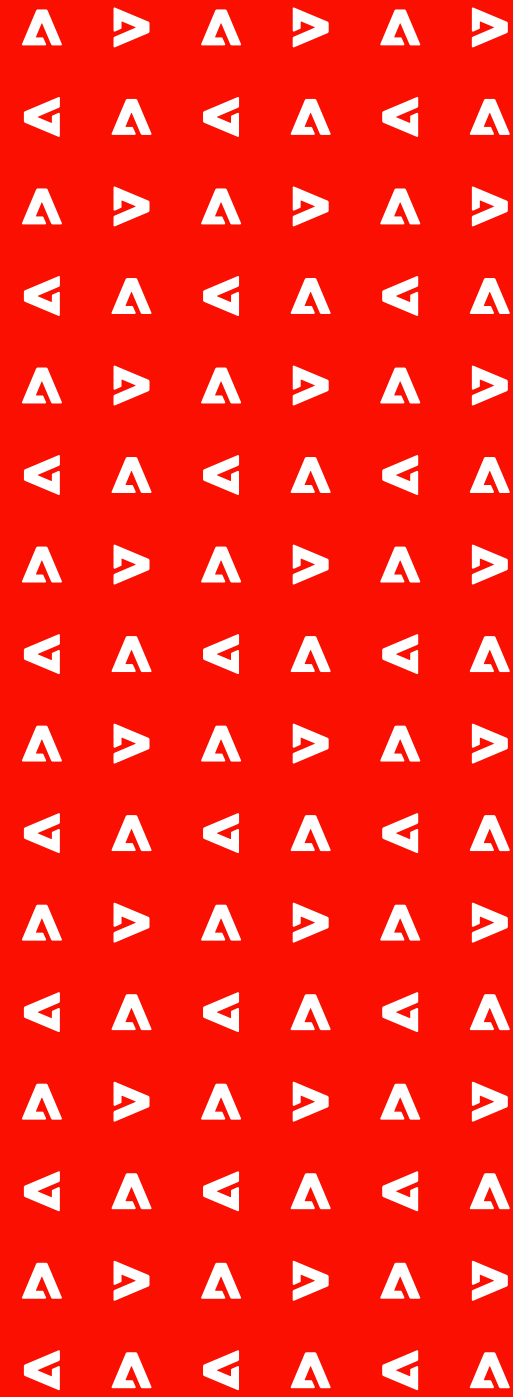




Data Distiller Workshop

August 2024

Adobe Technical Validation Team



Data Distiller Overview

- Lightly customized workshop showcasing Data Distiller features
- Guided hands-on exercises
- Light technical depth
- Leave-behinds include session recordings and presentation decks
- Post-lab survey

Hands-on Lab Agenda

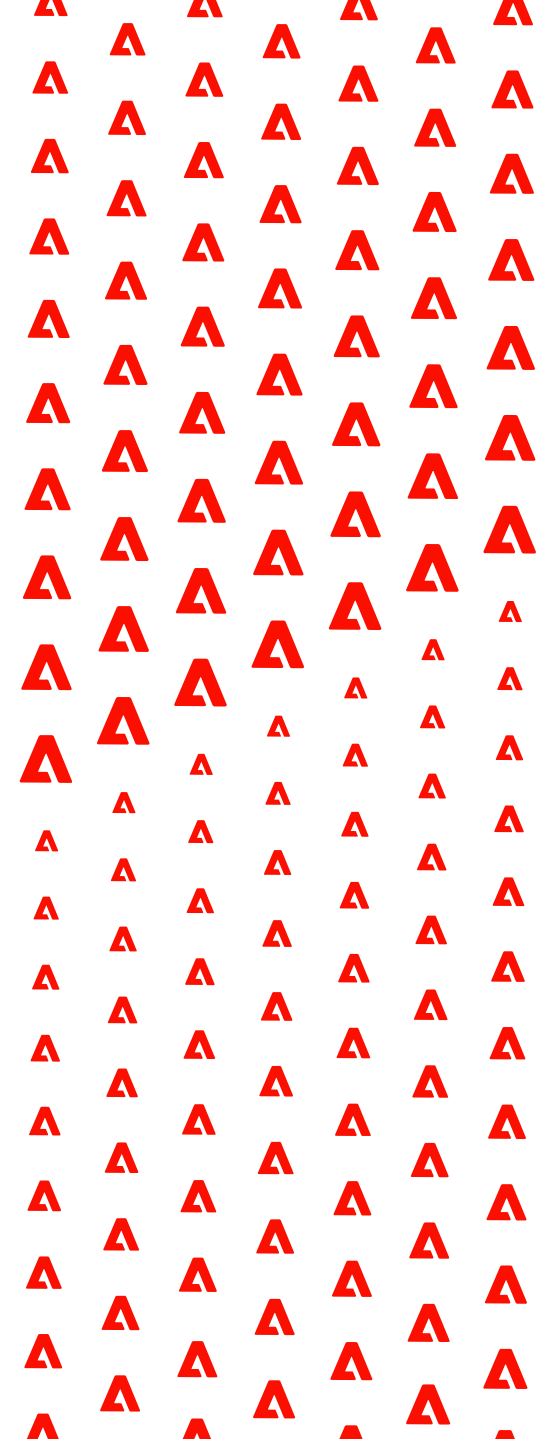
Adobe Experience Platform Day 1 Agenda	
Intros & Access	15 mins
Data Distiller	15 mins
Query Editor UI	30 mins
Client Connection	15 mins
Data Collection, Schemas, Datasets	30 mins
Wrap-up + Q&A	10 mins

Reminders as we get started

- ✓ This lab is lightly customized to showcase the capabilities of Data Distiller and its related application services
- ✓ This lab is conducted in a demo environment and does not reflect performance of the application in production.
 - ✓ Instance will not contain data representative of a full-scale production system
 - ✓ Will use fictional data
- ✓ This session will be recorded. The recording will be shared after the lab.



Introductions



Adobe Technical Validation Team

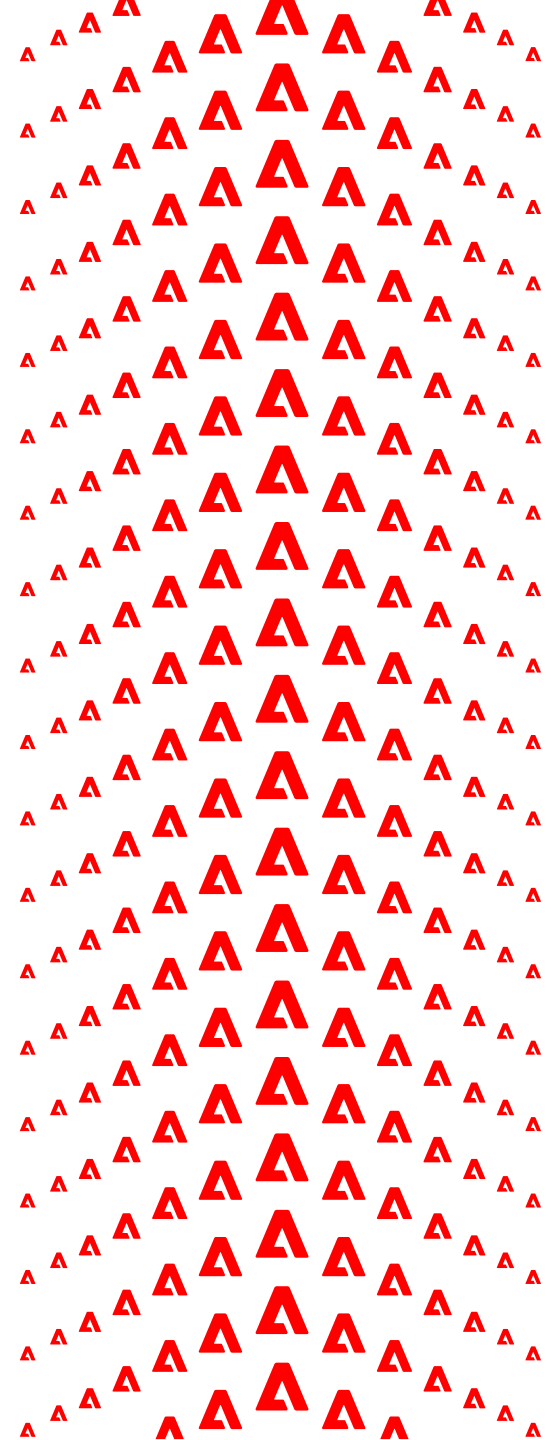


Jamar Smith
Principal - Technical Validation
gesmith@adobe.com



Andrew Rex
Sr. Practice Lead - Technical Validation
arex@adobe.com

Lab Housekeeping



Archbee

Login information for this hands-on lab and step by step instructions:

<https://tvdocs.adobe.com/data-distiller-august-2024>

Password: DataDistillerLab08!

