

# DQ Foundation

v15.4.x



ataccama

# Agenda

## DQ Foundation

- Rules
- Lookups
- DQ Monitoring Projects
- Data Reconciliation
- ONE Data
- Data Transformation Plans



**There are 4 practical lab exercises on this course**

# Material GOALS

## Who is this for:

- This session is meant for power users that will be configuring Rules and Monitoring Projects in Gen 2

## Prerequisites/what should be known:

- Catalog & Glossary for Power Users

## Target group:

- Power Users or Developers

## Training Duration:

- This session is estimated for 1 day duration.

## Gen 2 version:

- 15.2.x



# Rules



# Rules – Overview

Rules are used to **detect terms** or **evaluate DQ results in ONE**.

**Rules in ONE are split into two categories:**

1. **Detection rules**
2. **Data quality evaluation rules (DQ rules)**
  - DQ rules are further split into **dimensions**.
  - You can define which dimensions will contribute to the **overall quality** calculation.

Both rules can be accessed and created from **Data Quality > Rules** but serve very different purposes in the platform.

The screenshot shows the Ataccama Data Quality interface. On the left, there is a sidebar with various icons and a main menu. The main area is titled "Data Quality" and "Rules". A purple box highlights the "Rules" section under "Data Quality". The "Rules" section is divided into "Published", "Unpublished", and "All" tabs. There is a search bar and filters for "Terms", "Input Attributes Terms", and "Stewardship". Below these are two views: "Standard view" and "Hidden columns". The "Standard view" table lists various rules with columns for Name, Type, Dimension, Rule definition source, Stewardship, Terms, and Description. Some rows have small icons next to them. The "Hidden columns" view is partially visible below it.

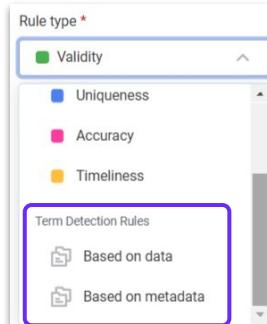
Name	Type	Dimension	Rule definition source	Stewardship	Terms	Description
Academic title	Detection		Ataccama Default Rules	Data Office	Academic title	Academic title
Airport code	Detection			Data Office		Detect a three-letter g...
Airport code	DQ Evaluation	Validity		Data Office		A three-letter geocode...
Boolean Completeness	DQ Evaluation	Completeness	Ataccama Default Rules	Administrators		This rule checks Boole...
CINS Detection	Detection		Ataccama Default Rules	Finance	CUSIP Number	CUSIP is a nine-chara...
CUSIP Detection	Detection		Ataccama Default Rules	Finance		Looks for CINS pattern.
Continent	Detection		Ataccama Default Rules	Data Office	Continent	Detection rule for cont...
Country	Detection		Ataccama Default Rules	Data Office	Country	Country

# Rule Logic Types

## Detection Rules

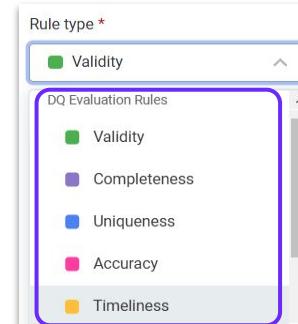
**There are two types of detection rules, which can be used together or separately on a term:**

- Data-based rules
  - Metadata-based rules
- 
- Detection rules are used for rule-based term detection during profiling and data discovery.
  - Detection rule creation and configuration are in **Data Quality**, while most interactions are done in **Business Glossary** and **Data Catalog**.



## DQ Evaluation Rules

- DQ evaluation rules run during DQ evaluation in the **data catalog**, in **monitoring projects**, or as **part of data observability**.
- DQ evaluation rules are used to evaluate data based on a specific **dimension**.
- **Overall Quality metric** includes results from contributing dimensions and can be found on the Data Quality tab for a catalog item or for the term itself, or in monitoring project results and reports.



# Rule Logic Types

## Detection Rules

- Are used to **detect glossary terms** of Catalog Item attributes.
- Can be combined **with AI to detect terms** in Catalog Items.
- **Cannot be used in monitoring projects.**
- **Cannot be used in Catalog Items** for evaluating the quality of the data.
- **Cannot be downloaded** and used as components in ONE Desktops plans.
- **Can be created in the web application only** - component detection rules are not possible.

## DQ Evaluation Rules

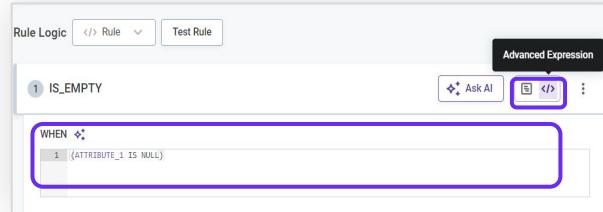
- Are used to **evaluate data quality** of Catalog Items.
- Can be used in **projects to monitor the quality of data.**
- **Can be used in Catalog Items** by attaching to relevant glossary terms.
- **Can be downloaded** and used as components in ONE Desktops plans.
- **Can be created in the web application** or through Validation Components in ONE Desktop.

# How to implement Rule Logic?

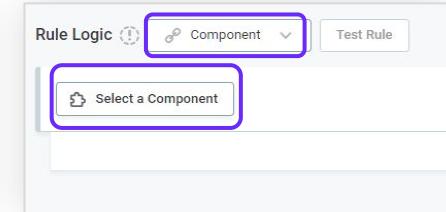
Condition Builder



Advanced Expression



Component



- Uses simple logical operators to implement the rule logic.
- Multiple conditions can be applied in a rule.
- Applicable to both Detection and DQ rules.

- Uses expressions and functions to develop more complex rule logs.
- you can use AI to generate the expression from a text prompt or using Python syntax or SQL.
- Applicable to both Detection and DQ rules.

- Creates components in the ONE Desktop.
- Allows combining data from different types of sources.
- Only applicable to **DQ Evaluation Rules** of Validity dimension.

# Creating DQ Evaluation rule

**There are multiple ways to create a DQ evaluation rule in ONE:**

## A. Rule creation in the **rule library**

- This method gives you access to all the possible rule-creation options, including variables, parameters, and scoring.

## B. Quick rule creation from the **attribute**

- The workflow is optimized for fast and simple rule creation from your data.
- The rule template comes prefilled with input and test data, featuring only the most commonly used options.
- For advanced options like variables, you can edit the rule in the rule library after creation.

## C. Quick rule creation from the **profiling results**.

- Create rules quickly using the profiling results.
- You can incorporate mask, pattern, and frequency results directly into rule conditions for faster and easier implementation.

The screenshot displays the Ataccama Data Quality interface with three main panels:

- Rules Library Panel:** Shows a list of rules including Detection Rules and DQ Evaluation Rules. A red box highlights the "Create" button in the top right corner of the header.
- Attribute-based Rule Creation Panel:** Shows attributes for a dataset (e.g., src\_name, src\_gender, src\_birth\_date) with a red box highlighting the "Create DQ Rule" button.
- Profiling Results-based Rule Creation Panel:** Shows a "Create a new rule" dialog with sections for "Mask Analysis" and "Pattern Analysis". A red box highlights the "Use results in rule" button in the Pattern Analysis section.

# Create DQ Evaluation rule in the rule library (1/2)

1. Go to **Data Quality >Rules** section.
2. Click the **Create** button.
3. Provide a name.
4. Specify the **Rule type**.
5. Click the **Save** button.

- To implement its logic, the user should go to the **Implementation tab**.
- Glossary terms or Relations with other rules can be added in the Overview tab.
- After a rule is created and applied, additional metadata like Occurrence Statistics and History are available in the Rule section.

The screenshot shows two views of the Ataccama Data Quality interface. The top view is the 'Rules' library, where a 'Create' button is highlighted with a purple box. A pink arrow points from this button down to the 'Create Rule' dialog in the bottom view. The 'Create Rule' dialog has a purple box around the 'Rule type' dropdown, which is set to 'Validity'. A callout bubble with the text 'You should specify the Rule type' also points to this dropdown. The 'Name' field contains 'xy\_Product Line'. The bottom view also shows sections for 'General information', 'Description', and 'Term Detection Rules' (with options for 'Based on data' and 'Based on metadata').

