

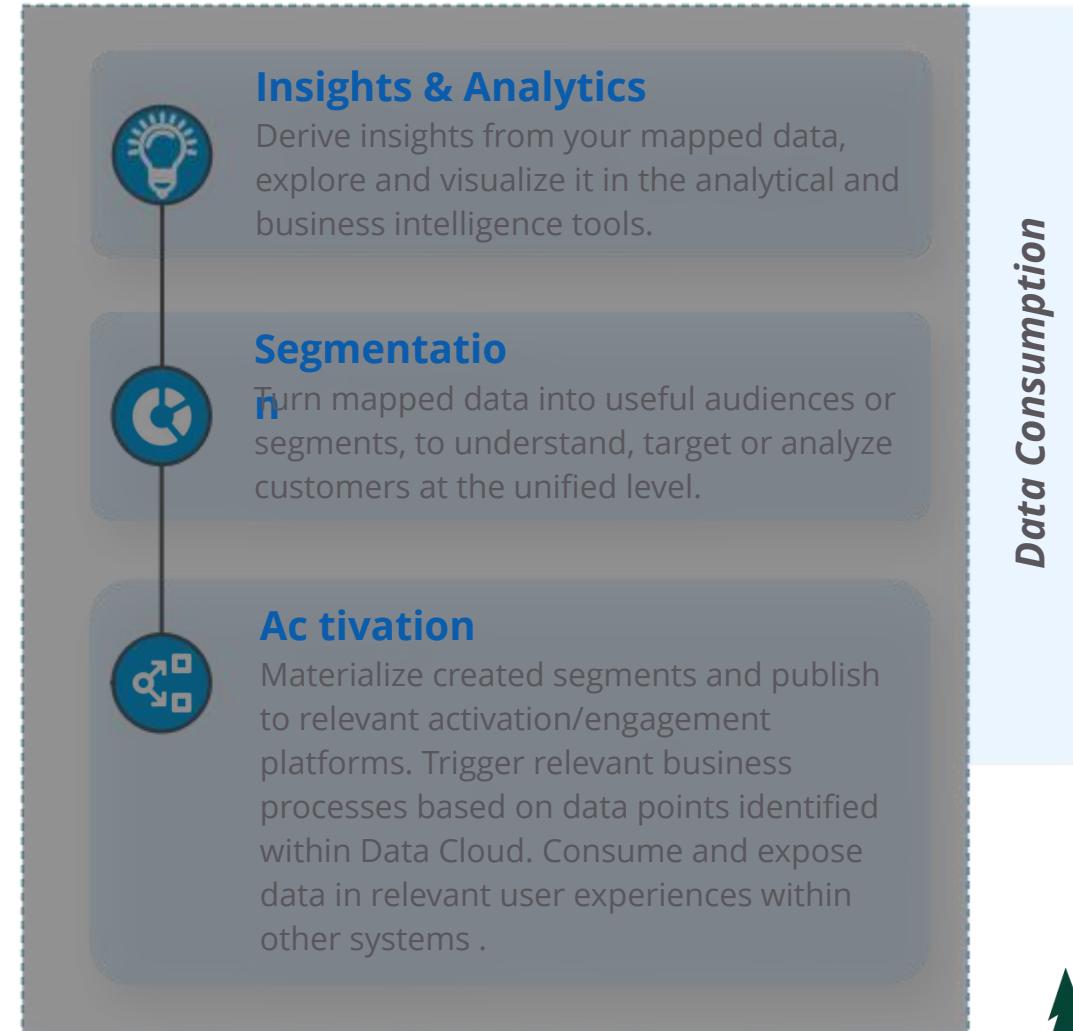
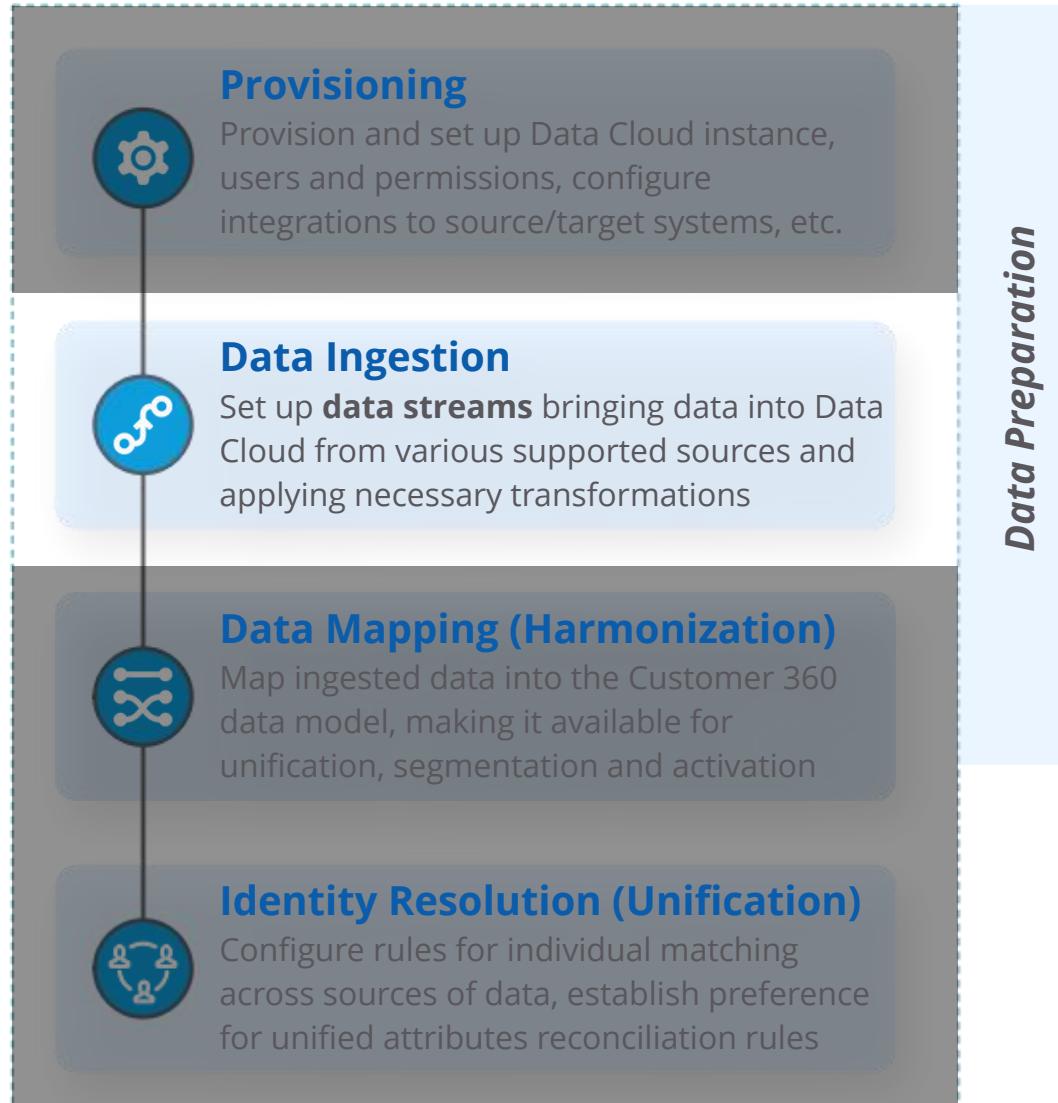