Data Distiller - M...

☾

Ask a question... 36 K

Login Information (cloned)

Login Information

Table of contents

data-distiller

Connect Clients

Query Editor UI

getting-started

Lab Content

schemas-and-datasets

Dataset and Data Ingestion

Schemas

workshop-takeaways

Workshop Content

Login Information (cloned)

🕒 3min

The Hands-on Lab URL is:

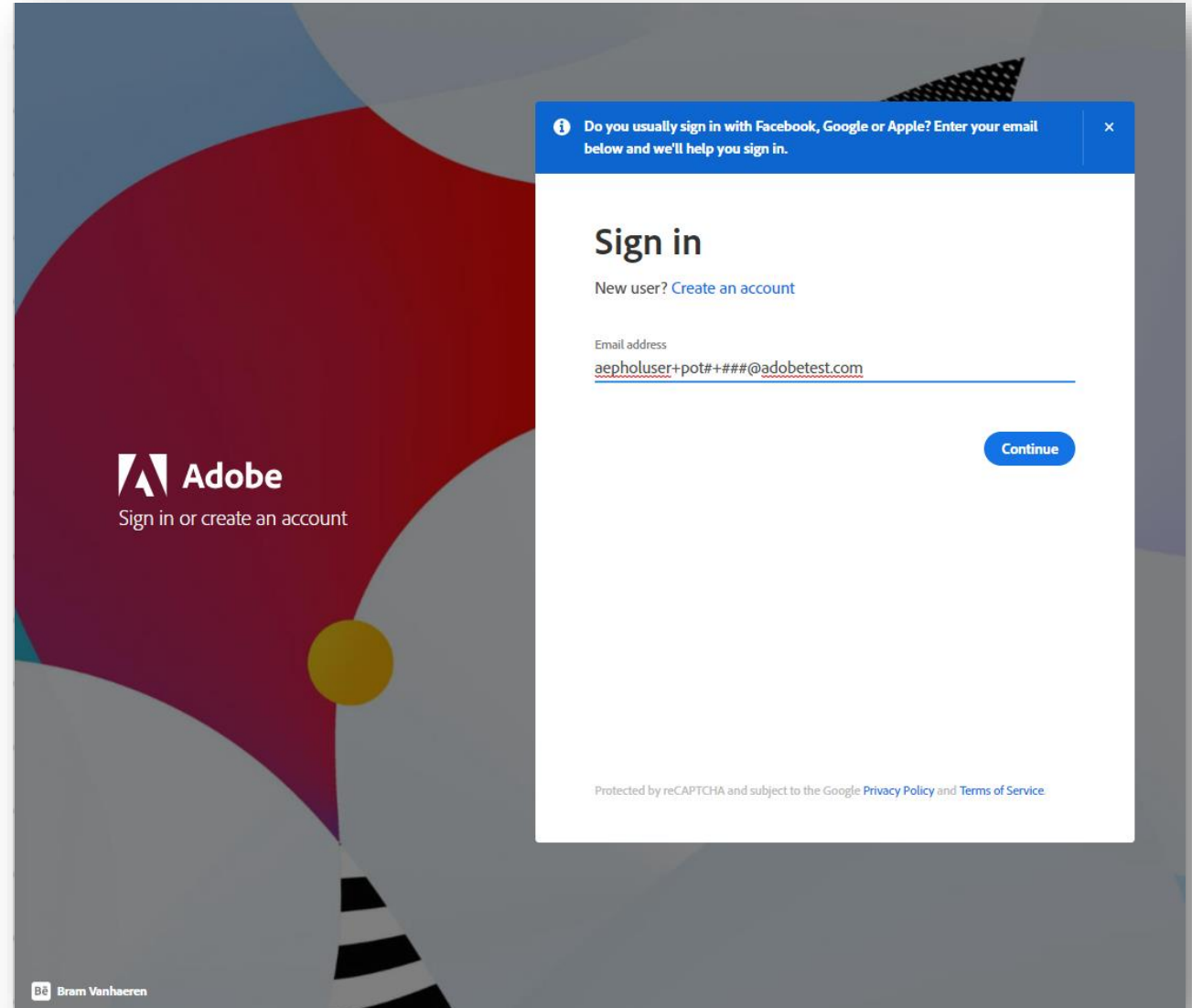
<https://platform.adobe.com/>

After navigating to the URL above, you must see a screen like this (If you see anything else, then make sure you sign-out, and try again):

Lab Housekeeping



- Attendee number
- Log in to Adobe Experience Platform
 - <https://platform.adobe.com/>
 - Email address:
datadistiller08u01@pot5.aephandsnolabs.com
 - Password: **DataDistillerLab08!**
 - DEVELOPMENT
- Sandbox: Data Distiller 1

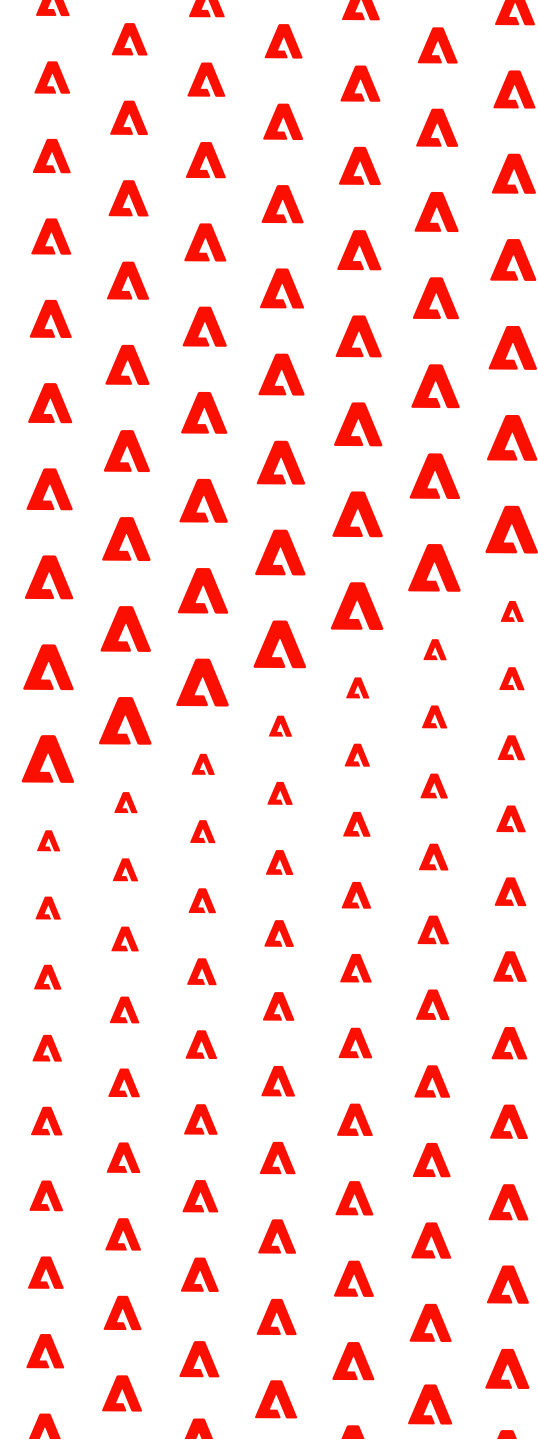




Adobe Experience Platform Foundations

Overview | Data Ingestion | Schemas | Datasets | Data Distiller

Adobe Experience Platform Overview



Adobe Experience Platform Overview

How It Works



Select the data sources you want to connect and unify



Map the data using Adobe's industry-leading experience data model



Govern and segment audiences of unified people and account profiles



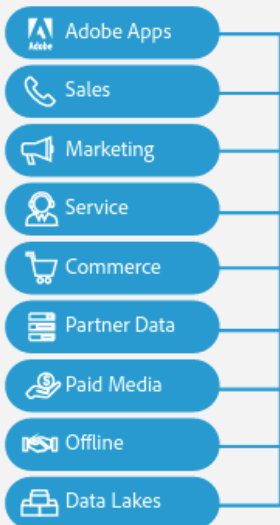
Select and connect the data destinations you want to activate to



Activate audiences across destinations for engagement, reporting and other B2C, B2B and hybrid use cases

