to be used when a large file is split into smaller file to identify the order of split. Refer to the next section to decide whether file split is required.
* .gz format is a compressed format supported by S3 uploads in AAM.

Example of uncompressed file name: **ftp\_dpm\_180427\_1103623890.sync**.

Example of compressed file name: **ftp\_dpm\_180427\_1103623890.sync.gz**.

## File Size

* For an uncompressed file, maximum size is 5 GB and optimal size is 1 GB.
* For a compressed file, maximum size is 1 GB, and optimal size is 200-300 MB.
* For an uncompressed file larger than 1 GB, it is recommended to split it into multiple files, each adhering to recommended optimal size.

E.g. If a file **ftp\_dpm\_180427\_1103623890.sync** is around 2 GB, then it is recommended to split it into two equally-sized files before uploading. The names of the split files should be **ftp\_dpm\_180427\_1103623890.sync.1** and **ftp\_dpm\_180427\_1103623890.sync.2** respectively.

## S3 Details

Adobe Audience Manager supports file ingestion via S3 upload only. (FTP upload is no longer supported for an Inbound data file). AAM ingestible feed needs to be put in S3 bucket created for AAM Inbound data.

Here are the S3 credentials for CRM Inbound data file into AAM:

**S3 prefix**: s3://demdex-s2s-clients/vmware/crm/<yyyy-mm-dd>/

(s3://<AWS\_directory>/<partner\_name>/<yyyy-mm-dd>/)

**Secret Name**: vmware-inbound

**AWS Username**: vmware-inbound

**Access key Id**: AKIAIQ3GDRYAEP4NOMGQ

**Secret Access key**: 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

**Bucket folder**: /demdex-s2s-clients/vmware/crm/<yyyy-mm-dd>/

# **Appendix**

1. <https://marketing.adobe.com/resources/help/en_US/aam/c_onboarding_data.html>
2. <https://marketing.adobe.com/resources/help/en_US/aam/c_async.html>
3. <https://marketing.adobe.com/resources/help/en_US/aam/c_id_sync_in.html>
4. <https://marketing.adobe.com/resources/help/en_US/aam/index.html?f=c_file_based_id_sync>
5. <https://marketing.adobe.com/resources/help/en_US/aam/inbound-s3-filenames.html>
6. <https://marketing.adobe.com/resources/help/en_US/aam/inbound-file-contents.html>
7. <https://marketing.adobe.com/resources/help/en_US/aam/c_inbound_crm_data_ingestion.html>