# Data Ingestion



# The Big Picture: Implementation Themes

Related to the components of Data Cloud

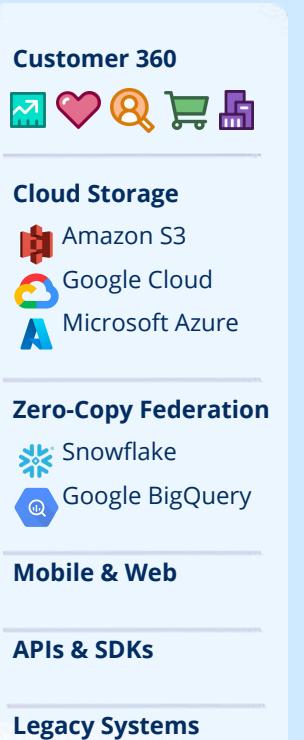


# Let's walk through how this works

A “day in the life” of customer data



## Data Sources



A vertical sidebar titled "Data Sources" containing a list of data source categories with corresponding icons:

- Customer 360**: Icons for Analytics, Loyalty, Search, Commerce, and Data Quality.
- Cloud Storage**: Icons for Amazon S3, Google Cloud, and Microsoft Azure.
- Zero-Copy Federation**: Icons for Snowflake and Google BigQuery.
- Mobile & Web**
- APIs & SDKs**
- Legacy Systems**

- What Data would you load?
- What connectors do you need?
- How much data would you load?

# Let's walk through how this works

A “day in the life” of customer data

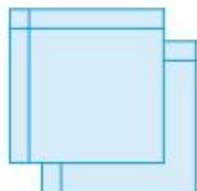
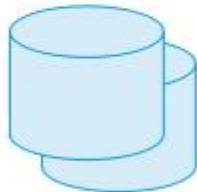


# Data Ingestion Flow Overview



## Data Streams:

An entity that can be extracted from a variety of Data Source Systems where data resides (e.g. CRM, SFMC, etc.)



## Data Source Objects (DSOs):

Original, ingested data



Partitions of your prepared data and its utilized components

## Data Spaces:

### Formulas

Ways to perform minor adjustments at the time of ingestion



### Bulk / Streaming Transformation

Ways to perform major joins / filters / transformations on DLOs

### Data Lake Objects (DLOs):

Source data hydrated with transformations

### Data Model Objects (DMOs):

Either materialized or views on top of the Data Lake Objects.

# Data Cloud Objects



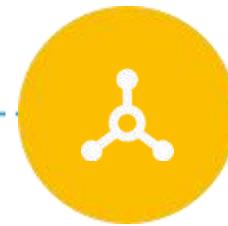
## Data Source Object

An object of data ingested as-is from the original data source. This is the original file format (e.g. CSV) or transient data storage in case of built-in connectors (e.g. Marketing Cloud).



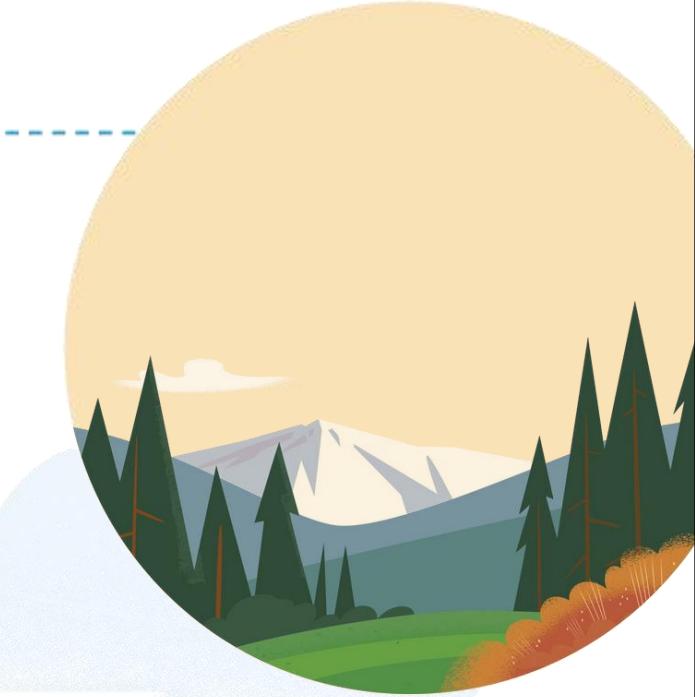
## Data Lake Object

When a data stream is created, Data Cloud automatically creates a locally stored Data Lake Object (DLO) with the same name. The DLO includes transforms and system source ids in addition to the raw source data.



## Data Model Object

A harmonized grouping of data created from data streams, insights, and other data sources.



Data must first be mapped from a DLO to a DMO for use in segmentation, activation, and analytics.

# Data Streams & Sources



New Data Stream

Select a source for the data stream to unify your data.

Connected Sources

- Marketing Cloud  
Starter data bundles and Data Extensions
- Salesforce CRM  
Import objects from Salesforce CRM
- Mobile App  
Import events from iOS and Android applications
- Website  
Import events from your connected websites.
- Ingestion API  
Stream and/or bulk upload data from external sources
- Interaction Studio  
Marketing Cloud Interaction Studio Bundle

Other Sources

Load a sample file in order to teach the system about your file's structure. At the end of this set up flow, you'll be able to specify where data should be retrieved from on an ongoing basis.

- Amazon S3  
Retrieve a file from Amazon Simple Storage Service
- Installed Data Kits & Pack...  
Import data streams from preconfigured data kits and packages.

Mulesoft Anypoint Exchange

Access these pre-built connectors and more from Mulesoft Anypoint Exchange. [Learn more](#)

- Adobe Marketo
- Amazon DynamoDB
- Amazon EC2
- Amazon Kinesis
- Amazon Lambda
- Amazon RDS
- Amazon Redshift
- Amazon S3

# Data Object Type Categories



## Profile

Segment oriented data set. A data set which contains any population you wish to segment by, or use as the starting population for a segment.