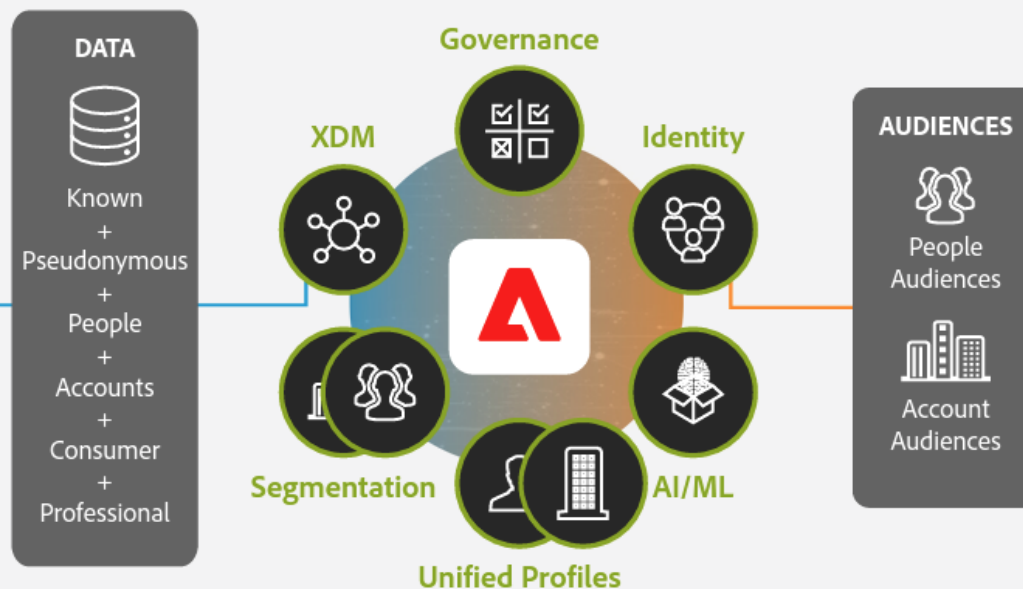
Sources

B2C & B2B Data Sources



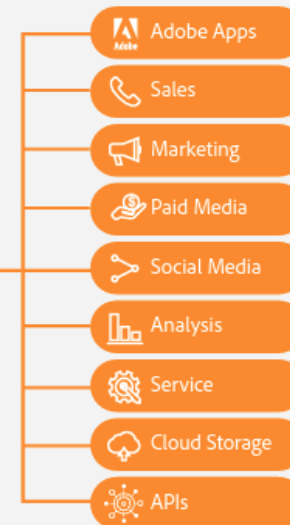
Adobe Experience Platform

B2C, B2B & Hybrid Profiles

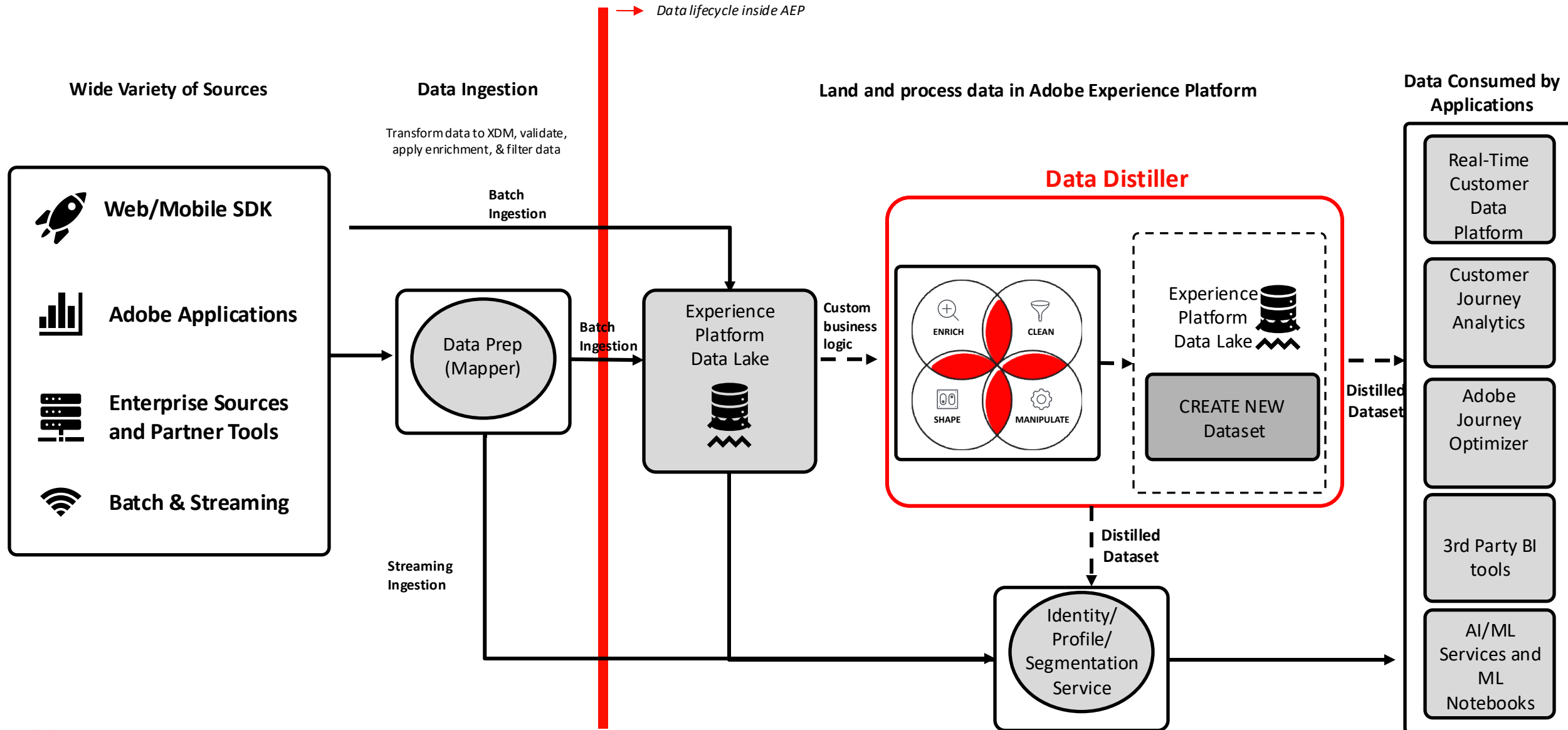


Destinations

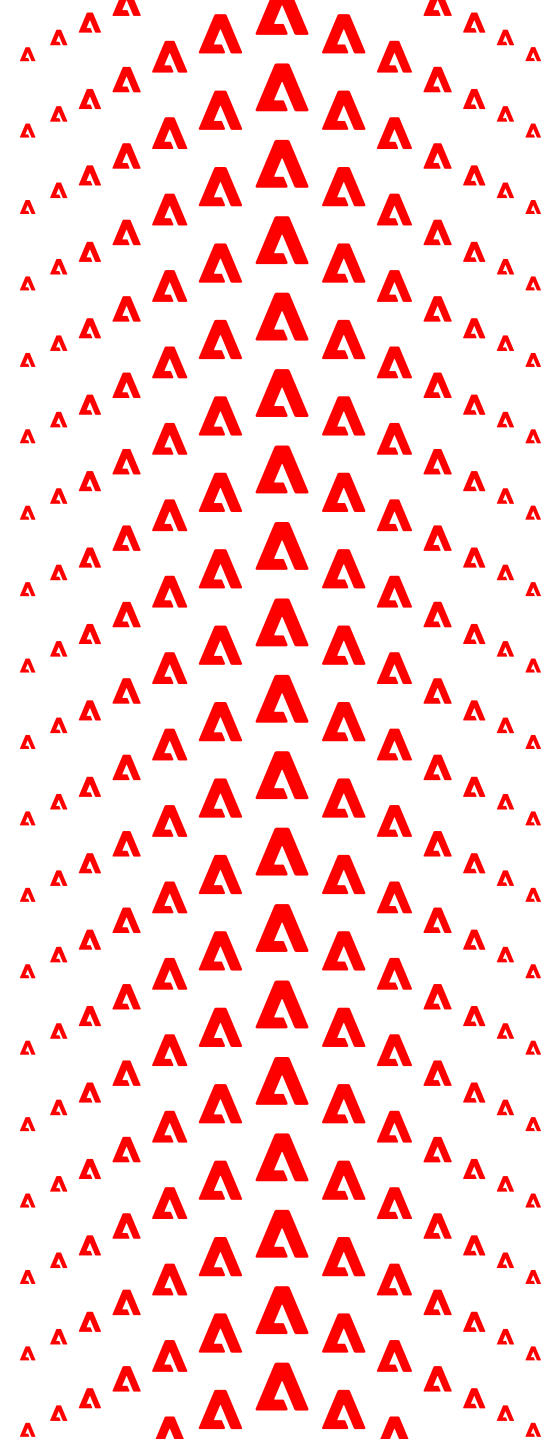
B2C & B2C Activation



Adobe Experience Platform Data Value Chain



Data Distiller



Adobe Experience Platform Data Distiller Overview



Experience Data



Adobe Experience Platform



Data Distiller Features



Call Center



Point Of Sale



CRM



Mobile



Web

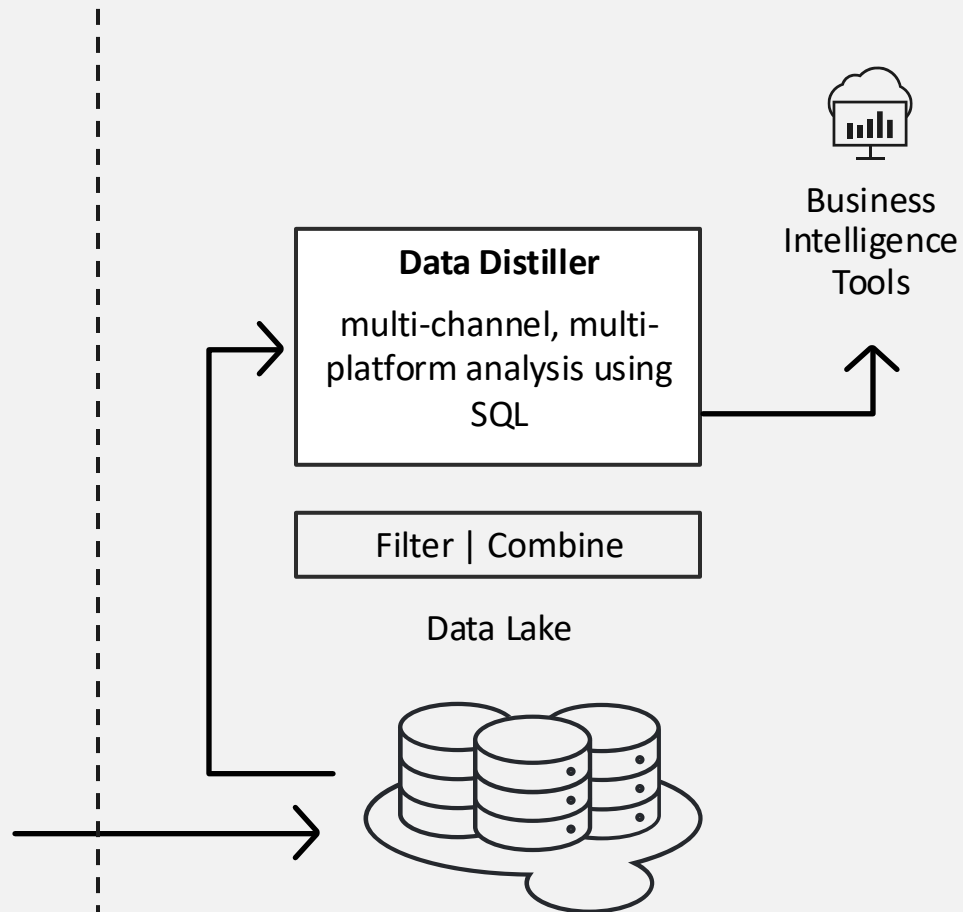
Campaign

Audience Manager

Analytics

Target

Advertising Cloud



- Adobe-defined functions
- Join any datasets in the Data Lake
- Capture query results as new dataset
- UI or API
- PostgreSQL CLI & other BI tools
- Scheduled Queries

Use Case Categories



Clean

- Filter, Standardize and Normalize data based on user defined rules



Shape

