

MAID Onboarding

02/01/2025 7:52 am EST

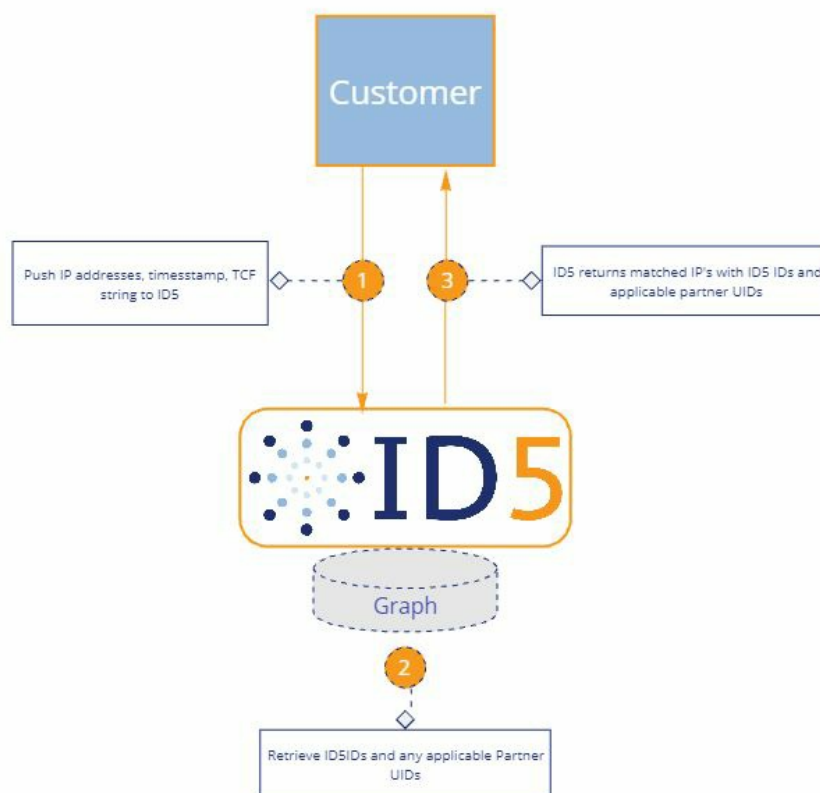
Overview

Customers may have a wealth of collected MAIDs that they would like to translate to online identifiers to facilitate the monetization of cookieless inventory via bid enrichment as well as enable audience building activation use cases. MAID onboarding allows customers to connect their MAIDs to ID5 IDs or other applicable Partner User IDs providing a privacy-safe view of their traffic.

How does the MAID onboarding work?

A customer can send ID5 a list of consented MAIDs and receive back a matching file with MAIDs associated with the ID5 IDs and any applicable Partner IDs. Depending on whether the customer desires scale, ID5 can optionally return the most recent ID5 ID seen for a MAID in the lookback window or indeed all ID5 IDs seen for a given MAID via the lookback window (configurable). The MAID onboarding can be done on a daily or weekly basis.

Solution Overview



1. The customer sends ID5 signal data: MAID, timestamp (optional), and consent string (optional and only for GDPR countries).
2. ID5 matches the signals against its global footprint. ID5 will not create ID5 IDs from MAIDs, rather simply look up ID5 IDs for a given set of MAIDs.
3. ID5 returns back to the customer a file with MAIDs matched to ID5 IDs and other Partner User IDs

- 4. The customer uses the IDs to enrich the bidstream, activate audiences or leverage them for other addressability use cases.

Expected input

Paths

The customer has to make sure the files containing the MAIDs to match are placed either

- into a separate directory every day with the naming convention `DATE=<ISO date>`
- into one single location which gets overwritten to (frequency of overwrite doesn't matter)

The files can have any name but the file name cannot be used to partition by day.

Examples:

Single directory, files get overwritten. Eg: `/maids/overwritten/all-maids_1.csv` , `/maids/overwritten/all-maids_2.csv` , ...

Separate directories partitioned by day. Eg: `/maids/DATE=2024-04-04/maids_1.csv` , `/maids/DATE=2024-04-04/maids_2.csv` , `/maids/DATE=2024-04-05/maids_1.csv` , ...

Any other partitioning scheme is not acceptable. Eg: `/maids/maids_2024_04_01.csv` , `/maids/maids_2024_04_02.csv` , ...

Format and content

The input format can be either

- Parquet
- CSV with headers, comma separated and optional quotes

We expect the following columns. Column names must be respected.