## Engagement

Time-series oriented data set. An Event Time field must be defined as part of set-up. The date field chosen for Event Time should be immutable otherwise records will be duplicated.



## Other

Data sets which are not related to either Profile or Engagement data.

Time-series data sets which do not have an immutable date field.



# Data Field Types



## Text

Stores any kind of text data. It can contain both single-byte and multibyte characters that the locale supports.

Zero length strings ("") and no value are treated as empty strings.

## Number

Stores numbers with a fixed scale of 18 and precision of 38.

Scale represents the number of fractional digits. Precision represents the count of digits, regardless of the location of the decimal point.

## Date

Holds the calendar date without a time part or time zone.

If the incoming data record includes a time part for a field configured as type date, the time part is ignored.

## DateTime

Stores an instant in time expressed as a calendar date and time of day.

A valid datetime must include the time part and time zone (following ISO-8601 standard). If time part and time zone are not included, it's inferred as 00:00:00 UTC.

See the [full list of expressions](#) by field type.

# Data Field Types

Added February 2024



## Email

Stores email addresses. The email data type is modeled on the text data type. You can use any valid text value for ingestion into an email data type field. Data Cloud doesn't validate the format.

## Percent

Holds percentage values. The percent data type is modeled on the number data type. Only valid numeric values are accepted for ingestion into a percent data type field.

## Phone

Stores phone numbers. Data Cloud doesn't validate the format of the phone number. The phone data type is modeled on the text data type. Any valid text value is accepted for ingestion into a phone data type field.

## URL

Stores URL values. Data Cloud doesn't parse or interpret the ingested URL value. It also doesn't validate the value for correctness. Data Cloud doesn't store any metadata related to the URL.

See the [full list of expressions](#) by field type.

# Data Field Types

Added February 2024



## Boolean

Possible values are  
'true', 'false', and blank.

Fields with boolean  
data type can't be used  
as primary keys, record  
modified fields, and  
internal organization  
fields.

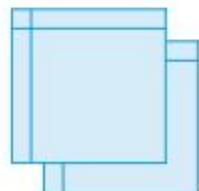
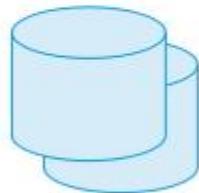
See the [full list of expressions](#) by field type.

# Data Ingestion Flow Overview



## Data Streams:

An entity that can be extracted from a variety of Data Source Systems where data resides (e.g. CRM, SFMC, etc.)



## Data Source Objects (DSOs):

Original, ingested data



## Data Spaces:

Partitions of your prepared data and its utilized components

### Formulas

Ways to perform minor adjustments at the time of ingestion



### Bulk / Streaming Transformation

Ways to perform major joins / filters / transformations on DLOs

### Data Lake Objects (DLOs):

Source data hydrated with transformations

### Data Model Objects (DMOs):

Either materialized or views on top of the Data Lake Objects.

# Formulas

## Leverage Formula Fields to Enhance or Enrich Source Data for Mapping

### Primary Keys

Create needed primary keys for mapping.

#### Transformation Formula

```
CONCAT('SFMC_523008784_', sourceField['Email'])
```

Consider functions like CONCAT()

### Set Picklist-type Values

Create fields that bucket values to simplify segmentation.

#### Transformation Formula

```
IF(sourceField['Spend']>=1000,'Platinum',
IF(AND(sourceField['Spend']<1000,
sourceField['Spend']>='500),'Gold',
IF(sourceField['Spend']<'500,'Silver','Unknown')))
```

Consider functions like IF(), AND(), or NOT()

### Standardization

Ensure consistent data formatting for activation.

#### Transformation Formula

```
PROPER(sourceField['First_Name'])
```

Consider functions like PROPER() or REPLACE()

## Transform Data Sources with Formula Fields

Supplemental fields can be hard-coded or derived from other fields in the data stream.

# Let's See Data Ingestion in Action



## DEMO TIME|



# Configure Data Ingestion



## 1. Service Cloud Data Bundle

- 1.1 Provide Access to fields – Case & Contact
- 1.2 Update Field label
- 1.3 Add Formula Fields
- 1.4 Remove Fields that are not required
- 1.5 Remove Account Data Stream

## 2. Provide access to Booking, Rental Preferences and Rental objects

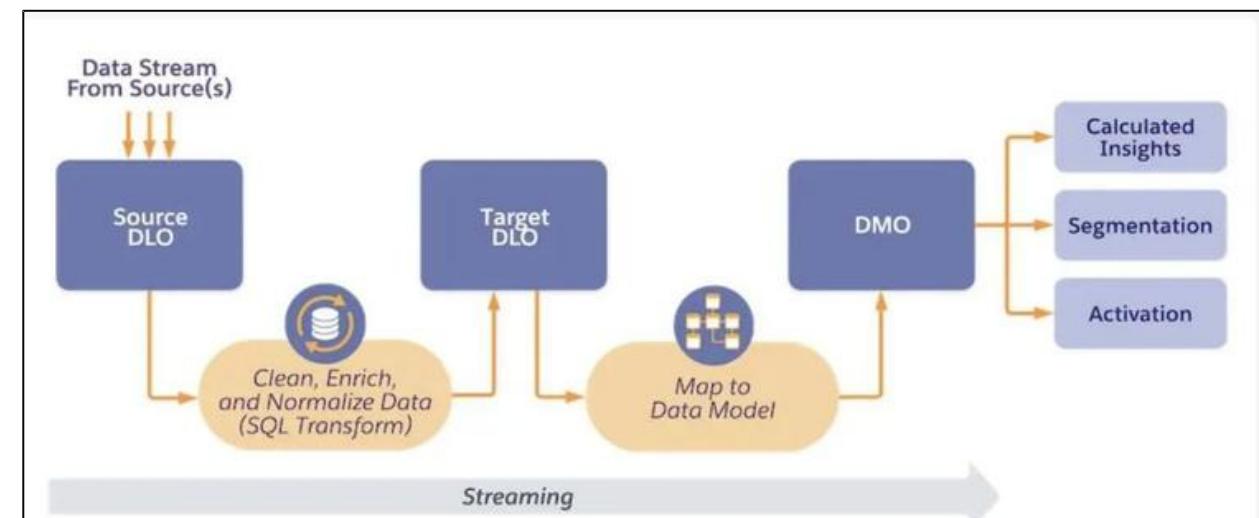
- 2.1 Add Formula fields in Booking Object
- 2.2 Add Formula fields in Rental Object
- 2.3 Add formula fields in Rental Preference Object
- 2.4 Create new DLO & add it to Data Space
- 2.5 Create new Streaming Data Transform

## 3. Configure S3 Connector & create Batch Data Transform

# Streaming Data Transforms

Transform data in near real-time

- A streaming data transform lets you clean and enrich your data in near real-time, as it enters the system.
- New records are ingested and are immediately transformed and appended to the output object.
- A streaming data transform reads records from a source data lake object (DLO) and runs a SQL query that modifies incoming data.
- It then maps the target data lake object to the Data Cloud data model in a Data Model Object (DMO).