# Create DQ Evaluation rule in the rule library (1/2)

5. Define **Input Attribute**(s) and their types.
6. Choose a **Rule Logic**: Rule, Component (for Validity type), Aggregation rule (for Validity type).
7. Select the rule condition, for both rule types (except for DQ Component rules), **Condition Builder** or **Advanced Expression** can be selected to implement the rule's logic.
  - When creating rules, instead of having to write ONE Expressions from scratch, you can use **AI to generate the expression** from a text prompt.
  - You can also generate the expression based on Python syntax or validate, build on or edit an existing expression.
  - Rule can be **tested** before it is Published.

The screenshot shows the 'DQ Evaluation Rule' creation interface. At the top, there is a toggle switch labeled 'Use AI to generate Rule logic and inputs' and a 'Beta' button. Below this, the 'DQ Evaluation Rule' section has a dropdown menu set to 'Validity'. The 'Inputs' section shows an attribute 'Abc String' with a value 'VALUE'. The 'PARAMETERS' section has a 'Rule' parameter added. The 'VARIABLES' section has a 'Component' variable added. The 'Rule Logic' section shows a condition 'IS\_EMPTY' with a 'value' input. A 'Test Rule' button is also present. Callout boxes provide the following information:

- 'You can generate full DQ evaluation rule implementations with Gen AI using natural language'
- 'You should define at least one input for your rule.'
- 'You can use AI to generate the expression from a text prompt.'
- 'Rules can have multiple conditions.'
- 'Switch between condition builder or Advanced Expression'

# Aggregation Rules

- **Aggregation rules** in ONE web application can be utilized in **all DQ Rule types** (Validity, Uniqueness, Completeness, Timeliness and Accuracy).
- Aggregation rules are used when there is a need to perform **Group by, Aggregations, or Group by + Aggregation logic.**
- Group by allows **splitting data into multiple groups of rows**, based on the **inputs** provided.
- Aggregation rules can have one or more Input Attributes and their condition can be set either using **Condition Builder** or **Advanced Expression.**

The screenshot shows a user interface for configuring a DQ Evaluation Rule. At the top, there is a toggle switch for "Use AI to generate Rule logic and inputs" and a "Beta" indicator. The "DQ Evaluation Rule" dropdown is set to "Validity".

**Inputs:** This section is highlighted with a purple border. It contains two attribute entries: "123 Long" mapped to "CUSTOMERNUMBER" and "1.2 Float" mapped to "CREDITLIMIT". Each entry has a "Add Term" button. Below these are "Add attribute", "PARAMETERS", and "VARIABLES" sections, each with an "Add [type]" button.

**Rule Logic:** This section is also highlighted with a purple border. It includes tabs for "Rule Logic" (selected), "Aggregation rule", and "Test Rule".

**Group By:** This section is highlighted with a purple border. It shows a header with "NAME" and "PHONE" fields. Below is a condition builder for the "IS\_EMPTY" operator. The "WHEN" section contains two AND conditions: "is empty" for "NAME" and "is empty" for "PHONE". The "THEN" section defines the result as "Valid" with a score of "1000000" and the condition "IS\_EMPTY".

# Scoring records

- **Scores** are used as a **numerical expression** of each record's **invalidity severity level** according to the rule configuration. Score is not self-explanatory by default and should be explained via Explanation Codes.
- **The higher the score** is – **the more severe** the problem is.
- The Score can be **assigned** through the **web application** in the rule's logic, or in the **ONE Desktop** via **Validation Components**. To be able to see the total score of each record, it is necessary to configure **export** project results.

The diagram illustrates the process of generating a scoring table from a rule configuration and a monitoring project.

**Left Panel (Scoring Rule Configuration):**

- A screenshot of a web-based configuration interface for a rule named "IS\_INVALID".
- The "WHEN" condition is set to "not length(VALUE) = 3".
- The "THEN" section shows a table with two rows:
  - Result:** Valid (radio button)
  - Score:** 1000 (text input field)
  - Explanation:** IS\_INVALID (text input field)

**Middle Panel (Monitoring Project):**

- A screenshot of the "Training project" tab in the Ataccama interface.
- The "orders\_transformed" dataset is selected.
- A component named "export\_dq\_orders\_transformed" is configured.
- "Post-processing results" show two shared files:
  - customers\_out.csv (114.81 KB)
  - customers\_out.csv.metaData (3.8 KB)

**Right Panel (Export Table):**

- A screenshot of a table showing the generated data. The columns are labeled O, P, Q, R, S, T, and score.
- The data rows include:

O	P	Q	R	S	T	score
email	valid_rule	invalid_ru	invalid_ru	0	0	0
atelierra	validation	String	Con	accuracy	C	0
signalifts	validation	String	Con	accuracy	C	0
australian	accuracy	FString	Con	accuracy	C	1000
larochelle	validation	String	Con	accuracy	C	0
NULL	String	Con	validation	accuracy	C	1000
minigiftsc	accuracy	FString	Con	accuracy	C	1000
N/A	String	Con	validation	accuracy	C	1000
blauersee	accuracy	FString	Con	accuracy	C	1000
minihwee	accuracy	FString	Con	accuracy	C	1000
landoftoy	accuracy	FString	Con	accuracy	C	1000
euroshopj	validation	String	Con	accuracy	C	0
volvomod	accuracy	FString	Con	accuracy	C	1000
danishwh	validation	String	Con	accuracy	C	0
savelyhe	accuracy	FString	Con	accuracy	C	1000
dragonsol	accuracy	FString	Con	accuracy	C	1000

**Ataccama Logo:**

ataccama

# Scores – Best Practices

## Four scoring (quantification) levels:

Scoring Level	Scoring Description	Scoring Result
0	No modification was done to the input value.	VALID
<10,000	Small modification was done to the input value (trim, squeeze spaces, uppercase, remove unsupported characters, safe replacements, etc.).	VALID/CORRECTED
<10,000,000	Huge modification was done to the input value (not found in the dictionary, does not meet official rules, major error, unsafe replacements, etc.).	UNSAFE/UNKNOWN
>10,000,000	Input value or pre-cleansed value (after a small modification) is null.	NOT VALID

It is recommended to use multiplications of 10 if you want to emphasize/differentiate the importance within one scoring level.

# Topic Highlights

- **Rules in ONE** are used for detecting terms or evaluating **Data Quality** (DQ) results.
- **Data Quality Evaluation rules** assess attribute and catalog item quality, and can be used in monitoring projects.
- DQ rules are divided into **Dimensions**, such as Validity, Completeness, Uniqueness, Accuracy, and Timeliness.
- Implementing rule logic can be done using the **Condition Builder** and **Advanced Expression**, allowing for flexible logic design.
- **AI** can also be used to generate expressions from text prompts, Python syntax, or SQL.
- **Aggregation Rules** are applicable to all DQ rule types, and used for Group by, Aggregations, or Group by + Aggregation logic.

# Memory Refresher #1 Rules



# Lookups



ataccama

# Lookups

**Dictionary files (lists or reference data) called Lookups** are placed in the Data Quality section - Lookup Items.

- Lookup items allow you to use unmanaged **reference data** in ONE.

**Lookup items can be compared to raw data in order to help with:**

- Data cleansing
- Verification
- Matching
- Enrichment.

The screenshot shows the Ataccama Data Quality interface. On the left, there is a vertical navigation bar with various icons and sections: Home, Search, Components, Transformation Plans, DQ Firewalls, Monitoring Projects, Reconciliation Projects, and a highlighted 'Lookup Items' section. The main area is titled 'Lookup Items' and includes filters for 'Published', 'Unpublished', and 'All'. A search bar allows searching for full-text. Below the filters, there are dropdown menus for 'Source Catalog Item', 'Key Attribute', 'Stewardship', and 'Name'. The main table displays 'Lookup Items' in 'Standard view', showing columns for Name, Source item, Key, Rebuild strategy, Duplicities, and options to remove accents and repeats. The table lists several entries, such as 'Academic Titles\_Key\_45f989b' and 'academic\_titles'.

Name	Source item	Key	Rebuild strategy	Duplicities	Do remove accents	Do remove repeats
Academic Titles_Key_45f989b	Academic Titles	Key	ON DATA CHANGE	FIRST	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Academic Titles_Key_51871fb	Academic Titles	Key	ON DATA CHANGE	FIRST	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Academic Titles_Key_8d30a84	Academic Titles	Key		FIRST	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Academic Titles_Key_c6c3c00	Academic Titles	Key	ON DATA CHANGE	FIRST	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
academic_titles					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
acc_academic_titles					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
acc_airports					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
acc_alpha_2_code					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
acc_alpha_3_code					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

# Create lookup item ( 1 / 2 )

- Currently, only **one column** Lookup Items are supported in the ONE Web application.
- Lookups can be used for both DQ or Detection rules through Advanced Expressions or Condition Builder.
- To employ Lookup Items in Rule Logic conditions, user can choose the **"is (not) from lookup"** option when using the Condition Builder.