- Reshape ingested data sets based on join, lookup or transpose logic
- Shaping data going into identity graph



Manipulate

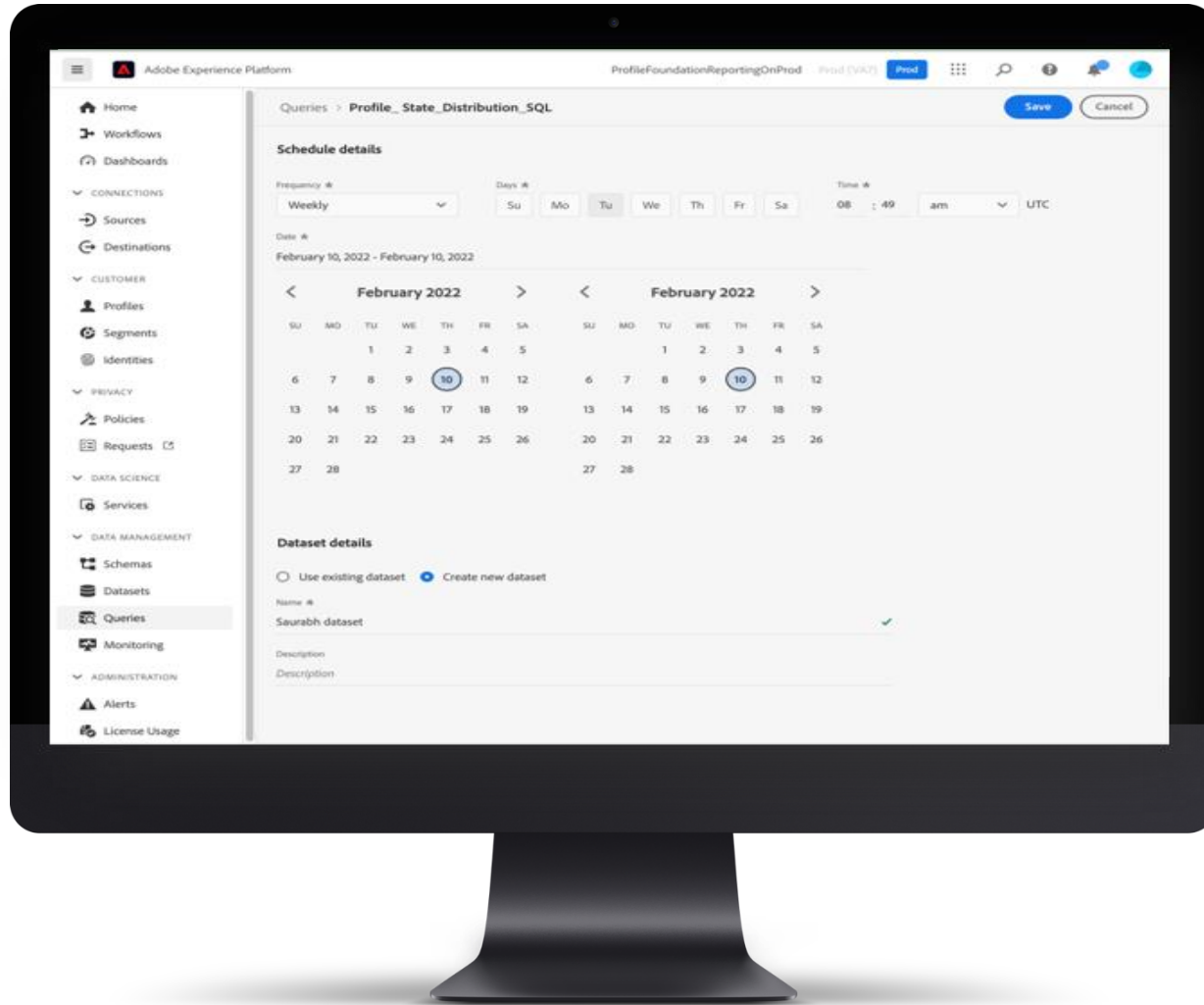
- Perform data aggregation, and window operations for BI reporting and audience/campaign activation use cases



Enrich

- Derive key additional attributes for downstream audience/campaign activation or analysis use cases e.g. brand/product affinity, LTV, etc

Data Distiller Capabilities



Powerful SQL Based data processing engine

- Scalable compute infrastructure
- SQL Functionality focused on customer experience use cases
- Standard spark functions
- SQL Adobe-Defined Functions for time series manipulations such as pathing, window functions
- Templates
- Anonymous Block
- Snapshot Incremental Processing
- Sampling



Operationalization of data processing workflows

- Easy automation and scheduling,
- Monitoring and Alerting
- License Usage Monitoring