# Streaming Data Transforms

Transform data in near real-time

A streaming data transform reads one record in a source data lake object, reshapes the record data, and writes one or more records to a target data lake object.

The source and target objects must be different objects.

A streaming data transform runs continuously as a streaming process, picking up new or changed data.



## Use Case:

Normalize Data with UNION Use Case

CUSTOMERID	NAME	MOBILEPHONE	HOMEPHONE	WORKPHONE
C1	John	408-512-2234		650-342-4612
C2	Peter		310-453-3421	915-870-9900

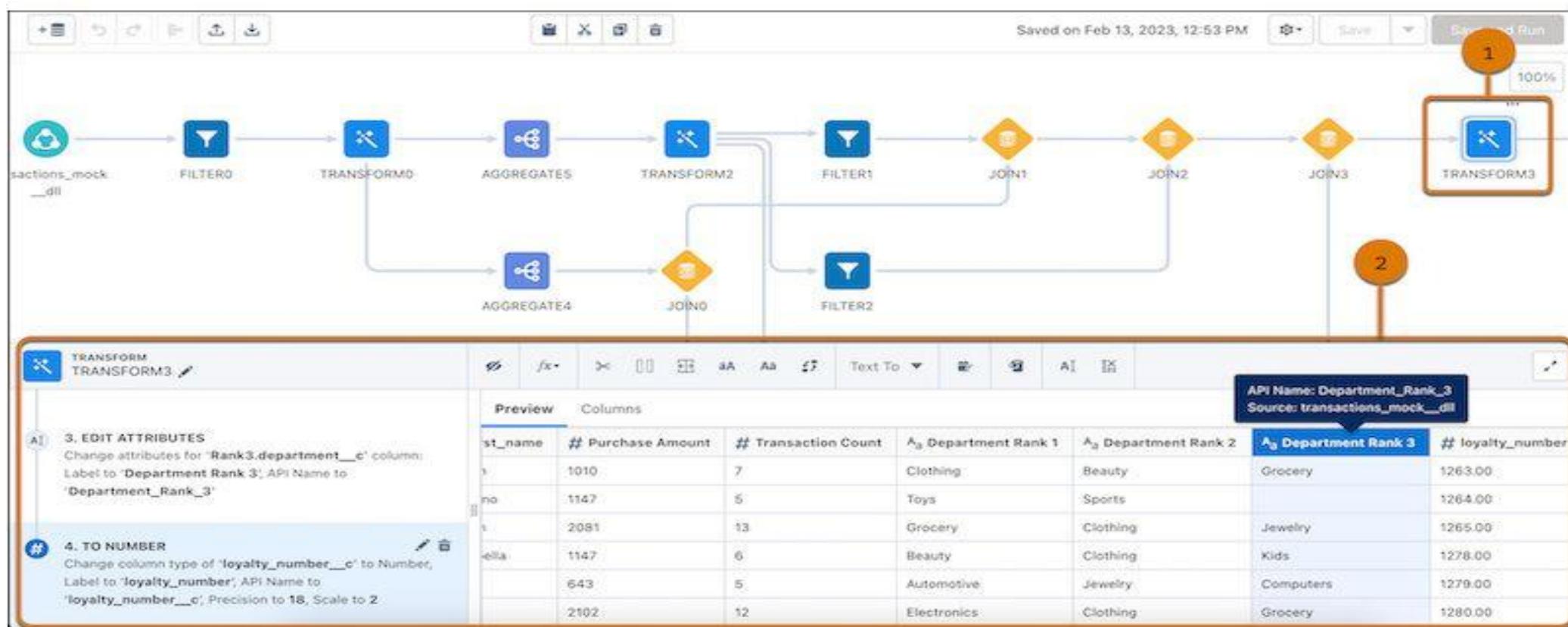


# Batch Data Transforms



Use a batch data transform to create a repeatable series of operations to transform your data and load it into a target for further usage. In contrast to a streaming data transform, which runs continually, a batch data transform runs on a scheduled basis.

- Does a full refresh of data
- Can use multiple source objects
- Can be used with DLOs or DMOs





## Batch Transforms

- Does a full refresh
- Repeatable process, can be scheduled or triggered manually
- Works with DLOs or DMOs as source objects
- Provides a visual editor tool and richer functionality
- Does not replace Calculated Insights

## Streaming Transforms

- Acts on one row of data at a time
- Transforms data as it's ingested
- Works only with DLOs
- Does not replace Calculated Insights

# Recap: Ingestion

## Steps for Configuring Data Streams

### 1. Select Data Source

Choose previously connected or authenticate new data source  
(cloud storage)

### 2. Select Data Source Object (Dataset)

Choose starting bundle or select object or specify filename

### 3. Define Data Stream Properties

Name source, set label, developer name and data category

### 4. Confirm Data Source Object Schema

Verify fields and data types, set primary key

### 5. Apply Transforms Data & Space

Optionally add formula fields to cleanse your source data or derive new fields and assign to a data space

### 6. Configure Updates to Data Source Object

Configure refresh mode and set the schedule

**Note:** Data spaces are not currently mentioned in PLC

# Hands-on Exercise Use Case



## Scenario

RAV Group is a company that has several lines of businesses operating under one main brand. They use the following products:

- Salesforce Marketing Cloud
- Salesforce CRM: contains data from vehicle rental business
- eCommerce platform: with transactions from their retail brand selling sports equipment

## Requirements

The RAV Group wants to combine data from two systems: Salesforce CRM that contains data from their vehicle rental business and eCommerce platform with transactions from their retail brand selling sports equipment. These brands operate as independent entities from the customer experience perspective. After the data is brought into the Data Cloud platform the RAV Group wants to identify and merge records of the people that exist in both systems.