To create a new lookup, user can:

- Create a new lookup from an existing Catalog item using the **Create** button.
  - Select an existing catalog item's attribute and transform it into a lookup.
- **Upload a lookup file (.lkp)** using the Upload button.

The screenshot displays the Ataccama ONE web application's Data Quality module. On the left, a sidebar lists various project components: Rules, Detection Rules, DQ Evaluation Rules, DQ Dimensions, Components, Transformation Plans, DQ Firewalls, Monitoring Projects, Reconciliation Projects, and Lookup Items. The main workspace is titled 'Rules' and contains a single entry for 'Country'. The 'Country' entry has a dropdown menu with options like 'ABC String value' and '+ Add Term'. Below this is a 'VARIABLES' section with a 'Add variable' button. The central area is the 'Rule Logic' editor, which is currently active. It features a 'Condition' section with two expressions. The first expression is 'ABC value value is from lookup countries\_eng'. The second expression is 'ABC value value is not from lookup blacklist'. A tooltip 'All expressions must be true' is displayed above the second expression. Below the expressions is a '+ Add expression' button. The 'THEN' section contains a 'Detect Term' field with the value 'Country'. The top right corner of the interface includes standard editing tools like 'Edit', 'Save', and 'Cancel'.

# Create lookup item ( 2 / 2 )

1. In **Data Quality > Lookup Items** select **Create**.
2. Provide general information about the lookup item, such as **name** and **description** (optional).
3. Select whether you want to create the lookup from a **catalog item attribute**, or **upload a lookup file**.

- A. If you select **Choose from Catalog Items**, select the required **catalog item** and **attribute**.
- B. If you select **Upload from computer**, **drag and drop the lookup file** to the space provided, or **browse files**.

4. In Data Configuration, define the following:
  - A. **Rebuild strategy**: Choose when the lookup data is updated.
  - B. **Duplicates handling**
5. Select **Save**.
6. Publish the **draft**.
7. Select **Build Lookup**.

The screenshot shows the Ataccama Data Quality interface. On the left, there's a sidebar with various icons and a navigation menu. The main area is titled 'Knowledge Catalog' and shows a 'Sources > xy\_training > postgres > public > customers city' path. The 'Overview' tab is selected. The 'Description' section contains fields for 'Add' and 'Generate'. Below it is a 'Glossary terms' section with two entries: '100% [USA] City' and '51% Surname'. The 'Data Quality' section is currently empty with a note: 'Data Quality has not been evaluated yet. To evaluate Data Quality you need to either add Rules or add Terms that are connected with Rules to attributes of your interest.' On the right, there's a detailed view of the lookup item with tabs for 'General information', 'Column type: varchar', 'Data type: STRING', 'Comment: Customer registration city', 'Supported: Yes', 'Order: 8', and 'Column size: 200'. A context menu is open at the top right, with the 'Create new lookup' option highlighted.

The lookup is built. You can now edit lookup metadata, manage access, view and restore historical versions, and edit, delete, or update the lookup.

# Topic Highlights

- Lookup Items are **reference data**.
- Lookup files extension – **.lkp**.
- Lookups used in **Detection** or **DQ rules** to validate attribute values.
- Upload a lookup file (.lkp) using the Upload button.
- Currently supports only **one column Lookup Items** in the ONE Web application.

The screenshot shows the Ataccama ONE Web application interface. On the left, a sidebar titled "Data Quality" contains navigation links for Rules, DQ Dimensions, Components, Transformation Plans, DQ Firewalls, Monitoring Projects, Reconciliation Projects, and Lookup Items. The "Lookup Items" link is highlighted. The main content area is titled "Lookup Items" and shows a list of items. The header of the list table includes columns: Name, Source item, Key, Rebuild strategy, Duplicities, Do remove accents, and Do remove repeat. Below the header, there are 12 rows of data, each representing a lookup item. The first row is for "Academic Titles\_Key\_45f989b" with "Academic Titles" as the source item and "Key" as the key type. Subsequent rows show other lookup items like "academic\_titles", "acc\_academic\_titles", etc., with various settings for rebuild strategy, duplicates, and accent removal.

Name	Source item	Key	Rebuild strategy	Duplicities	Do remove accents	Do remove repeat
Academic Titles_Key_45f989b	Academic Titles	Key	ON DATA CHANGE	FIRST	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Academic Titles_Key_51871fb	Academic Titles	Key	ON DATA CHANGE	FIRST	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Academic Titles_Key_8d30a84	Academic Titles	Key		FIRST	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Academic Titles_Key_c6c3c00	Academic Titles	Key	ON DATA CHANGE	FIRST	<input checked="" type="checkbox"/>	<input type="checkbox"/>
academic_titles					<input checked="" type="checkbox"/>	<input type="checkbox"/>
acc_academic_titles					<input type="checkbox"/>	<input type="checkbox"/>
acc_airports					<input type="checkbox"/>	<input type="checkbox"/>
acc_alpha_2_code					<input type="checkbox"/>	<input type="checkbox"/>
acc_alpha_3_code					<input type="checkbox"/>	<input type="checkbox"/>

# Memory Refresher #2 Lookups



# Lab Exercise #1

## Basic DQ Rules



# DQ Monitoring Projects



# Monitoring Projects

Data Quality Monitoring projects are used to **evaluate the data quality** of Catalog Items and monitor the changes **over time**.

## Within a project, the user can:

- Select **Catalog Items** to be evaluated.
- Apply **data quality rules**.
- Display the results in multiple formats.

Use monitoring projects for manual monitoring of your most critical data assets.

Monitoring projects provide reactive monitoring of selected data.

If you want to monitor entire data sources or schemas without needing to apply rules manually, consider using the Data Observability feature.

The screenshot shows the Ataccama Data Quality interface. On the left, a sidebar menu includes 'Data Quality' (selected), 'Rules' (Detection Rules, DQ Evaluation Rules), 'DQ Dimensions', 'Components', 'Transformation Plans', 'DQ Firewalls', 'Monitoring Projects' (selected), 'Reconciliation Projects', and 'Lookup Items'. The main area is titled 'Monitoring Projects' with tabs for 'Published', 'Unpublished', and 'All'. A search bar at the top says 'Type here to search full-text for Monitoring projects'. Below the search bar are filters for 'Terms' and 'Stewardship'. A table lists monitoring projects with columns: Name, Terms, Items, Latest run, and Stewardship. The table shows five rows: 'Customer DQ report' (PROD, Personal Data, 3 items, 4 months ago, Data Office), 'E-mail campaigns performance' (Marketing Department, 1 item, 8 months ago, Data Office), 'Financial Securities Failures' (Finance, 0 items, N/A, N/A), and 'xy\_Training Project' (N/A, N/A, 0 items, N/A). A callout box points to the search bar with the text: 'You can use the Search bar to browse the list of existing projects.' Another callout box points to the table with the text: 'Various highlights can be seen in the list of projects - name of the project, the project tagged Terms, number of Catalog Items, the time of the last run, stewardship.'

Name	Terms	Items	Latest run	Stewardship
Customer DQ report	PROD	Personal Data	3	4 months ago Data Office
E-mail campaigns performance		Marketing Department	1	8 months ago Data Office
Financial Securities Failures		Finance	0	N/A
xy_Training Project			0	N/A

► Projects are located under **Data Quality > Monitoring Projects**.

# Creating a new Monitoring Project

1. Go to the **Data Quality>Monitoring Projects**, click the **Create** button.
2. Specify a **Name** and optionally **Description** or **aim** of the project.
  - o **Configuration of the project** is done under the **Configuration & Results** tab.
  - o Once saved, it is possible to map **Glossary terms** to the project itself on the **Overview** tab.

The screenshot illustrates the process of creating and managing a monitoring project in the Ataccama Data Quality platform.

**Create Monitoring Project:** On the left, the "Create Monitoring Project" dialog is open. It shows the "General information" section where the "Name" field is filled with "xy\_trainingProject". Below it, the "Description" section contains rich text editor controls and a preview of the text: "New Monitoring Project: xy\_trainingProject CatalogItems - Customers, Products".

**Data Quality Sidebar:** A sidebar on the left lists various data quality components: Rules (Detection Rules, DQ Evaluation Rules), DQ Dimensions, Components, Transformation Plans, DQ Firewalls, Monitoring Projects (selected), Reconciliation Projects, and Lookup Items.