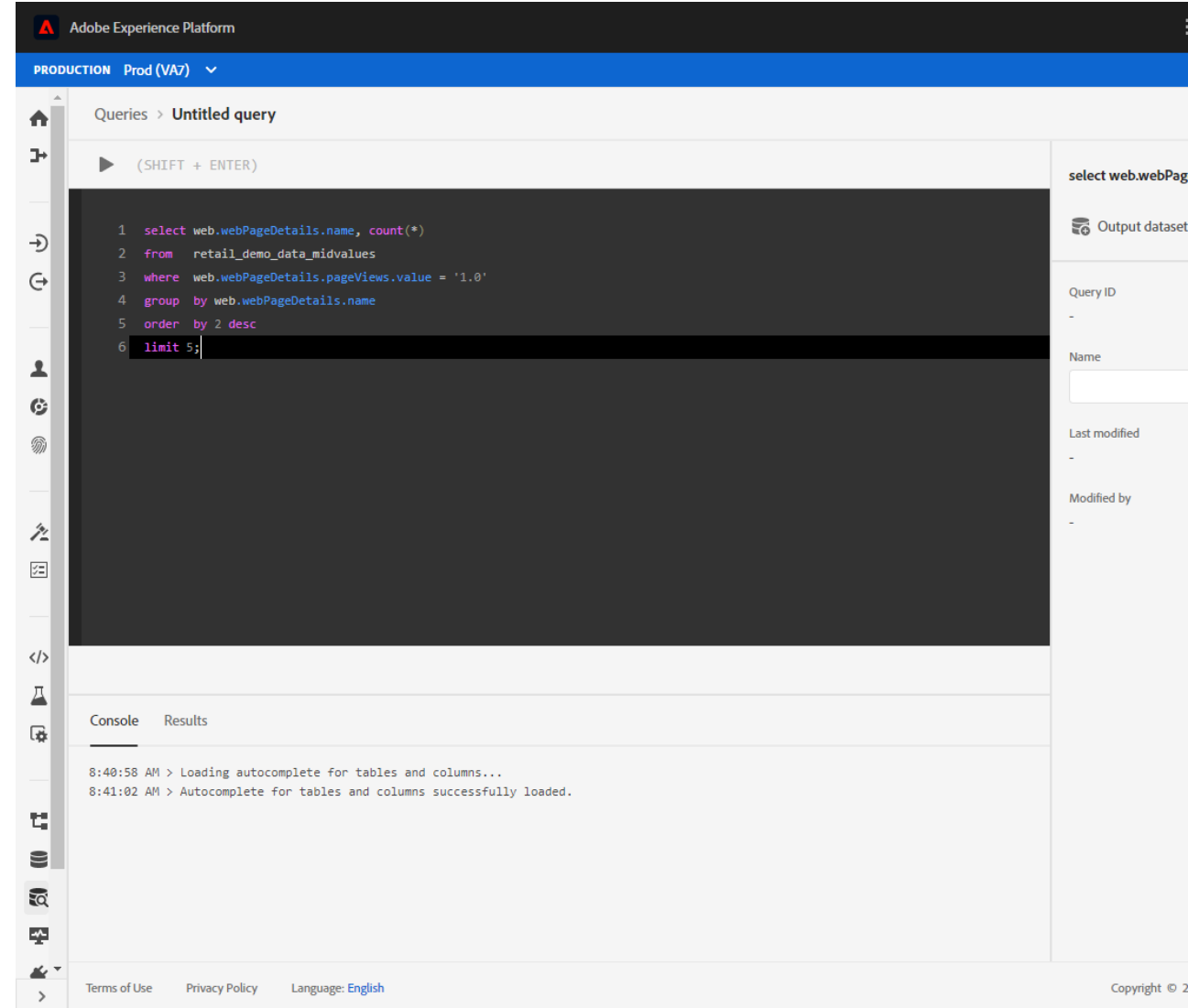
Integration workflows for delivering extended insights

- BI Connectivity
- Extend Unified Profiles/AJO reporting data models
- 3rd Party Clients

Query UI

Intended to be used as sandbox to test queries before executing them on wider dataset.

- Automatic syntax highlighting
- SQL keyword autocomplete
- Table and field autocomplete
- Error detection



Client Connection

Used to develop queries, perform analysis and exploration.

- Connect to Query Service outside of AEP UI
- Command-line interface
- Installed along with PostgreSQL

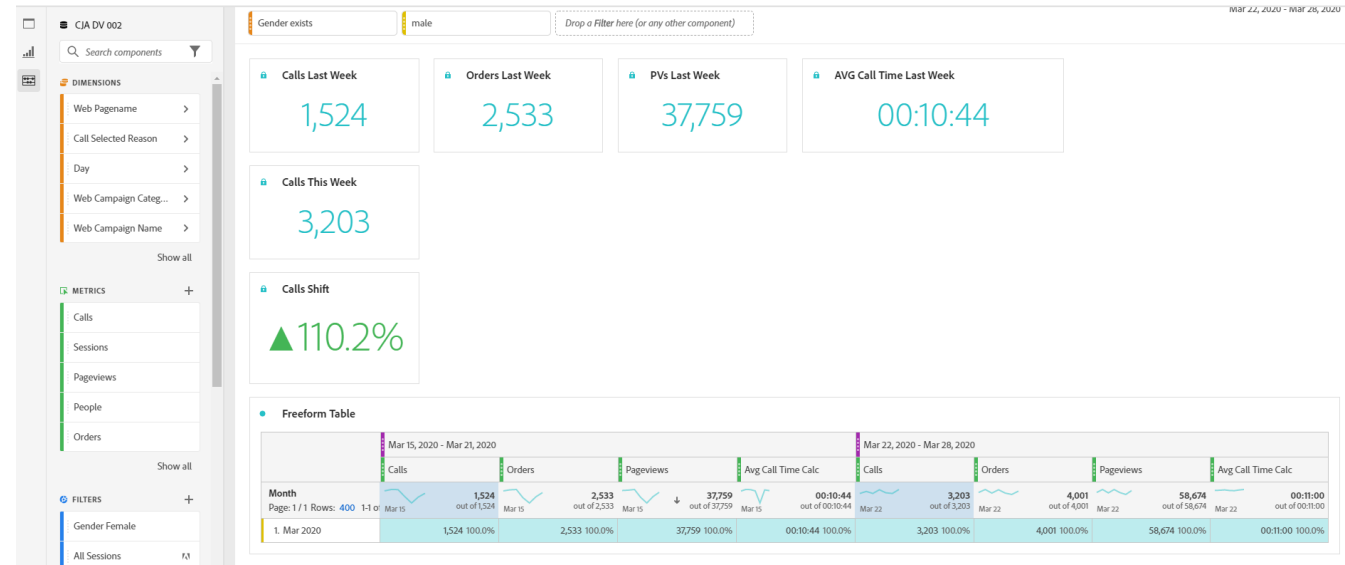
```
Command Prompt - psql "sslmode=require host=adobeamericaspot5.platform-query.adobe.io port=80 db=
Microsoft Windows [Version 10.0.19042.928]
(c) Microsoft Corporation. All rights reserved.

C:\Users\dumiller>psql "sslmode=require host=adobeamericaspot5.platform-qu
62A55DE67F540A495CAC@AdobeOrg password=eNqrVkrMSVeyUgoKNjI1U9JRqjAtBfIyc4v
J6nj3JFW2kDLpUuBQmnVqFLk2NPWJYmj20kWEP--Tlr2AfZh_TQ-czwzZ2b8bnBmnB10gFHouW
tOBQqawPAgeakbrWaCuhOWIowiPkBrE_Okf4PHBR4m17dP4zYSKHm6EmS0UUaOr7ytjJgud7aF
yEMYB27vliAlF1VP2CpVy7PBgJfSqohjky5tITa8sqko03hwog_2zmBSpsXd1Qili2e1nI23ky
Cb6fzS-nu-R1ulu-TUYj1N0fn8fJdPG4X5bz8W0yMj60XPVad2oJpTpvpsQLVBolUmOnIrsr-e
0Hnku2Hs0BaCSA8acGDFOXEs4gc5Q3mEfRScBl0Krp_HeeiGPoo0AIEaNYAzXmk8CjzUnS4PFU
2bqtaId9c9aN2AFVwB4awVqqLkSl4KC-8RpRgPyKmqdemf3sudA7yj9tW-9rAxsuVfN6qqtleq
QjW1Ca3_2qwkmsKNItUnZJ62rvtCXXRgGVff52N8_AEkGidu.eNoBAAH__oIc-o11QMxvExl0r
jwxqA-c1hKGou03BL6ctI_m8k-5MEZjbVNTEPHa_WitiCg_vzaYG5pAwfaY_YiAujbVJRCv-mi
JHn8HHEPilhKLEmlr1VKaaYrNNSZkys7H1HbinW3DzDjdj50pP_JQqHwI37ZpG6uXgL8sP6Gct
A0fXT9UKwsdzgWvjYXfJbjm933qshLZ4u2aFnhiFQLION"
psql (11.6, server 2.0.0-20210506230610-8bb8b7ede3)
WARNING: Console code page (437) differs from Windows code page (1252)
         8-bit characters might not work correctly. See psql reference
         page "Notes for Windows users" for details.
SSL connection (protocol: TLSv1.3, cipher: TLS_AES_256_GCM_SHA384, bits: 2
Type "help" for help.

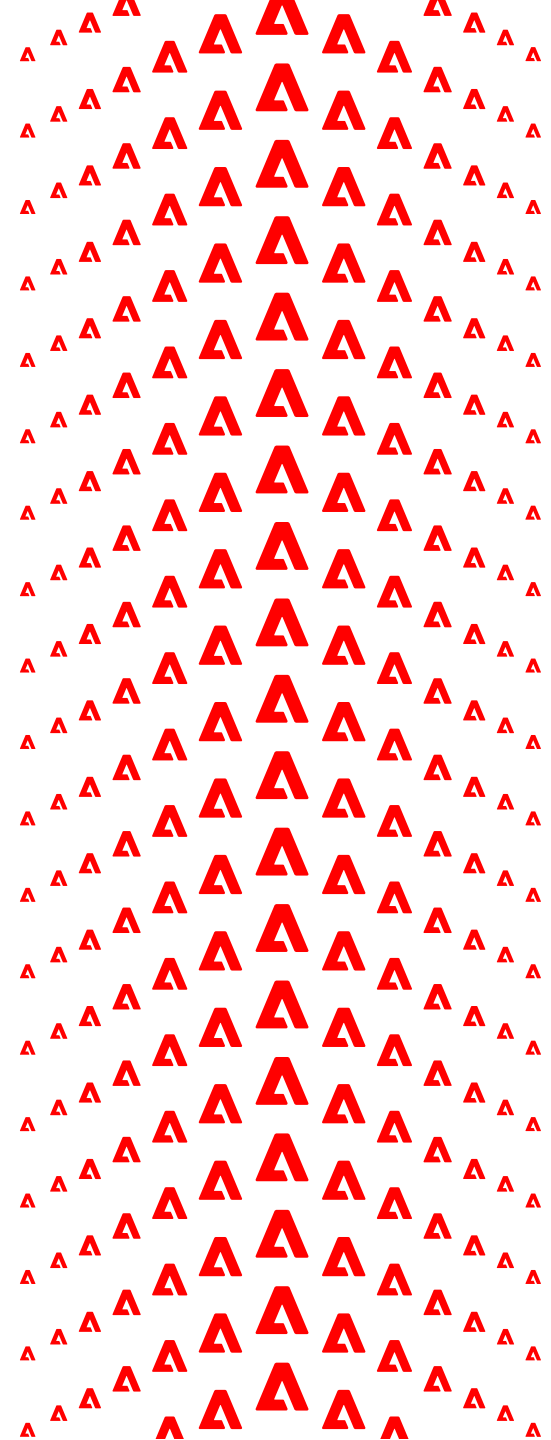
prod:all=>
```