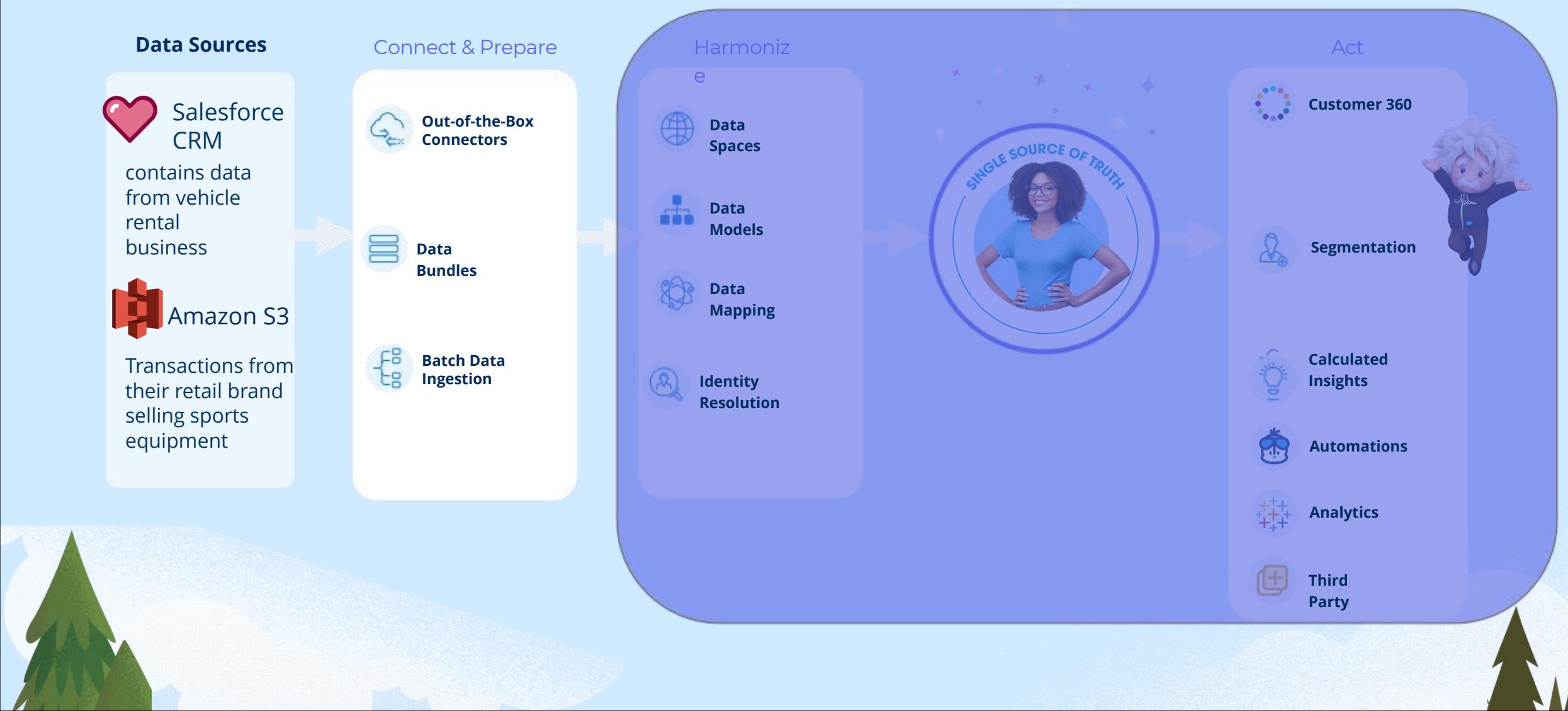
The combined audience should enable users from either of the brands to create segments of customers with certain characteristics that are sourced from either of the systems. These segments then can be consumed for further actions, such as analytics, retargeting on the eCommerce site , activating segments on Marketing cloud and creating a segment that can be used in a campaign in Marketing Cloud. Create Other segments and calculated insights for analysis.

## Solution

- Use an independent Data Cloud org that will ingest data from multiple data sources, and post segmentation activated audiences into relevant target systems. Ingest Data from all data sources Extend the standard Data Model. Perform harmonization unification of the data. Create required insights and segments from unified profiles. Activate segments to be consumed by a target system.