**Monitoring Projects Overview:** The main right-hand panel shows the "xy\_Training Project" details. The "Overview" tab is active, displaying the project's description ("xy\_Training Project") and general information. It includes sections for Monitored Catalog Items (2), Last Run (November 21, 2024, 2:00:37 AM), Last Modified (November 21, 2024, 1:27:59 AM), Purpose (+ Add Purpose), and Stewardship (Edit). A callout bubble points to the "Overall Quality Over Time" chart, which shows a line graph with a single data point at 74%.

**Callout Text:** A purple callout bubble contains the text: "If you re-run the project multiple times, you can monitor the data quality changes over time."

# Project configuration (1 / 4)

- The first step to configure a project is to **add Catalog items** to be monitored.
- That is done under the **Configuration & Results tab** in the **Items to Monitor** section using the '**+ Add**' button.
- Once the catalog items have been added, you'll need to select each item individually to begin **adding checks** by clicking on the item name.

**Three types of checks can be applied:**

- Structure checks
  - Anomaly detection
  - DQ rules
- 
- To remove an existing item, use the three dots menu and select **Delete**.

The screenshot shows the Ataccama Data Quality interface for the 'xy\_Training Project'. The 'Configuration & Results' tab is selected. At the top right, there is a 'Run monitoring' button. In the 'Items to monitor' section, there is a table with two rows: 'customers' and 'products'. For 'customers', there are 15 checks, 2 anomalies, and 5 DQ rules, resulting in an overall score of 74%. For 'products', the status is 'Not calculated yet'. An 'Add' button is located in the top right corner of the 'Items to monitor' table area. A callout box labeled 'Run the project' points to the top right button, and another callout box labeled 'Add a new Catalog Item to be monitored' points to the 'Add' button.

# Project configuration (2 / 4)

- After both adding or removing checks, make sure to **publish** your changes.
- After **Publishing** the project, evaluation should be triggered by pressing the **Run Monitoring** button.
- Configuration & Results** tab offers an overview of the data quality, issues found, etc.

To remove wrongly applied checks: Hover over the applied check in the table and delete using the X icon.

98% validation Surnai X

Switch between the historical evaluations.

Great! Your data is in good shape

Structure OK 1 check on 1 table

Anomaly Detection OK 2 checks on 1 table

Data Quality OK 5 checks on 1 table Overall VAL ACC 74%

Show invalid samples

Name	Structure	Anomalies	DQ Rules	Origin
customers	1 check	2 checks	5 DQ Checks 11+ suggestions	Postgres
products	15	9	3 DQ Checks 1+ suggestion	Postgres

74% 84% Postgres

customers xy\_training > postgres > public 122

products xy\_training > postgres > public 9

1+ suggestion

Not calculated yet Postgres

Check a sampling of invalid records.

# Project configuration (3 / 4)

- By clicking on each Catalog item, the details of the item's data quality as well as its configuration will become available.
- There, the user can assign **DQ rules** to **attributes**, turn **Structure Checks** on and enable **Anomaly Detection**.

Data Quality

Monitoring projects  
xy\_Training Project

Overview Configuration & Results Report Export Notifications History

November 21, 2024, 2:00:37 AM (Latest)

Great! Your data is in good shape

Tables 2 Attributes 24 Checks 8

Structure OK 1 check on 1 table

Anomaly Detection OK 2 checks on 1 table

Data Quality OK 5 checks on 1 table Overall VAL ACC

Items to monitor

Name	Structure	Anomalies	DQ Rules	Overall	VAL	ACC
customers	1 check	2 checks	5 DQ Checks	74%	74%	84%
products	9	1+ suggestion	Not calculated yet	Not calculated yet	Not calculated yet	Not calculated yet

Data Quality

Monitoring projects  
xy\_Training Project

Overview Configuration & Results Report Export Notifications History

xy\_Training Project > customers

November 21, 2024, 2:00:37 AM (Latest)

Great! Your data is in good shape

Records 122 Attributes 15 Checks 8

Structure OK 1 check

Anomaly Detection OK 1 attribute and 1 catalog item check Open profile inspector

Data Quality OK 5 checks Overall VAL ACC

Filter Attributes and Rules All Attributes

Standard view Hidden columns

Name	Terms	Filter by	Structure	Anomaly Detection	Overall Quality	Applied Rules
123 customernumber			Mandatory	+ Enable Detection		+ Add
abc customename			+ Make Mandatory	+ Enable Detection		+ Add
abc contactlastname	Summa		+ Make Mandatory	+ Enable Detection	84%	+ Add 84%
abc contactfirstname	firstname		+ Make Mandatory	+ Enable Detection	0%	+ Add 0%

# Project configuration (4 / 4)

The screenshot shows a project configuration interface for an 'XX\_Training project'. On the left, a sidebar lists monitoring projects. The main area displays a table of attributes with columns for Name, Terms, Filter by, Structure, Anomaly Detection, Overall Quality, and Applied DQ Checks.

**Anomaly Detection** discovers any inconsistencies or irregularities in the data and plays a key role in maintaining up-to-date and good-quality data.

**Validity** results are available after running the project.

To assign custom **DQ checks** use the plus icon under Applied DQ Checks and select the rules from the list.

**Terms** are imported from Catalog Items. They can also be manually added here.

**Attribute** values can be used for filtering data quality results in Reports. This action will need to run an evaluation first.

**Structure Checks** confirm whether all mandatory attributes have been successfully imported and are available for further checks.

# Rule suggestions

- The application can **suggest rules** for **DQ checks**.

## Rules are suggested in the following situations:

- When the **rule** is assigned to a **term** which is assigned to an **attribute**.
- When a **term** assigned to an **attribute** is **also assigned** to the **rule input** or **rule itself**.
- When the **rule** is assigned to the same **attribute** in **another monitoring project**.

Accept or reject suggestions one-by-one, using the relevant icon (tick for accept and cross for reject)

This screenshot shows a summary dashboard for the 'customers' dataset. It includes sections for 'Structure', 'Anomalies', and 'DQ Checks'. The 'DQ Checks' section displays 1 check, 2 checks, and 11 suggestions. A prominent yellow warning icon indicates 'DQ Issues Detected'.

This screenshot shows a detailed view of rule suggestions for the 'customers' dataset. The table lists attributes under 'Name' and their corresponding 'Applied DQ Checks', 'Terms', and 'Suggested Rules'. Each row shows a list of validation rules with icons for acceptance or rejection.

Name	Applied DQ Checks	Terms	Suggested Rules
customers		Personal Data	
123 customernumber			<input checked="" type="checkbox"/> Customer number Uniqueness <input checked="" type="checkbox"/> Greater than or equal to (Long) <input checked="" type="checkbox"/> Less than
Alic customername			<input checked="" type="checkbox"/> Length not longer than <input checked="" type="checkbox"/> Not empty (String) <input checked="" type="checkbox"/> Mask not match
Alic contactlastname	validation Surname accuracy Surname	Surname	<input checked="" type="checkbox"/> Length not longer than <input checked="" type="checkbox"/> Pattern not match <input checked="" type="checkbox"/> Not empty (String)
Alic contactfirstname	validation First name accuracy First name	First name	<input checked="" type="checkbox"/> Length not longer than <input checked="" type="checkbox"/> Pattern not match <input checked="" type="checkbox"/> Pattern match
Alic phone	[North America] Phone Number		<input checked="" type="checkbox"/> Length not longer than <input checked="" type="checkbox"/> Not empty (String) <input checked="" type="checkbox"/> String Completeness
Alic addressline1	Street with number		<input checked="" type="checkbox"/> Length not longer than <input checked="" type="checkbox"/> Not empty (String) <input checked="" type="checkbox"/> Mask not match
Alic addressline2	Street with number		<input checked="" type="checkbox"/> Length not longer than <input checked="" type="checkbox"/> Mask not match <input checked="" type="checkbox"/> Length not shorter than
Alic city			<input checked="" type="checkbox"/> Length not longer than <input checked="" type="checkbox"/> Pattern not match <input checked="" type="checkbox"/> Not empty (String)
Alic state	[USA] State		<input checked="" type="checkbox"/> Length not longer than <input checked="" type="checkbox"/> Pattern not match <input checked="" type="checkbox"/> String Completeness
Alic postalcode			<input checked="" type="checkbox"/> Length not longer than <input checked="" type="checkbox"/> Pattern not match <input checked="" type="checkbox"/> Mask not match