Data Distiller | Key Use Cases

- Join Experience Event datasets with other customer data to create reporting or BI inputs
- Customer Journey reporting and insights across channels, devices and environments
- Better understanding of omni-channel attribution
- Unify datasets under a single Identity for Intelligent Services
- Data collection debugging



Data Ingestion



Open and Extensible Data Ingestion

Easily configurable integrations allowing customers to ingest data from a range of applications including cloud storage, advertising, sales, and CRM.

Centralize data ingestion from disparate sources with prebuilt connectors.

SOURCES CATALOG

Choose from dozens of sources connectors including advertising, cloud storage, CRM, databases, and marketing automation tools.

REAL-TIME STREAMING & BATCH DATA COLLECTION

Stream data in real time to and from Adobe and non-Adobe client- and server-side devices for faster insights, quicker time-to-market, and more timely responses to customer engagement.

USER INTERFACE & APIs

Configure new sources directly in the UI or leverage RESTful APIs and SDKs to build bespoke connections

Sources Viewer

Sources

Catalog Accounts Dataflows System View

☒ All sources

☐ My sources

CATEGORIES

Adobe applications

Advertising

Cloud storage

CRM

Customer success

Databases

eCommerce

Search

Google Google Cloud Storage

Configure



Google PubSub

Configure



Oracle Object Storage

Configure



SFTP

Configure



Out of the Box
Connections



Catalog
View



Data Flows



Monitoring



Run Notifications



API Accessible

Data Ingestion

Multiple methods are available for data ingestion.



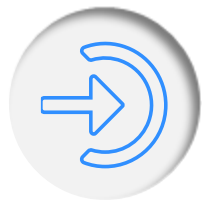
Batch Ingestion



Streaming Ingestion

Data Ingestion | Sources

Easily configurable integrations allow data ingestion from a range of applications including advertising, cloud storage, CRM, sales, etc.



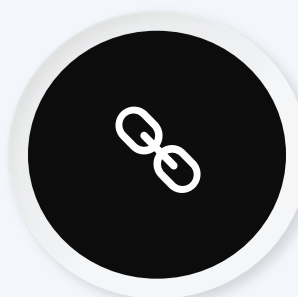
CUSTOMER ENGAGEMENT
(CRM, Customer Service,
Marketing Automation...)



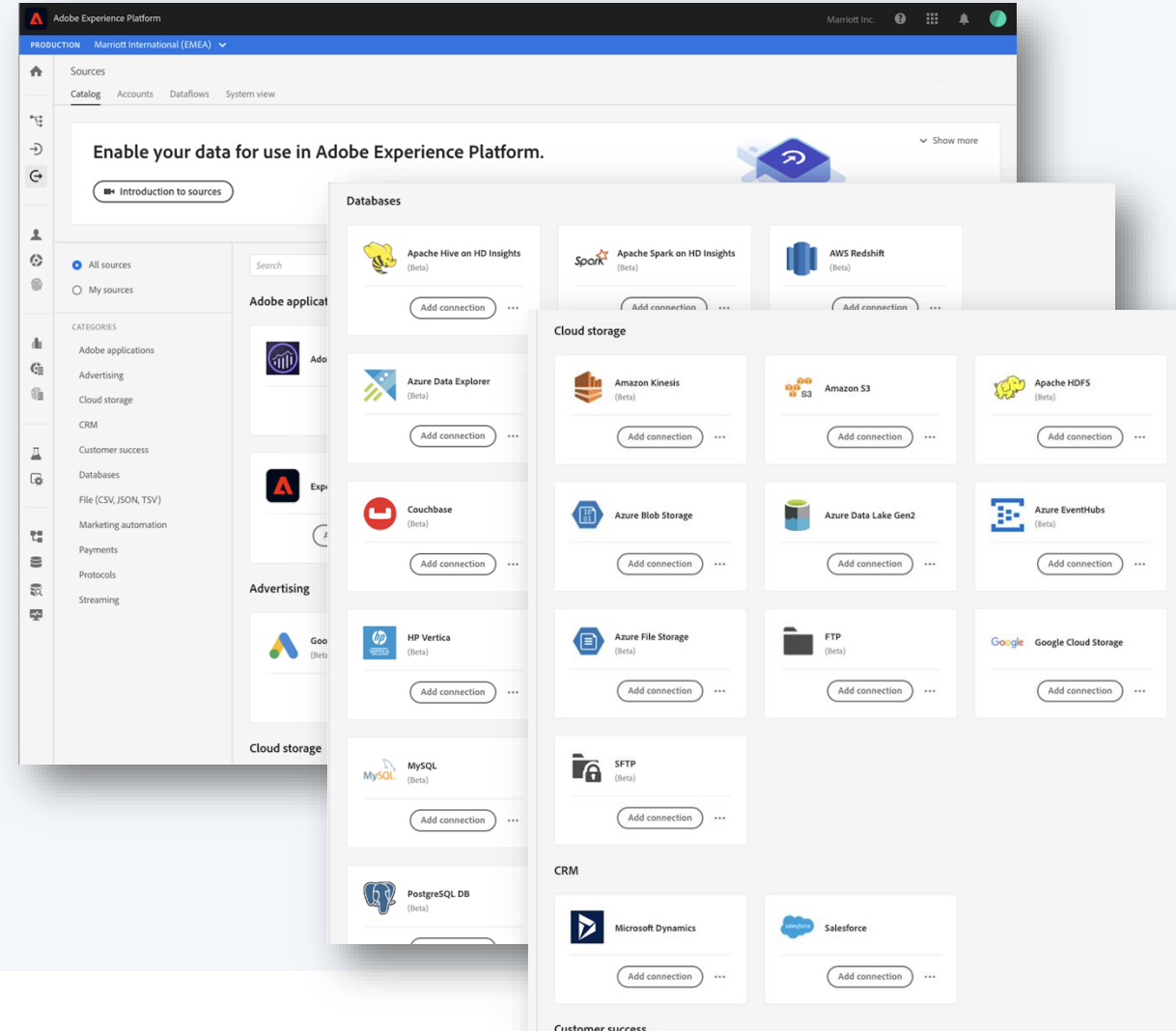
FILE STORES & DATA WAREHOUSES
(Amazon S3, Google Cloud Storage,
Azure Blob, Google BigQuery, Amazon
RedShift)