Field	Type	Description	Sample Value
maid	String	The MAID to onboard	<code>03516956-ab79-4801-bec3-fc366a05d795</code>
consent_string	String (optional, nullable)	The TCF V2 consent string	<code>CP9mGvtP9mGvtPIAAAENCZCAAAAAAAAAAAAAAAAAAAAAAAAAA.II7Nd_X_bX9...</code>
timestamp	String (optional, nullable)	The timestamp associated to the MAID in ISO-8601 format	<code>1994-11-05T13:15:30.114Z</code>

The timestamp is optional and is not part of the processing but is supported for user convenience: ID5 will copy the content of the timestamp column to the output.

Delivery Frequency

ID5 will deliver the matching file to the customer's S3 bucket on a daily or weekly basis. The matching happening on a given day will read and write using the date of the previous day / week. For example:

- On a daily match performed on the 5th October 2024, ID5 will read the data of `2024-10-04`
- On a weekly match performed on the 12th September 2024, ID5 will read the data of `2024-09-05`

The delivery date used to name the output files will correspond to that of the input files.

Matching File Content

The columns in the table below are included in the ID5 matching file returned to the customer.

The customer has the option to choose whether to match MAIDs to multiple ID5 IDs or just to the most recent ID5 ID. The customer will receive one data set (which may be comprise of one or multiple files with the same format) with the results of the matching process.

If some ID5 partner cookie matching has been requested, additional rows with ID5 partner user identifiers are delivered into the output data set. So the same maid / id5id pair can be present multiple times but every time with a different partner_id / partner_uid. The fields partner_id / partner_uid will be `null` if no matching ID5 partner identifier is found for that maid / id5id pair.

MAIDs that cannot be matched (there may be several reasons for that) will not be included in the output.

By default, we will only output the ID5 IDs matched that have had consent via the TCF V2 consent string.

By default, we will output the TCF V2 consent string associated with the ID5 ID, if the input does not already contain it.

Field	Type	Description	Sample Value
maid	String	The MAID related to the specific ID/IDs	03516956-ab79-4801-bec3-fc366a05d795
id5id	String	The Associated ID5 ID	ID5-bafePflpB9wsv5hp-ct4NcB5vIQD-G9MAja-Lm7f1g
partner_id	Long integer (optional, nullable)	The identifier of the ID5 Partner we matched with	264
partner_uid	String (optional, nullable)	The user ID according to the matched partner	4360196598752625072 (Example for The Trade Desk)
timestamp	String (optional, nullable)	The timestamp passed in the input	1994-11-05T13:15:30.114Z
consent_string	String (optional, nullable)	TCF V2 consent string associated with the ID5 ID that was matched	CP9mGvtP9mGvtPIAAAENCZCAAAAAAAAAAAAAAAAAAAAAA.II7Nd_X_bX9...

Output File Formats and Paths

ID5 supports two file formats when delivering the MAID matching file: CSV and Parquet. The output will be stored on a partitioned basis using the format:

`<configurable prefix>/DATE=<ISO date>/<some random filename>.[csv.gz|snappy.parquet]`

Examples:

- `/id5/output/DATE=2024-04-09/part-00000-tid-5623552081912624944-0b8fafa8-51a6-4f35-978c-2cd65b4957f3-85138-1-c000.snappy.parquet`
- `/id5/csv_output/DATE=2024-04-07/part-00000-tid-897652208191266533-ca6f3876-557d-4de3-bf76-32330850a2d8-78652-1-c000.csv.gz`
- Parquet
 - A binary, data efficient file format which will contain all of the supplied data fields in a flat file structure. The files are compressed using the `snappy` algorithm.
- CSV
 - The CSV file contains all of the supplied fields in a flat file structure. The separator is `,` (comma) and the quotes `"` are added only if required. The output files are compressed using the `gzip` algorithm.

CSV Example

In this example, we're using the CSV format without timestamps. First row in the example is the headers, second one an maid / id5id which additionally has an ID5 partner cookie match and the third one has no partner cookie match.

```
maid,id5id,partner_id,partner_uid
03516956-ab79-4801-bec3-fc366a05d795,ID5-ZHMOx88UySFiiNN0lpFvLeZ9yAyhi_P0xfGc9fw7yQ,264,
11cde477-3ab7-429b-ba3e-a851d5d35539,ID5-a9ddxNN39oI_silnqpcuG9bYnZrOcmfCt804_cNL0w,,
```