# Update rule references

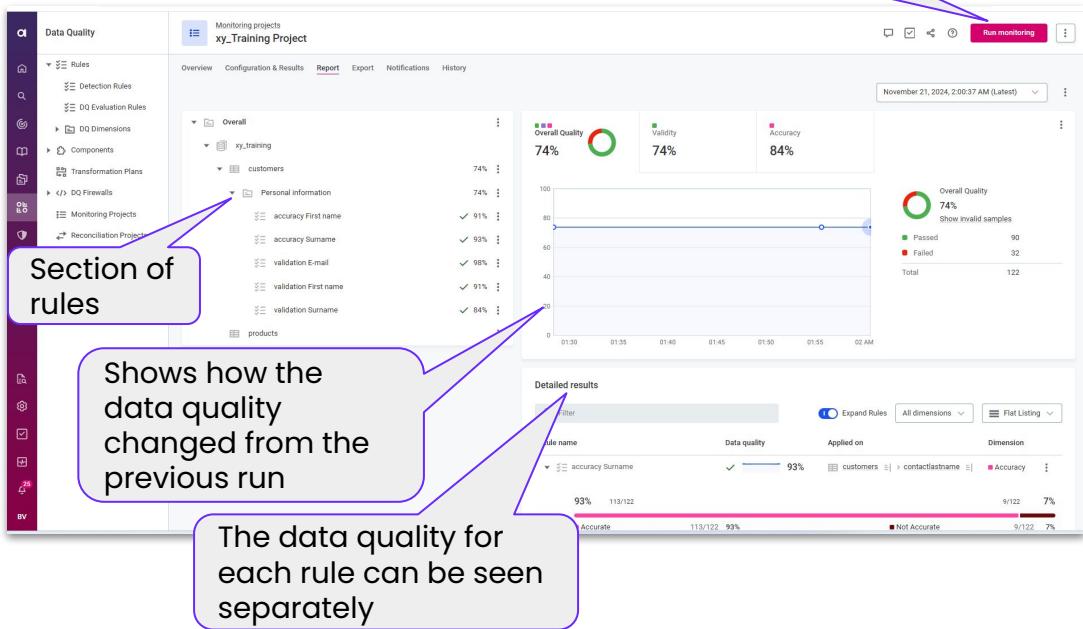
- If your monitoring project contains mappings to rules that have since been edited, you are notified on the **Configuration & Results** screen.
- To *update all rules* to the latest version, select **Update for this project** for the monitoring project or **Update for this catalog item** for a specific catalog item.
- To *update only selected rules*, find the outdated rules in the list of attributes, marked with a **lightbulb icon**.
- Select the rule to open the rule information, and then select See the details.
- You are redirected to the Preview changes screen: review the changes and then select Update Rule.
- Application admins can turn on and off these notifications for all users through the core:reference trait.
- If you change the trait after applying rules in the project, the trait doesn't apply to those rules.

The screenshot shows the 'Configuration & Results' tab of the Ataccama interface. A notification at the top states '1 DQ check is outdated.' with a lightbulb icon. Below it, a button labeled 'Update for this project' is highlighted with a purple border. The 'Applied Rules' section lists two rules: 'validation LONG Unique' (100% success) and 'test rule' (100% success). Both rules have lightbulb icons next to them.

This notification isn't visible if you have **view-only** permissions for monitoring projects.

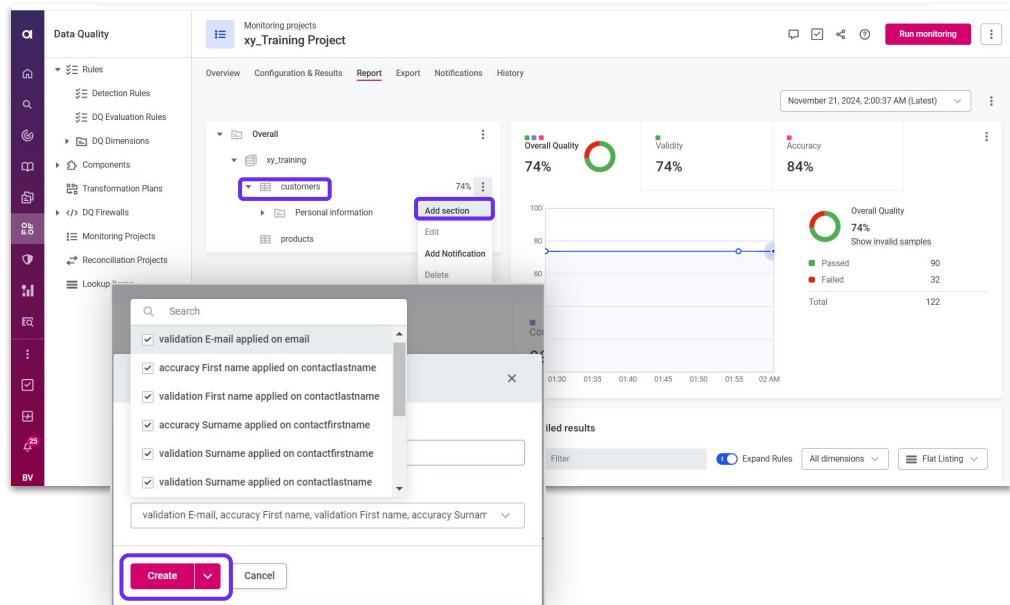
# Reports

- Once monitoring has finished, the DQ report is available on the **Report tab**
- DQ Reports** show the DQ results and detected anomalies **over time** for all monitoring runs.
- After the report configuration, the project should be **rerun** to see the results.
- Reports** can be customized to monitor DQ results for selected rules, rather than the whole Data Quality checks (see next slide).
- DQ Stewards are notified when a **new report** is available, and can see **DQ issues** (and expand details), **anomalies**, and the corresponding **profiling results**.



# Configuring Reports

- To create a rule **section**, click the Catalog Item option menu (3-dots), or an already existing section to create a sub-section, and select **Add section**.
- To report the data quality **over time** for separate rules, the user needs to create **Sections**.
- After implementing the changes, it is necessary to run the monitoring again to see the resulting reports.
- To see the data quality for all rules used in the monitoring project, expand the rules in the **Detailed results** list.



# Sample of invalid records (1 / 2)

- Once the **DQ monitoring** of a project is finished, the user can observe a **sample of invalid records** by clicking on the **Show invalid samples** either from the project **Result Overview** or from the **Report tab**.
- When there are issues detected in your monitoring project, investigating a sample of records that failed the applied DQ rules gives you a clearer idea of the issue at hand.

Data Quality ⓘ

**⚠️ Issues in 3 tables detected**

43 checks on 3 tables

[Show invalid samples](#)

The screenshot shows the 'Invalid Samples' section of the Data Quality tool. At the top left, there's a navigation bar with 'customers' and 'products'. On the right, it says '74%' and 'Sources > xy\_training > postgres'. Below this is a dropdown menu 'Attributes 3/15'. The main area is a table with 15 rows, each representing a customer record. The first few rows have red highlights over certain cells, indicating validation errors. A callout bubble points to one of these red-highlighted cells with the text 'Invalid values are highlighted with red color.' Another callout bubble points to the 'customers' table header with the text 'All Project Catalog Items can be accessed quickly.' A third callout bubble points to the '74%' metric with the text 'Filters can be applied to narrow down the list of displayed records.'

1	Piestrzewicz	Zbyszek	N/A INVALID_FORMAT: validat...
1	Murphy	Julie	miniwheelsco@gmail...
2	de Castro	Isabel	portoimportsco@gma...
2	Ranc  -	Martine	daedalusdesignsimpo...
	Bertrand	Marie	lacornedabondance,c...
2	Devon	Elizabeth	ukcollectables,ltd....
	Citeaux	Fr  -d  -rique	test@test.com
	Kloss	Horst	natrlichautos@gmail...
	Fresni  ire	Jean	qubechomeshoppingne...

# Sample of invalid records(2 / 2)

- You can configure how many invalid results are included in the sample or turn off this option completely.
- To access the configuration, on the **Configuration & Results** tab, expand the Data Quality three dots menu and select **Configuration of Invalid Results Samples**.

The screenshot shows two panels. The top panel is a summary card for 'Data Quality' with a green checkmark icon and the text 'OK'. It also displays '5 checks on 1 table' and a link to 'Show invalid samples'. The bottom panel is a modal dialog titled 'Invalid Samples configuration' with a checked toggle switch for 'Invalid samples enabled on Monitoring Project'. It includes a dropdown for 'Max number of invalid records per Catalog Item' set to '100', and two buttons at the bottom: 'Save' (pink) and 'Cancel' (blue).