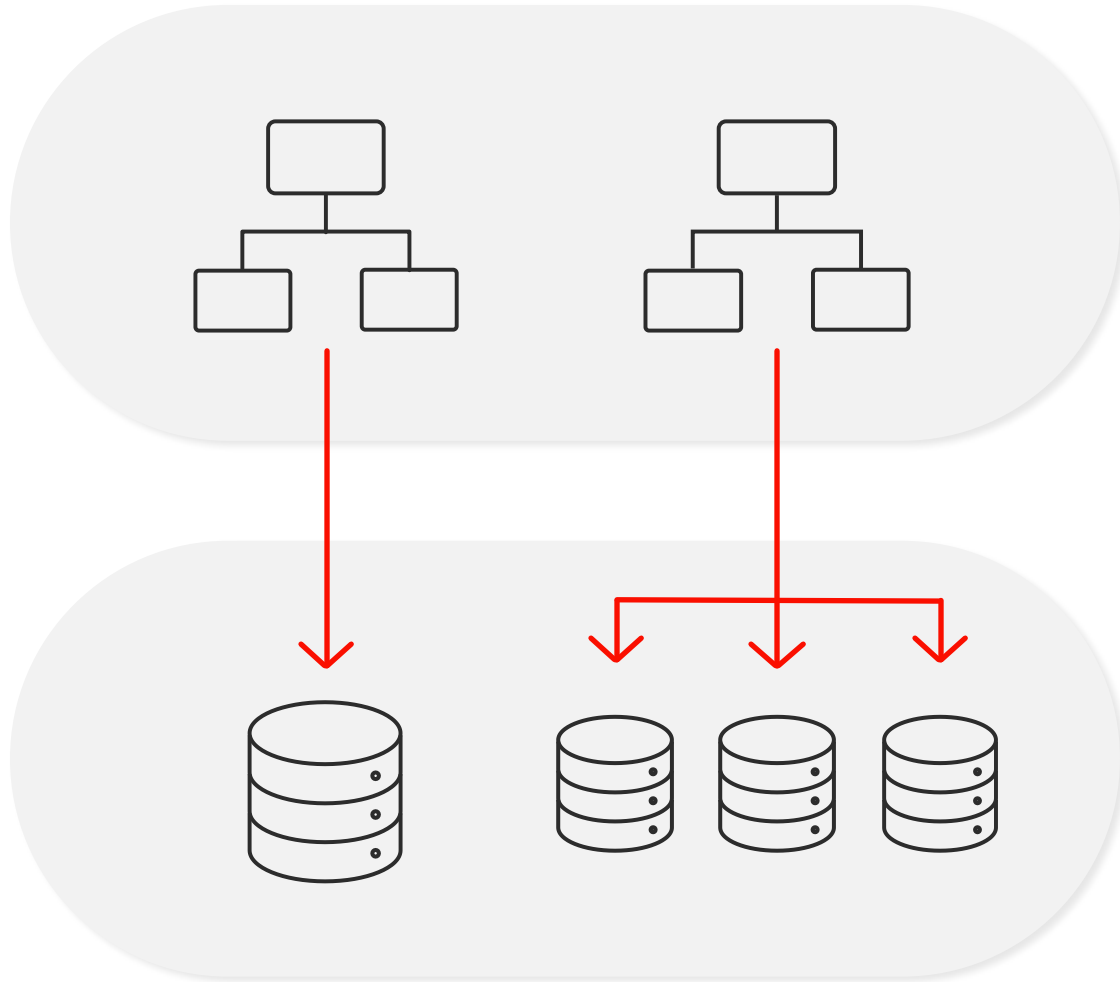
STREAMING
(Amazon Kinesis, Azure
EventHub, Google PubSub)



ETL PARTNERS
(Microsoft SQL, SoftwareAG)
Representing 200+ Sources



Datasets vs. Schemas



Schema

A set of definitions that describe the blueprint of the data, including data types, validations, and relationships

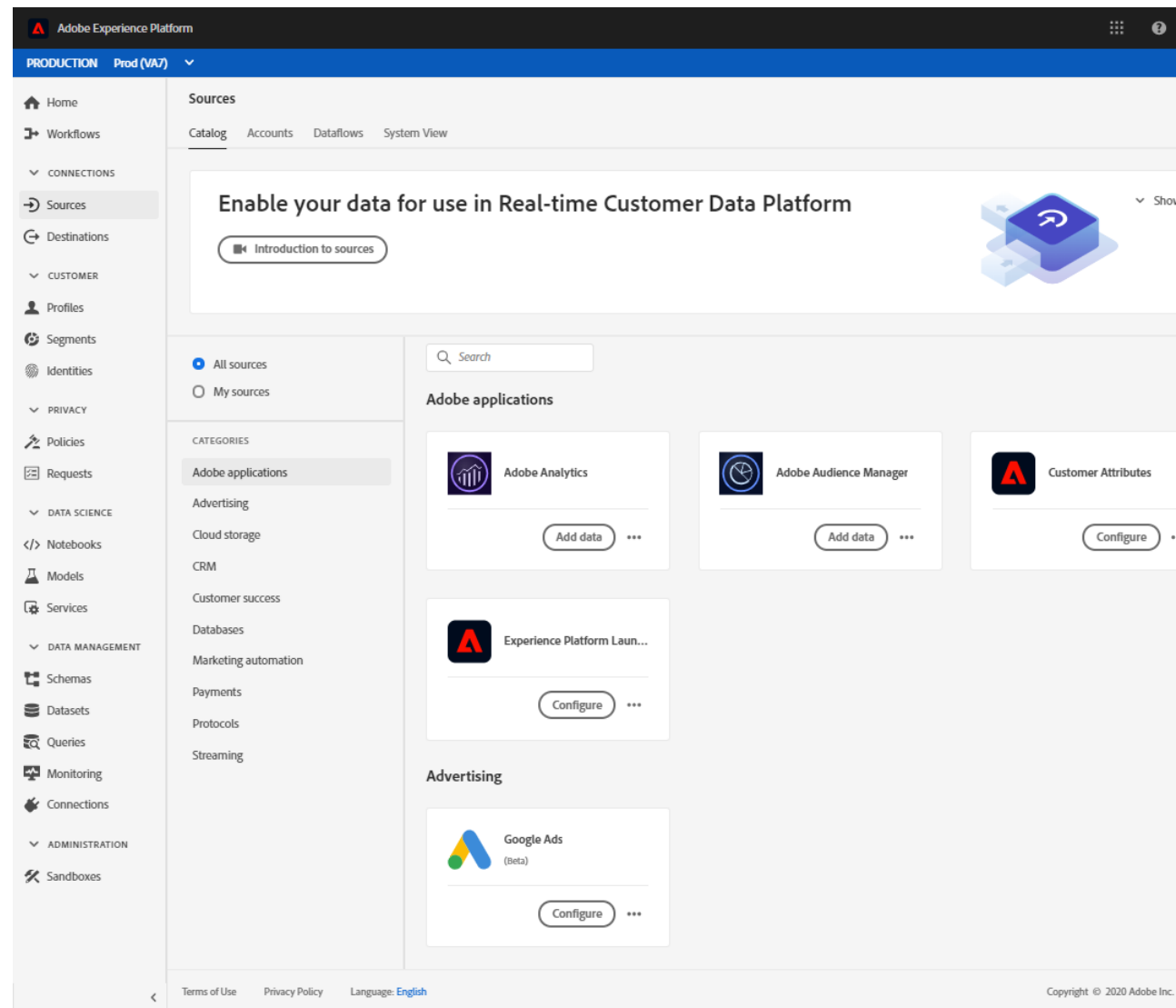
Dataset

An observable collection of data defined by a schema

Data Ingestion | Explore UI



- Environment PRODUCTION Prod
- Explore Sources
 - Catalog | Accounts | Dataflows | System View
 - Streaming
- Explore Workflows
 - Map CSV to XDM schema
- Explore Monitoring
 - Batch end-to-end | Streaming end-to-end

