Working with AudienceDB and EventDB

This article describes how to access your AudienceDB and EventDB services.

These services are also offered directly by Tealium, but have to be licensed separately. As the service was not yet fully developed and does not cover all requirements, we at CDP decided to build it ourselves.
In order to use this service, the EventStore and AudienceStore services must first be activated. For additional information, contact your CDP account manager.

This article covers the following topics:

How It Works

AudienceDB and EventDB services are used to store your structured audience and event data in a Datalake that can be queried by SQL (Amazon S3, EMR and Athena). From Amazon Athena, you can then query and analyze the data directly using your preferred SQL client (e.g. SQL Workbench) or Business Intelligence (BI) tool (e.g. Tableau). In addition, this data can be analyzed with a Python (Jupyter) or Spark (Zeppelin) notebook. This can be done with Amazon Sagemaker.

When a service is activated, a datalake is created in Amazon S3 to store your data. For additional information, contact your account manager. The new datalake contains a table for each data type that it can store. Data associated with visit-level data is stored in tables that contain the expression "_event_" (e.g. dm1_f_event_events_map_visitorid_30day). Data associated with visitor-level data is stored in tables that contain the expression "_audience_" (e.g. dm1_f_audience_map_visitorid_30day). In addition to the tables, several reference tables are also created to make it easier to write queries.

Data Type

Data Type refers to the format in which the attribute *value* can be stored. Attributes come in a variety of types, from the basic number and string, to the more powerful tally and badge.

The following data types are supported:

Data Type	Description
Number	Stores numerical values such as order total, lifetime event count, or number of days since last visit.
String	Stores text values such as first/last name, address, favorite product, and page name.
Boolean	Stores only one of two values: 'true' and 'false'. Boolean can be used for indicating the status of a visitor action or visit.
Array of Numbers	Stores multiple numeric values as an array. The array may contain unique or duplicate numeric values.
Array of Strings	Stores multiple string values as an array. The array may contain unique or duplicate string values.
Array of Booleans	Stores multiple Boolean (true/false) values as an array. The array may contain unique or duplicate Boolean values.

Tally	Stores one or more key-value pairs.
Set of Strings	Stores a collection of unique string values as a Set.
Date	Stores the date of a visitor event or particular visit.
Funnel	Tracks the status of every step in a multi-step event, such as payment funnel and registration.
Timeline	Records all occurrences of a visitor action within a time range.
Badge	Assigns a visual mark or symbol to visitors who meet certain criteria or browsing behavior.
Visitor ID	Stores visitor's unique traits as their secondary identifier.

Tables, References

The columns in the tables are named according to the attribute type and the external attribute name. For example, if you have a badge attribute with an internal id of "30", the table will contain a column named "Visitor bild". Tables and references of data are created to make it easier to write queries.

- Since the number of attributes can change daily, a Generic Model was used to avoid creating a separate column for each attribute.
- For certain categories of attributes columns with Complex Datatypes (e.g. Maps) were used.
  Example: "audiences_map"
- Since the attributes can no longer be recognized by the column name, but are processed within the complex datatype, reference tables were created to determine the possible attributes.
  Example: "dm1_d_badges"

AudienceDB Mart Tables

The following describes the AudienceDB table types used for audience data:

Table Type and Description	Table Name	Attributes
Last Visit Lists for each user of the last 30 days Current visit Set of Strings attributes, tally attributes, Badge attributes, Set of Strings attributes, Dates attributes, Flags attributes, audiences etc.	`dm1_f_audience_map_visitorid_30day`	date_str visitor_id last_visit_date expire_date audiences_map audiences_map_cnt badges_map badges_map_cnt properties_map properties_map_cnt dates_map dates_map_cnt flags_map flags_map_cnt metrics_map metrics_map_cnt
Count Visitor per Badge Derived view of dm1_f_audience_map_visitorid_30day that determines the number of visitors per badge.	`dm1_f_audience_badges_usage_visitorid_30day_v`	effectiv_date badge visitor_cnt
Reference Badges List of all Badges	`dm1_d_badges`	`badge`
Reference Dates List of all Dates	`dm1_d_dates`	dates
Reference Flags List of all Flags	`dm1_d_flags`	flags
Reference Metrics List of all Metrics	`dm1_d_metrics`	metrics

Reference Properties List of all Properties	dm1_d_properties	properties
Reference Metric-Sets List of all Metric-Sets		metric_sets
Reference Property-Sets List of all Property-Sets		property_sets

AudienceDB Core Tables

The following describes the AudienceDB table types used for audience data in the Core Schema:

Table Type and Description	Table Name	Attributes
Last Visit Lists for each user of the current day Map of the last Audience of the current day for each user	co_tl_visitor_s_audience_audiences	
Last Visit Lists for each user of the current day Map of the last Badges of the current day for each user	co_tl_visitor_s_audience_badges	
Last Visit Lists for each user of the current day Map of the last Dates of the current day for each user	co_tl_visitor_s_audience_dates	
Last Visit Lists for each user of the current day Map of the last Flags of the current day for each user	co_tl_visitor_s_audience_flags	
Last Visit Lists for each user of the current day Map of the last Metric-Sets of the current day for each user	co_tl_visitor_s_audience_metric_sets	
Last Visit Lists for each user of the current day Map of the last Metrics of the current day for each user	co_tl_visitor_s_audience_metrics	
Last Visit Lists for each user of the current day Map of the last Properties of the current day for each user	co_tl_visitor_s_audience_properties	
Last Visit Lists for each user of the current day Map of the last Property-Sets of the current day for each user	co_tl_visitor_s_audience_property_sets	

Tables Containing Event Data

For tables containing event data, event data includes event attributes for all events in the event feed. Table columns are named according to the attribute type and name, with only some attributes referencing internal ID's. Standard Universal Data Object (UDO) variables are named with a "udo_" prefix and most column names match their corresponding attribute names eg. "udo_event_name".

Event data coming from the Tealium collect tag also includes information about which tags executed on the page and page performance metrics.

For additional information, see the Live Events and Feeds.

EventDB Mart Tables

The following describes the AudienceDB table types used for audience data:

Table Type and Description	Table Name	Attributes
Each Event for each User of the last 30 days	`dm1_f_event_events_map_visitorid_30day`	event_id, event_time, visitor_id, visit_id, pageurl_full_url, referrerurl_full_url, event_map, event_map_cnt
Reference Event Keys List of all Event-Keys	`dm1_d_event_key`	`event_key`

EventDB Core Tables

The following describes the AudienceDB table types used for audience data:

Table Type and Description	Table Name	Attributes
All events for each user of the current day Map of the all Events of the current day for each user	`co_tl_visitor_s_event_events`	

Database Credentials

You must use a third-party tool with AWS Athena support to connect to your database. These tools require authentication credentials to connect, which are provided by your CDP account manager and are organized in the Password-Tool.

Getting Database Authentication Credentials

Database access is created via AWS IAM. Here the corresponding rights for AWS Athena and AWS s3 must be deposited. In an existing cluster, corresponding roles already exist.