Data Quality ⓘ

OK

5 checks on 1 table

Show invalid samples

Overall VALID INVALID ACO

Configuration of Invalid Results Samples

14/0

Invalid Samples configuration

Invalid samples enabled on Monitoring Project

Max number of invalid records per Catalog Item

100

Save Cancel

# Use data slices in monitoring projects

- If data slices have been created in the catalog items included in your monitoring project, you can run monitoring just on those data slices rather than the whole item.

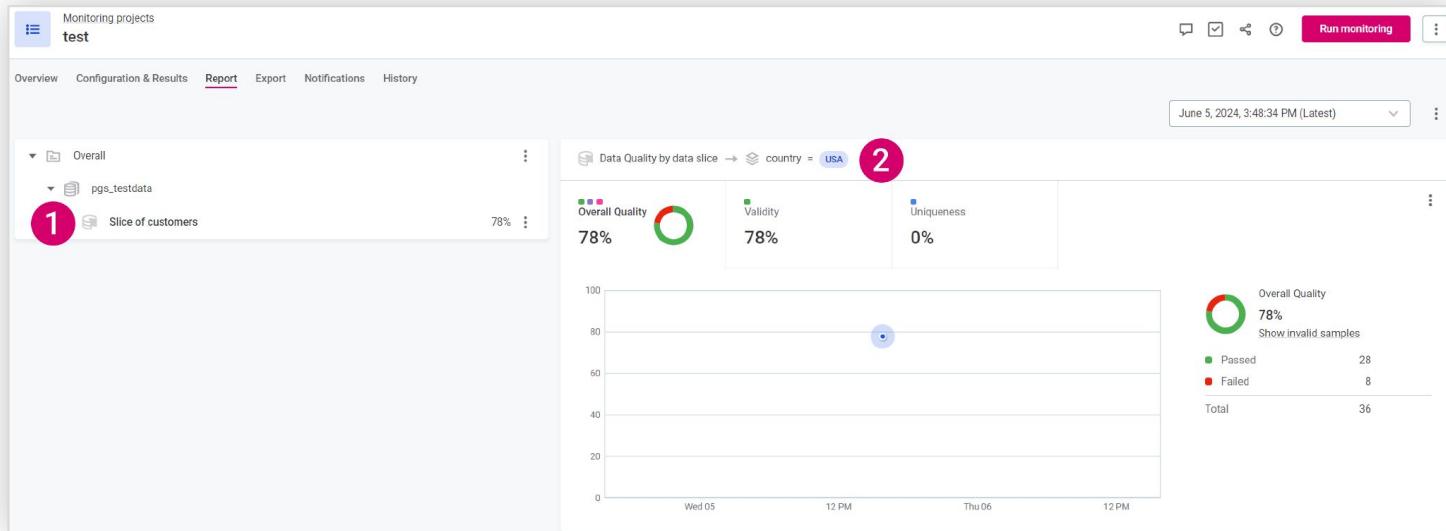
## To do this:

- On your monitoring project **Configuration & Results** tab, select the catalog item with data slices.
- Enable **Run monitoring project on data slice**.
- Click **Select**.
- Select the data slice you want to use, or create a new one.
- Repeat steps 3 and 4 to add additional slices.
- Publish** your changes.

The screenshot shows the Ataccama Monitoring Projects interface. The top navigation bar includes 'Monitoring projects', 'test', and various tabs like 'Overview', 'Configuration & Results' (which is active), 'Report', 'Export', 'Notifications', and 'History'. Below the tabs, the current view is 'test > customers DEV'. A pink box highlights the 'Run Monitoring project on data slice' toggle switch, which is turned on, and the 'Select' button next to it. The main content area displays monitoring details for the 'customers' slice, including 'Finish configuration first' (Attributes: 15, Checks: 0), 'Structure' (No checks applied), 'Anomaly Detection' (No checks applied), and 'Data Quality' (No checks applied). At the bottom, there's a 'Filter Attributes and Rules' section with 'All Attributes' and '18+ suggestions', and a table with columns 'Name', 'Terms', 'Filter by', 'Structure', and 'Anomaly D...'. The table has one row: '123 customernumber'.

# Use data slices in monitoring projects

- When you select Run monitoring, it now runs only on the selected data slices.
- This applies to all monitoring, that is, DQ evaluation, structure checks, and anomaly detection.
- Results from these slices can be seen on the Report tab: expand the results view and select your data slice (1).
- Details of the slice configuration can be found above the results for each run of the monitoring project (2).



# Notifications

**There are two available types of notifications:**

- A. General
- B. Specific

- Both **general** notifications and **specific** notifications are listed on the monitoring projects **Notifications** tab.
- From here, they can be **edited**, **turned on** or **off**, or **removed** as required.
- New general notifications can be added directly from the **Notifications tab** but new specific notifications are added from the **Report tab**.
  1. **Add** general notifications.
  2. **Edit** or **delete** general notifications.
  3. **Enable** or **disable** specific notifications.
  4. **Edit** or **delete** specific notifications.

The screenshot shows the Ataccama Monitoring interface with the 'Notifications' tab selected. The 'General' section contains two rows of notifications:

Name	Monitoring failed	New results	Structure error	DQ Issue	Anomalies	In App	Email
MP new results	✗	✓	✗	✓	✓	<span>+2</span>	john.taylor
MP failed	✓	✗	✓	✗	✗	<span>+2</span>	<span>+2</span>

The 'Specific' section shows one notification:

Name	Path	Threshold	Filters	Users	Active
Overall DQ Notification	< 70%			<span>1</span>	<span>3</span> <span>4</span> <span>⋮</span>

Annotations with numbers 1 through 4 point to specific UI elements: 1 points to the 'Add Notification' button, 2 points to the first row in the General table, 3 points to the user icon in the Specific table, and 4 points to the three-dot menu icon in the Specific table's footer.

# General Notifications

**Get automatic alerts about important monitoring events.**

## You Can Monitor:

- Monitoring failures
- Critical errors in data structures
- New results available
- Data quality issues and anomalies

## To Set Up General Notifications:

- Go to the "**Notifications**" tab in your monitoring project.
- Click "**Add Notification**" (or edit an existing one).
- Select the events you want to be notified about.

## How you want to be notified:

- In-app notification (select users)
- Email (select users)
- Slack (enter channel name)
- Microsoft Teams (insert webhook URL)

The screenshot illustrates the process of setting up general notifications in the Ataccama monitoring interface. It consists of three main panels:

- Monitoring projects xx\_Training Project**: Shows a table of monitoring configurations. A pink box highlights the "Run monitoring" button in the top right corner. Another pink box highlights the "Add Notification" button in the bottom right corner of the same row.
- Notification**: A modal window where a notification is being configured. The "Notification name" field contains "xx\_Notification1". The "Notify when" section lists several event types: "Monitoring failed" (selected), "A new version of the results is available" (selected), "Data Quality Issue" (selected), "Critical Error in structure" (unchecked), "Anomalies" (unchecked). A pink box highlights the "Monitoring failed" checkbox. A pink arrow points from this panel down to the "Channels" panel.
- Channels**: A modal window showing the selection of notification channels. It includes tabs for "In app" and "E-mail", with "E-mail" selected. Under "In app", "In app" is checked. Under "E-mail", "E-mail" is checked. Other options like "Slack" and "Microsoft Teams" are also listed. A pink box highlights the "In app" checkbox under the "In app" tab. A pink arrow points from this panel up to the "Notification" panel.

# Specific Notifications

- Get notified about very specific data quality issues.
- **Granular Thresholds:** Define the exact data quality percentage that triggers an alert.
- **Filter Conditions:** Narrow down notifications based on specific criteria (e.g., date range, attribute values).
- **Targeted Delivery:** Sent as emails to designated users or roles.

## How to Create Specific Notifications:

1. Go to the "**Report**" tab.
2. Select the instance or rule you want to monitor.
3. Click the three-dot menu and choose "**Add Notification**."
4. Configure the notification.
5. Click "**Save and publish**."

Name	Monitoring failed	New results	Structure error	DQ Issue	Anomalies	In App	Email	Slack	MS Teams
XX_Notification	✓	✓	✓	✓	×	bliss.varghese@ataccama.c...	bliss.varghese@ataccama.c...	×	×

Name	Path	Threshold	Filters	Users	Active
customers Notification 1	XX_training > customers	< 70%		No recipients	<input checked="" type="checkbox"/>

# Topic Highlights

- **Monitoring Projects** evaluate the **data quality** of catalog items and monitor **changes in data quality over time**.
- Monitoring projects with outdated rules display notifications on the *Configuration & Results* screen.
- **Report** on quality changes over time and monitor data quality results for selected rules.
- General notifications for broad monitoring alerts and specific notifications for targeted data quality thresholds and filter conditions.