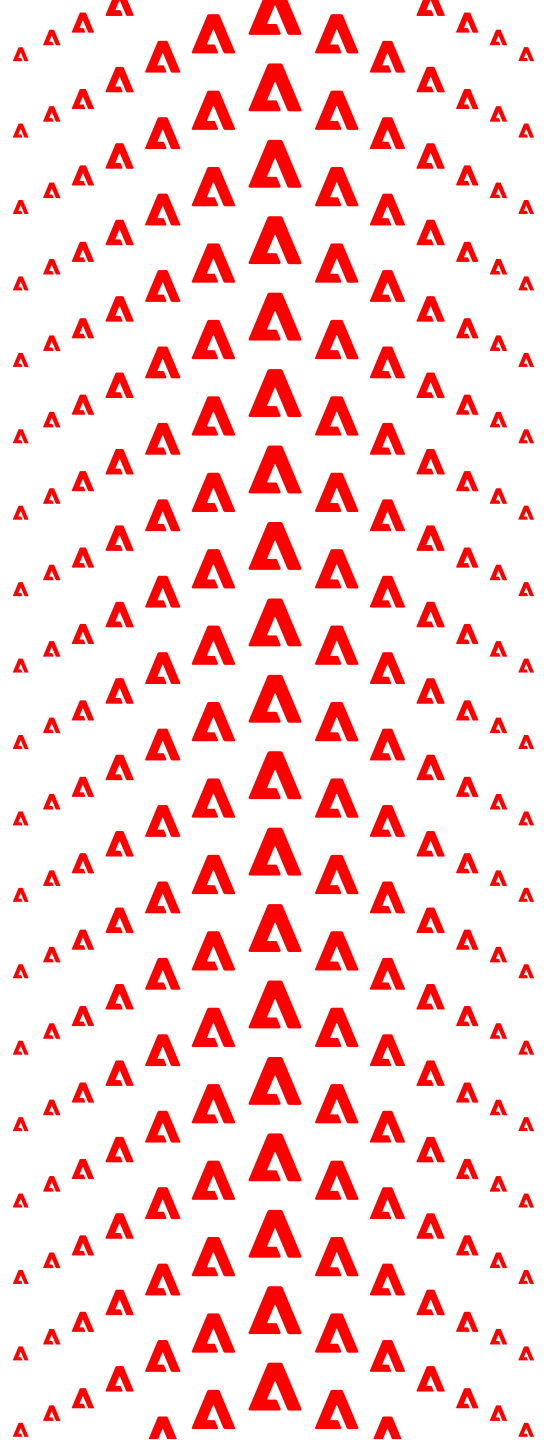


Schemas | Hands-on Exercise



- Environment DEVELOPMENT Sandbox##
- Create an ExperienceEvent schema
 - Add existing field groups
 - Create a field group (mixin)
 - Create Object data type
 - Add String and Integer fields

APIs



APIs

Adobe Experience Platform is built on open APIs.

- Multiple endpoints to list (GET), lookup (GET), edit (PUT and/or PATCH), and delete (DELETE)
- API key required as header in all API requests
- Access tokens required in the Authorization header of all API calls
- API Documentation
- Postman collections



Access Control API

Catalog Service API

Connectors Service API

Data Access Service API

Data Ingestion Service API

Dataset Service API

Flow Service API

Identity Service API

Mapping Service API

Observability Insights API

Offer Decision Service API

Access Control API^{1.0}

Access Control in Adobe Experience Platform allows IMS Organization administration to assign roles and permissions for various Platform capabilities. The Access Control API provides a public endpoint to retrieve effective policies for a user on given resources within a specified sandbox. All other access control capabilities are provided through the [Adobe Admin Console](#).

Related documentation:

- [Access Control documentation](#)

API paths:

- PLATFORM Gateway URL: `https://platform.adobe.io`
- Base path for this API: `/data/foundation/access-control`
- Example of a complete path for making a call to "acl/effective-policies":
`https://platform.adobe.io/data/foundation/access-control/acl/effective-policies`

Required headers:

- All service calls require the headers `Authorization`, `x-gw-ims-org-id`, and `x-api-key`. For more information on how to obtain these values, see the [authentication tutorial](#).
- All resources in Experience Platform are isolated to specific virtual sandboxes. All requests to Platform APIs require the header `x-sandbox-name` whose value is the all-lowercase name of the sandbox. The operation will take place in (for example, "prod"). See the [sandboxes overview](#) for more information.
- All requests with a payload in the request body (such as POST, PUT, and PATCH calls) must include the header `Content-Type` with a value of `application/json`.

ACCESS CONTROL POLICIES

Survey

<https://forms.office.com/r/rNLH6AddPN>





Data Management in Adobe Experience Platform

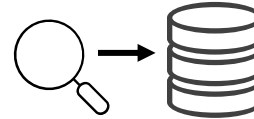


Data Standardization

Data Prep (Mapper)

Standard Data Mapping

- Map input data from batch and streaming source systems to XDM
- Intrinsic data validation during ingestion

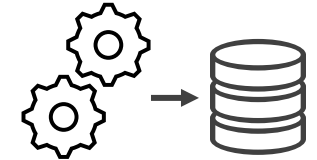


Data Exploration

Query Service (Ad hoc Query)

Ad-hoc Data Exploration

- Execute simple, short running (up to 10 minutes) queries on data ingested into Experience Platform data lake for exploration and validation (up to 10 minutes duration)



Data Transformation

Data Prep (Mapper)

Ingestion Data Preparation:

- Single dataset transformation operation (standard row level filtering, cleaning)
- Row based calculated fields

Query Service (Batch Query)

Custom Data Mapping

- Transform data from enterprise batch systems to ad hoc schemas into a derived dataset
- Post-ingestion data standardization

Advanced Data Exploration

- Execute long running (up to 24 hours) queries on data ingested into Platform data lake and store output of validation records for reporting and analysis

Post Ingestion Data preparation:

- Table level transformation
- Multi dataset transformation operations
- Multi-step operations
- Custom business logic

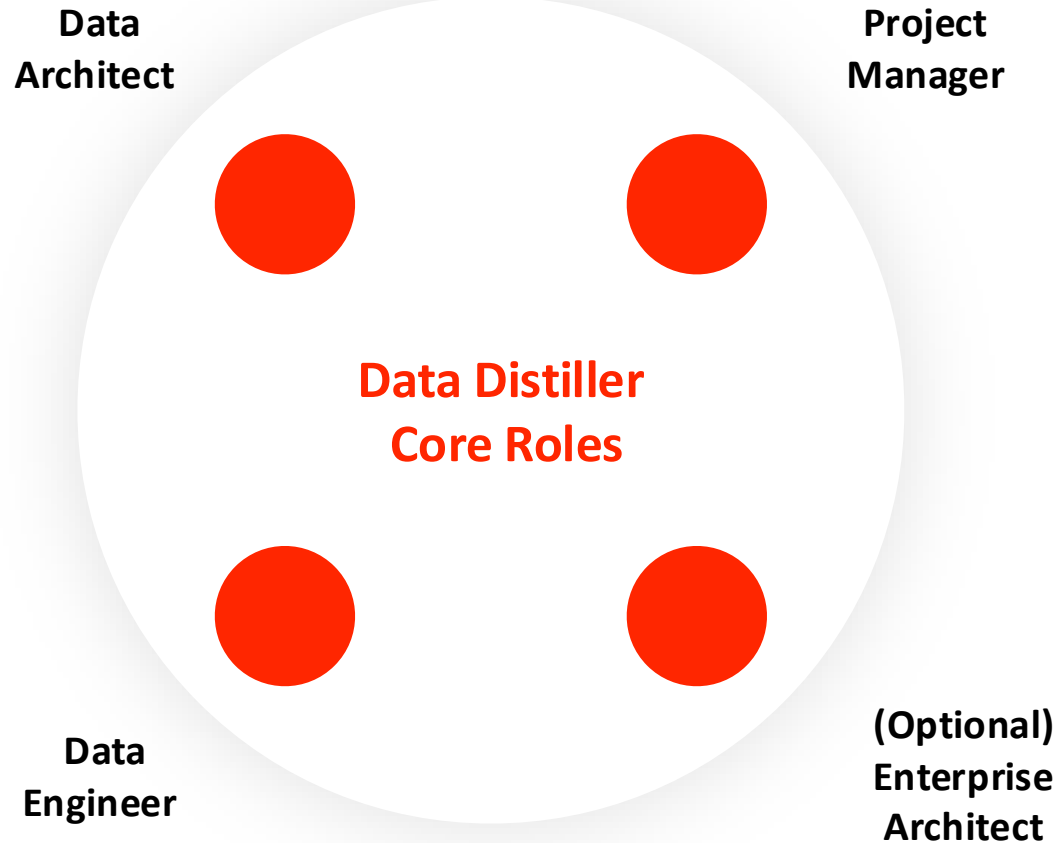
Foundational

(Native with Platform Based Apps)

Incremental

(Data Distiller Add-on)

Data Distiller Service Module: Core Team Roles



Our cross-functional team of Adobe data experts with experience across the Experience stack are tailored to **bring to life Query Service capabilities**

An example team design can be found here: [Full Service/Run and & Operate General Team](#)

3 Key Takeaways

1

Data Distiller adds value to platform-based applications

- Horizontal offering for cleaning, shaping, manipulating and enriching customer data for cleaner, richer, more actionable data with your platform-based applications
- Extremely vital to unlocking value for use cases where custom business logic is required

2

Data Distiller provides access to batch Query Service Capabilities

- Execution and Scheduling of SQL Query jobs to create derived data sets for a variety of use cases

3

Data Distiller Professional Services are available

- Our cross-functional team of Adobe data experts with experience across the Experience Platform and Platform-based application stack are tailored to bring to life Data Distiller capabilities