# Let's walk through how this works

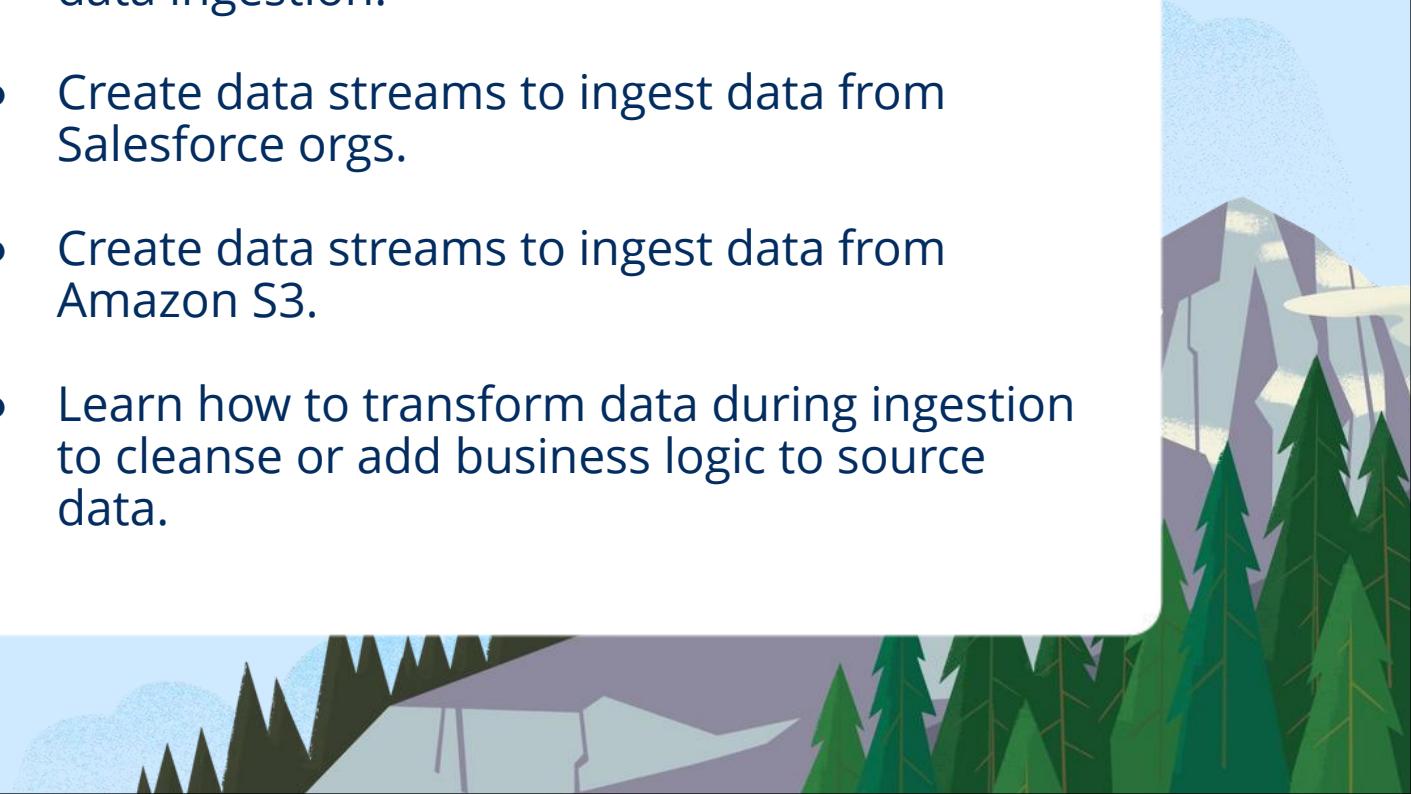
A “day in the life” of RAV Group





# Data Ingestion Exercise Goals

- Install a data bundle to assist with ingesting data from a Salesforce org.
- Connect Data Cloud to a Salesforce org for data ingestion.
- Create data streams to ingest data from Salesforce orgs.
- Create data streams to ingest data from Amazon S3.
- Learn how to transform data during ingestion to cleanse or add business logic to source data.

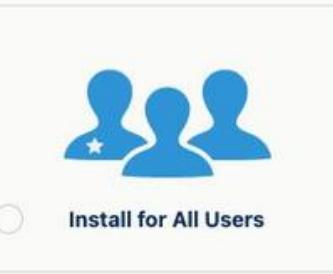


# Activity: Prepare Your Data

## Salesforce CRM Data

 **Install Data Cloud - learning data model**  
By Salesforce

 What if existing component names conflict with ones in this package?  
 Do not install.  
 Rename conflicting components in package.

  Install for Admins Only  
  Install for All Users  
  Install for Specific Profiles...  
  
**Install** **Cancel**

App Name	Publisher	Version Name	Version Number
Data Cloud - learning data model	Salesforce	Spring 2023 - Apr 19	1.4

**Description**  
Follow instructions in this guide to complete data initialization in your org.

**Additional Details** [View Components](#)



**Common Mistake** Many consultants skip the “Set Audit Fields” step in the Setup exercise, this will cause your installation to fail. Make sure to set this field!

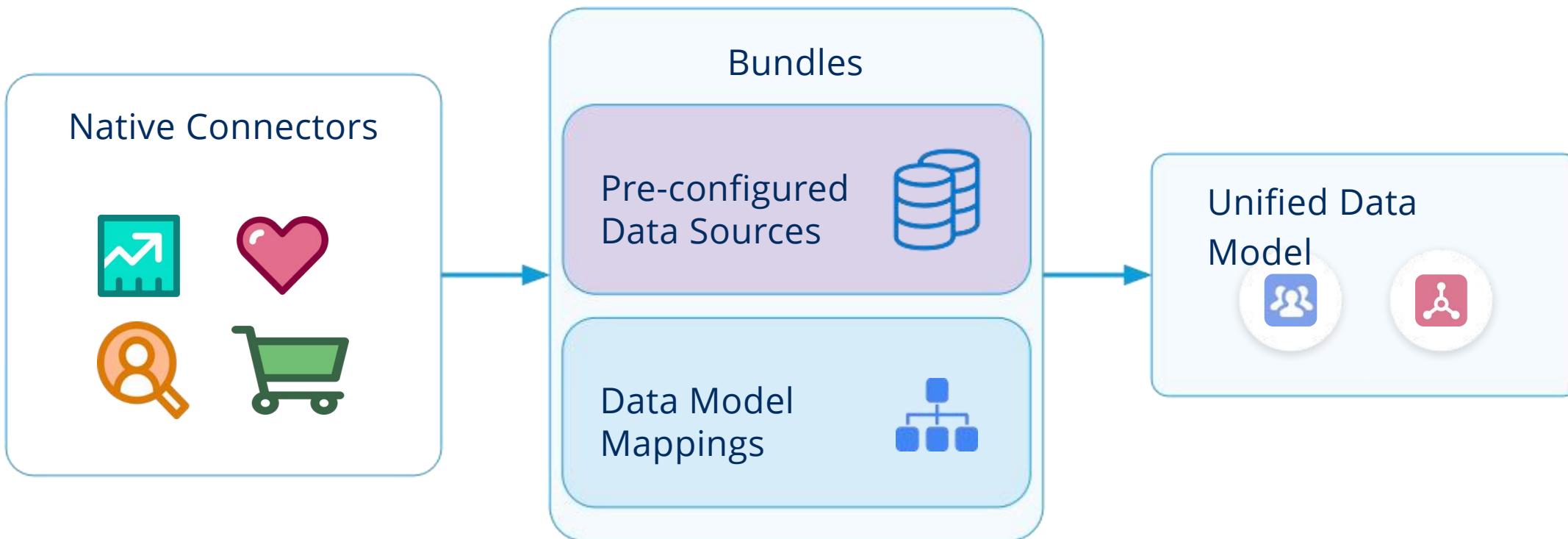
### Setup

Enable Enhanced Page Layout Editor  
 Enable Streaming API  
 Enable Dynamic Streaming Channel Creation  
 Enable “Set Audit Fields upon Record Creation”  
 Enable “Delete from Field History” and “Delete”  
 Enable Custom Object Truncate

Found in Setup  
Go to User Interface -> User Interface

# Explore Salesforce CRM Data Ingestion

Bundles: Quickly integrate common data sets from multiple sources



# Activity: Configure Data Ingestion

## CRM Starter Data Bundle



### Field Permissions

Field Name	Read Access	Edit Access
Account Name	<input type="checkbox"/>	<input type="checkbox"/>
Asset	<input type="checkbox"/>	<input type="checkbox"/>
Business Hours	<input type="checkbox"/>	<input type="checkbox"/>
Case Number	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Case Origin	<input type="checkbox"/>	<input type="checkbox"/>
Case Owner	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Case Reason	<input type="checkbox"/>	<input type="checkbox"/>
Case Source	<input type="checkbox"/>	<input type="checkbox"/>
Closed When Created	<input type="checkbox"/>	<input type="checkbox"/>
Contact Email	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Contact Fax	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Contact Mobile	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Contact Name	<input type="checkbox"/>	<input type="checkbox"/>
Contact Phone	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Created By	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Date/Time Closed	<input type="checkbox"/>	<input type="checkbox"/>
Date/Time Opened	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Description	<input type="checkbox"/>	<input type="checkbox"/>
Escalated	<input type="checkbox"/>	<input type="checkbox"/>
Internal Comments	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Last Modified By	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Parent Case	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Priority	<input type="checkbox"/>	<input type="checkbox"/>



**Expect a roadblock!** We have designed this exercise to show you what happens when the object you want to ingest has insufficient permissions with the integration user.