The screenshot shows the Ataccama Data Quality interface. On the left is a dark sidebar with icons for Home, Search, Configuration, Components, Transformation Plans, DQ Firewalls, Monitoring Projects (which is highlighted), Reconciliation Projects, and Lookup Items. The main area has a header 'Monitoring Projects' with tabs for Published, Unpublished, and All. It includes a search bar and filters for Terms and Stewardship. Below is a table with columns: Name, Terms, Items, Latest run, and Stewardship. The table contains four rows:

Name	Terms	Items	Latest run	Stewardship
Customer DQ report	PROD	Personal Data	3	4 months ago Data Office
E-mail campaigns performance		Marketing Department	1	8 months ago Data Office
Financial Securities Failures		Finance	2	over 1 year ago Finance
xy_Training Project			2	about 15 hours ago

# Memory Refresher #3 Monitoring Projects



# Lab Exercise #2

## Basic Monitoring Projects

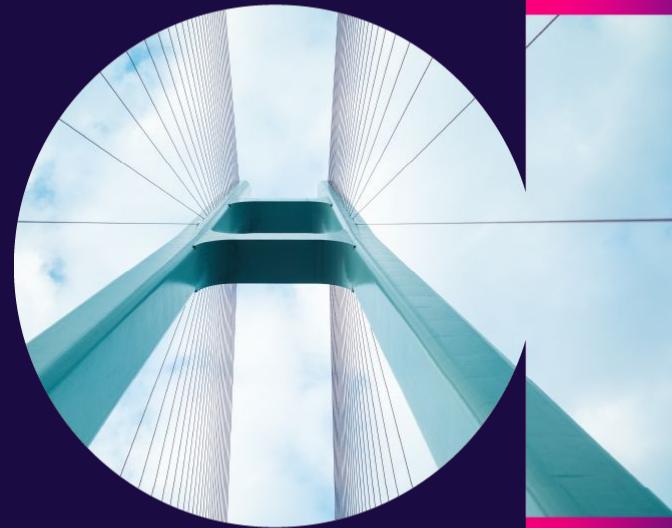


# Lab Exercise #3

## Aggregation Rules



# Data Reconciliation



# Reconciliation Projects

- The data reconciliation feature provides **high-level checks of data consistency** across multiple sources, comparing attribute profile statistics of selected data.
- **Fingerprints of attributes** are also **compared** (except for Snowflake data), which allows you to verify, for example, that the target data is the same as the source data when migrating from one database to another, so you know that the original data was transferred **without any data leakages or malformations**.
- Issues detected during data reconciliation are provided in a **human-readable** format.

The screenshot shows the Ataccama Data Quality interface. On the left, a sidebar menu is open, showing various options under 'Data Quality': Rules, Detection Rules, DQ Evaluation Rules, DQ Dimensions, Components, Transformation Plans (which is highlighted with a purple box), DQ Firewalls, Monitoring Projects, Reconciliation Projects (which is also highlighted with a purple box), and Lookup Items. The main area is titled 'Reconciliation Projects' and contains a table with three rows:

Name	Description	Stewardship
Oracle to Snowflake migration		
Employee data (Join test)		

# Create a reconciliation project in Data Quality

- To get started with data reconciliation, select **Data Quality**, then **Reconciliation Projects** and select **Create**
- If you have existing data reconciliation projects, they are listed here. Select the project name to open the project, **edit** the configuration, or **view** results.

The screenshot shows the Ataccama Data Quality interface. On the left is a sidebar with various icons and a count of 25 items. The main area is titled "Data Quality" and shows a list of project components: Rules (Detection Rules, DQ Evaluation Rules), DQ Dimensions, Components, Transformation Plans, DQ Firewalls, Monitoring Projects, Reconciliation Projects, and Lookup Items. A specific project, "xy\_trainingReconciliationP...", is selected, indicated by a blue header bar with "Reconciliation projects" and the project name. Below the header are tabs for "Overview" (which is active), "Results", and "History". A modal menu is open over the "Edit" button, listing options: "Edit", "Schedule", "Notification configuration", "Delete", and "Edit page template". The "Overview" section contains fields for "Description" and "Glossary terms" (with a "+ Add Term" button). The "Mapped sources" section is currently empty, showing a document icon and the text "No mapped sources" and "No mapped sources have been added yet, add one to get started." A pink "Add sources to compare" button is at the bottom of this section.

# Select data origin and target

The key element of data reconciliation is the data mapping between origin and target.

- You can select an **entire source** or **browse** and **select specific locations** or **catalog items** to compare a subset of the source.
- Once you have specified the **origin** and **target**, Ataccama ONE looks at the **currently imported catalog items** of the two sources and **matches** as many as possible by **name-by-name** and **data structure**.
- You need to map **unmapped** data **manually**, and then you can override the automatic mapping with manual mapping if required.

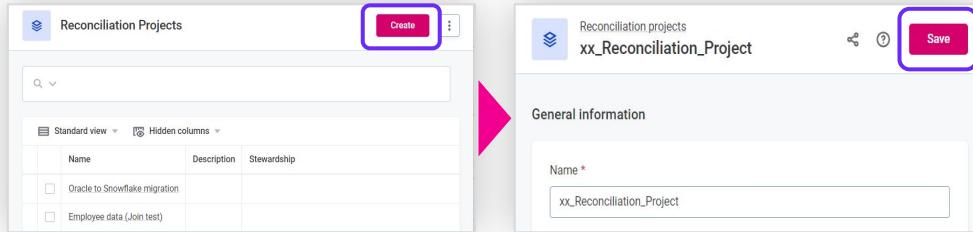
It is possible to compare:

- A single data source against another data source
- Multiple locations (schemas)
- Multiple catalog items
- A data source and a location
- A location and selected tables

# How to configure reconciliation project (1/2)

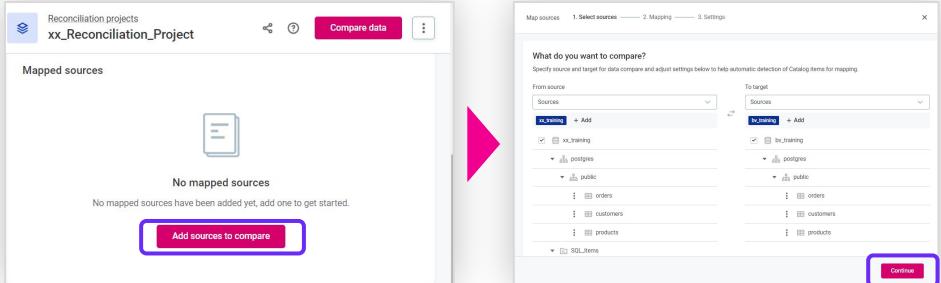
1. From **Data Quality > Reconciliation Projects**, Create and name a reconciliation project.

1



2. Specify the data **origin** and **target**:
  - Sources
  - Location
  - Catalog Items

2



# How to configure reconciliation project (2/2)

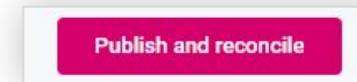
3. **Check** mapping/map unmapped items or **override** an automatic mapping.
4. Configure **Settings**:
  - Auto-mapping (for a new project auto-mapping is enabled by default).
  - Detect issues with data types.
  - Use existing profiling.
5. Run the reconciliation by selecting **Publish and Reconcile** for a new project or **Compare data** for an existing one.

Once data reconciliation finishes, results are available on the **Results** tab.

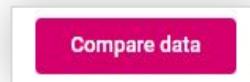
The screenshot shows the Ataccama reconciliation interface. At the top, there's a header with a logo, 'Reconciliation projects', and a 'Compare data' button. Below the header, there are three tabs: 'Overview', 'Results' (which is highlighted with a blue border), and 'History'. Underneath the tabs, it says 'You compare: Snowflake → Snowflake' and shows the date 'April 17, 2023 5:28:29 PM (Latest)'.