### Insufficient Permissions

You do not have permission to import records for 1 object. Please check with your Salesforce Admin to make the following object available for use. [Tell Me More](#)

**Case**  
1 Field unavailable: ParentId

**Ok**

# Activity: Prepare Your Data

## Amazon S3 Data Sources

Amazon S3 > Buckets > ravg-summer22 > ecommerce-data/ > Upload

### Upload Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files**, or **Add folders**.

**Files and folders (3 Total, 85.7 KB)**  
All files and folders in this table will be uploaded.

Find by name

<input type="checkbox"/>	Name	Folder	Type	Size
<input type="checkbox"/>	S3 Customer Profile.csv	-	text/csv	9.7 KB
<input type="checkbox"/>	S3 Order Headers.csv	-	text/csv	40.7 KB
<input type="checkbox"/>	S3 Order Line Items.csv	-	text/csv	35.3 KB



**Common Mistake** Many consultants skip the guide where we configure the Amazon S3 bucket, its Access Policy, and the User we will use to connect. Without this step complete, you won't be able to connect to your S3 Bucket! **[S3 Guide here:](#)** <https://salesforce.quip.com/Ge0zAXFPcYLE>

# Activity: Configure Data Ingestion

## Applying Data Transformations

```

SELECT Rental_c_<ORGID>__dll.Customer_c__c as Party__c
, Concat(Rental_c_<ORGID>__dll.Customer_c__c, "_", Rental_c_<ORGID>__dll.Airline_Mem
, "Loyalty Program" as Party_Identification_Type__c
, "Airline Membership" as Identification_Name__c
, Rental_c_<ORGID>__dll.Airline_Membership_Number_c__c as Identification_Number__c
FROM Rental_c_<ORGID>__dll
WHERE Rental_c_<ORGID>__dll.Airline_Membership_Number_c__c <> ''
AND Rental_c_<ORGID>__dll.Airline_Membership_Number_c__c <> ""

UNION

SELECT Rental_c_<ORGID>__dll.Customer_c__c as Party__c
, Concat(Rental_c_<ORGID>__dll.Customer_c__c, "_", Rental_c_<ORGID>__dll.Driver_Licen
, "Person Identification" as Party_Identification_Type__c
, "Driver License" as Identification_Name__c
, Rental_c_<ORGID>__dll.Driver_License_Number_c__c as Identification_Number__c
FROM Rental_c_<ORGID>__dll
WHERE Rental_c_<ORGID>__dll.Driver_License_Number_c__c <> ''
AND Rental_c_<ORGID>__dll.Driver_License_Number_c__c <> ""

```



**Common Mistakes** Assign a Data Space: with the addition of Data Spaces, make sure to assign this new DLO to your default data space! Reference your Org Id: Replace the placeholder ORGID fields with your actual Data Cloud Org Id.

Denormalized Record with 2 Data Points

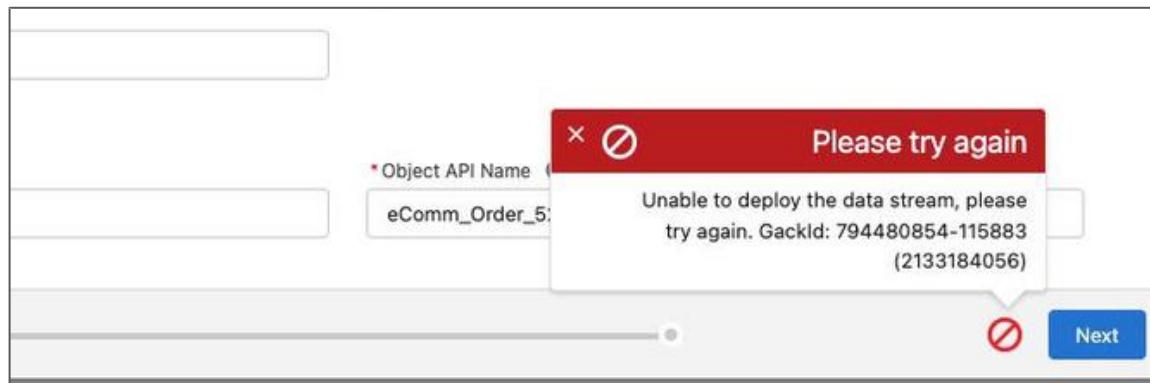
Data Transform

Data Point 1

Data Point 2

# Activity: Configure Data Ingestion

## Ingest Data from S3 Bucket



**Transformation Formula**

```
CONCAT(sourceField['Order Number']+'_'+sourceField['Order Line Number'])
```



### Common Mistakes

- Make sure your S3 User has the appropriate permissions to read the files in your bucket
- Set your refresh rate to none - this data will not change, and daily refreshes may impact your free S3 account
- Validate your data types! You cannot map two different types of data to the same field in a DMO
- Do not initially refresh the Product SKU Lookup and Related Activity Lookup data sources! We will make more configuration changes before this should be refreshed.



# Ingestion - Exam Tips



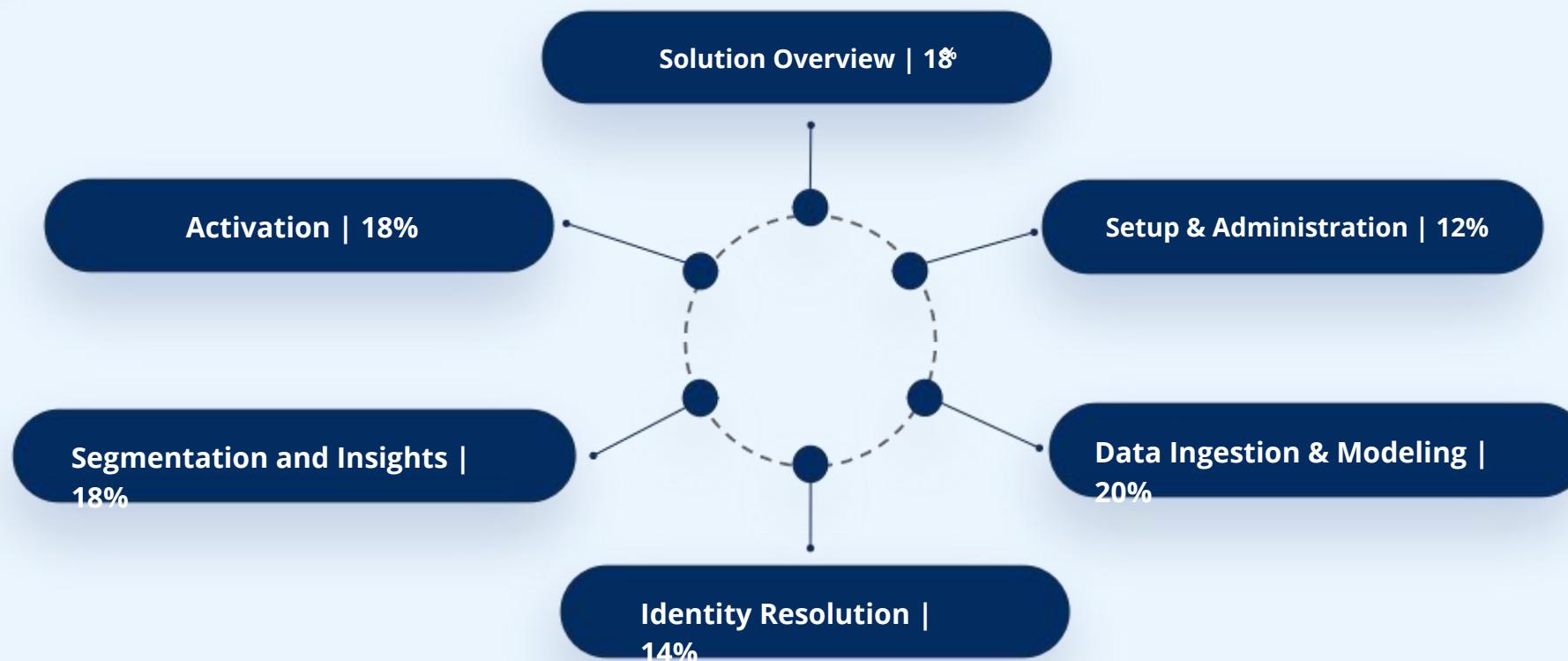


# Salesforce Certified Data Cloud Consultant

Total Question	60	Allotted Time	105 min	Passing Score	62 %
----------------	----	---------------	---------	---------------	------



## Exam Outline



*Test takers are strongly advised to complete the Data Cloud Partner Learning Camp Curriculum before attempting the exam*





## Salesforce Certified Data Cloud Consultant

Total Question	Allotted Time	Passing Score
60	105 min	62 %



### Study Tips for Ingestion

Things that require special attention - **study these concepts!**

- **Key Terminology:** Especially DSO, DLO, DMO
- **Data Spaces:** This is a new feature launched since the PLC was last revised
- **Native Connectors:** Understand what's available & review hard limits
- **Cloud Storage:** Understand how to ingest data from Cloud Storage & how S3 differs from Azure, GCS
- **Bundles:** Understand bundle offerings & objects ingested - [watch this video on SFMC bundles](#)
- **Schedules & Modes:** Understand ingestion timings & [differences between a full refresh and a upsert](#)
- **Data Categories:** Understand the three types & benefits
- **Formulas & Transformations:** Understand when & how they're applied & limitations
- **Field Types:** Understand what's available (esp. difference b/t Date & DateTime)

*Salesforce recommends a combination of hands-on experience, training courses in Partner Learning Camp, and self-study in the areas listed in the Exam Outline section of this exam guide*





## Salesforce Certified Data Cloud Consultant

### Study Tips for Ingestion

## Ingestion Consideration - remember these concepts!

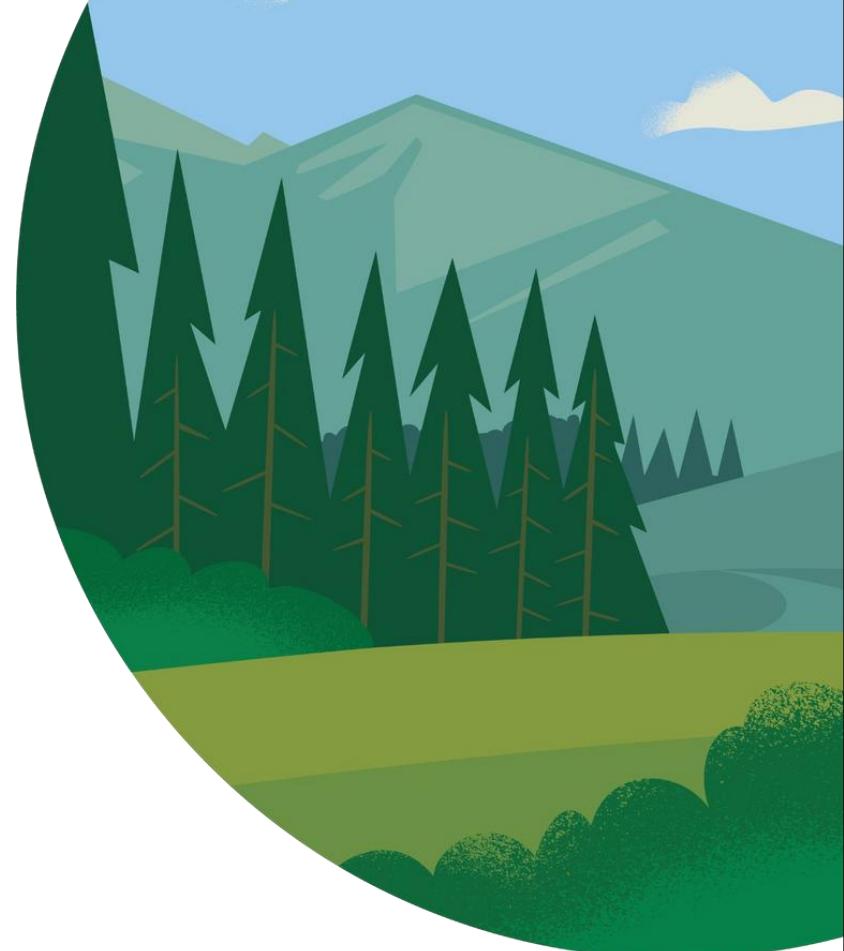
- Incorrect policy settings are the most common issue when facing errors with **S3 ingestion**.
- Understand **bundle** offerings & objects ingested Carefully select the data **category** (profile, engagement, or other).
- You cannot change after saving the data stream.
- Make sure to assign all new DLOs to the correct **data space** (default must be assigned if only one).
- Understand what **field types** are available and particularly the difference between Date & DateTime.
- Understand ingestion **timings** & differences between a full refresh and a upsert.
- Understand when & how **formulas and transformations** are applied and their limitations.

*Salesforce recommends a combination of hands-on experience, training courses in Partner Learning Camp, and self-study in the areas listed in the Exam Outline section of this exam guide*

Total Question	Allotted Time	Passing Score
60	105 min	62 %



# Q&A



salesforce

# Thank you