This screenshot shows the 'Mapping' step of the reconciliation configuration. It has a header 'Map sources' with tabs for '1. Select sources', '2. Mapping', and '3. Settings'. Below the header is a table with four columns: 'All items' (162), 'Unmapped items' (150), 'Mapped items' (12), and 'Ignored items' (0). The main area shows a list of mappings with columns for 'Mapping' (status: Manual, Automatic, Unmapped, AUTOMATIC\_PARTIAL), 'Source asset' (e.g., 00\_Orders\_Cancelled, orders, New\_customers, customers, products, test, customer, customer\_two, names), and 'Target asset' (e.g., orders, orders, ???, customers, products, ???, ???, names). A pink circle with the number '3' is positioned at the bottom center of the mapping table.

This screenshot shows the 'Data compare settings' and 'Data freshness' sections. In 'Data compare settings', there are two checkboxes: 'Enable auto-mapping suggestions' (checked) and 'Detect issues with data types' (unchecked). In 'Data freshness', there are sections for 'Source' and 'Target', each with a checkbox 'Use existing profiling' (checked) and a dropdown menu set to 'Any time'. Pink circles with the numbers '4' and '5' are positioned to the right of these sections.



5



ataccama

# Reconciliation Projects: Results

The screenshot shows the Ataccama Data Quality interface. On the left is a sidebar with various navigation icons and a list of projects: Data Quality, Rules, Components, Transformation plans, DQ Firewalls, Lookup items, Monitoring projects, and Reconciliation projects. The 'Reconciliation projects' icon has a red '35' badge.

The main area displays a reconciliation report for 'XX\_training' against 'XX\_training'. It shows 10 issues found, categorized by type: Inconsistent Catalog items (10/16), Missing Catalog items (5/16), and Failed jobs (3/5). A callout points to the 'Results' tab in the top navigation bar.

A table below lists reconciliation details for various assets:

Source asset	Target asset	Issues	# Records	# Attributes
TOP XX_Orders Cancelled	orders	Target has 320 more records (5333%)	6 +320	2
TOP payments_prime	products	163 records are missing in target (60%), 4 attributes are missing in target	273 -163	4
TOP employees_prime	customers	Target has 99 more records (430%), 7 attributes are missing in target	23 +99	8
TOP productlines_prime	KS_Orders_transformed	Target has 319 more records (4557%), 12 attributes are missing in target	7 +319	12

Callouts provide additional information:

- A callout points to the 'Results' tab in the top navigation bar with the text: "A table listing all issues can be found in the Results tab".
- A callout points to the 'History' button in the top navigation bar with the text: "View historical versions of results".
- A callout points to the 'Notification configuration' option in the context menu of a specific row with the text: "You can schedule or configure notifications".
- A callout points to the bottom right of the table with the text: "Results are shown on the attribute level, but not on a record level".

# View results

- A table listing of all issues can be found on the **Results** tab for a selected project, as well as an overview of the number of catalog items which have issues, or are missing in the target location.
- Results can be seen on the level of the attribute but not on a record level.

## Possible issues include:

- Mismatches in data types
- Missing records
- Missing values
- Missing columns
- Duplicated records
- Duplicated values in a field
- Unexpected null values
- Badly formatted values
- Other inconsistencies in values (by comparing min, max, length, and quantiles).

Inconsistent Catalog items	Missing Catalog items	Failed jobs
⚠ 3/162	⚠ 150/162	✓ 0/3
Source asset	Target asset	Issues
TOP ▾ 00_Orders_Cancelled	orders	Target has 320 more records (5333%)
123 ordernumber	ordernumber	Different number of unique values (5333%)
Abc comments	Abc comments	246 values are missing, 43 more duplicates, Attributes have different fingerprints,
TOP ▾ 00_Orders_transformed	orders	1 attribute is missing in target
123 processing_time	-	
▼ customers	customers	1 inconsistent attribute
123 creditlimit	123 creditlimit	1 fewer duplicates (<1%), Different number of distinct values (<1%), Different num

# Topic Highlights

- **Data Reconciliation** provides **high-level checks** for data **consistency** across **multiple sources**.
- **Data mapping** between **origin** and **target** is crucial for **reconciliation**.
- **Compares** attribute **profile statistics** and **fingerprints** of attributes (except for Snowflake data).
- Use Cases for Reconciliation include: **Verify data consistency** during database migration, **detects issues** such as **mismatches**, **missing records**, **duplicated values**, and **more**.
- You can override **automatic mapping** with **manual adjustments** if needed.
- Supports comparisons between **multiple locations**, **catalog items**, and **combinations** of **data sources** and **locations**.

# Memory Refresher #4 Data Reconciliation



# ONE Data



ataccama

# ONE Data

ONE Data is a module that allows to build trustworthy and governed data products.

## With ONE Data you can:

- Create, configure, and maintain reference data.
- Remediate and fix data errors directly from ONE.
- Leverage ONE features, such as data processing, DQ evaluation capabilities, collaboration and reporting options, and more.

## Additionally:

- ONE Data is a type of data source in ONE.
- You can access the metadata of each ONE Data table in the Knowledge Catalog as well.
- These tables are labeled as ONE Data catalog items.

The screenshot shows the Ataccama ONE Data Tables interface. On the left is a vertical toolbar with icons for Home, Search (highlighted with a purple box), Refresh, Import, Export, Data Processing, Security, and Analytics. Below the toolbar is a badge with '25' and 'BY'. The main area is titled 'ONE Data Tables' and displays a list of tables. Each table entry includes a 'Name' column (e.g., 'xy.US\_Customers', 'Turkey city', 'Saudi Arabia city', 'Airports', 'HTTP Request', 'R&D', 'HTTP Response Code', 'Italy city', 'Conference Attendees', 'Indonesia city', 'Country'), a 'Terms' column, an 'Anomalies' column, an 'Overall Quality' column (with a progress bar), a '# Attributes' column, a '# Records' column, an 'Origin' column, and a 'Loc' column. A search bar at the top is empty. A dropdown menu shows 'Standard view' and 'Hidden columns'.

Name	Terms	Anomalies	Overall Quality	# Attributes	# Records	Origin	Loc
xy.US_Customers	-	-	-	10	one-data		
Turkey city	-	-	-	2	167	one-data	
Saudi Arabia city	-	-	-	2	87	one-data	
Airports	International Airport Code	-	-	2	1,942	one-data	
HTTP Request	-	-	-	2	9	one-data	
R&D	First name, Surname, E-mail, +2	23%	-	7	1,000	one-data	
HTTP Response Code	-	-	-	2	48	one-data	
Italy city	ITA City	>99%	-	2	144	one-data	
Conference Attendees	First name, Surname, E-mail, +4	52%	-	11	199	one-data	
Indonesia city	-	-	-	2	230	one-data	
Country	Country, ISO-2 Country Code, +1	66%	-	5	240	one-data	

# ONE Data: Interface

The **key difference** between **Catalog Items** and **ONE Data tables** is that ONE DATA represents **actual data** stored in the application, which allows you to **modify the content**.

The screenshot shows the Ataccama ONE Data interface. On the left is a sidebar with icons for Home, Search, Catalog, and Data. The main area is titled "ONE Data Tables" and displays a list of tables with columns for Name, Terms, Anomalies, Overall Quality, # Attributes, # Records, and Origin. A callout points to the "Hidden columns" button in the toolbar above the table, with the text "Hide irrelevant information". Another callout points to the left edge of the table area, with the text "Switch between the Standard and the Compact view to change the column height to best fit your screen size.". A third callout points to the "Create table" dialog on the right, with the text "Create a new table manually or Import an existing a catalog item from the Knowledge Catalog or upload a local file from your machine.". The "Create table" dialog includes options for "From Catalog Item", "From file", and "From local file".

# ONE Data: Table details

- **Data:** Quickly view information such as table data, key metadata (attribute data types, glossary terms), and DQ results.
- **Overview:** Metadata information.
- **History:** Previous versions of the table.
- **Lineage:** A MANTA graph of connections between items in different data sources.
- **Data Quality:** DQ evaluation results.
- **Profile & DQ insights:** Profiling results, including DQ insights.
- **Relationships:** Connections between tables from the same data source.
- **Data Export Project:** An overview of data export projects.

The screenshot shows the ONE Data Tables interface for a 'Country' table. The top navigation bar includes tabs for Data, Overview, History, Lineage, Data Quality, Profile & DQ Insights, Relationships, and Data Export & Transformations. A purple box highlights the 'Data' tab. Below the navigation is a filter bar with a 'Filter' dropdown set to '66%', a note about DQ results being outdated, and an 'Evaluate Data Quality' button. A modal window titled 'Filter records' is open, showing a 'WHERE' clause with a dropdown menu labeled 'Select attribute...'. The main table lists 14 countries with columns for Name, Alpha-2, Alpha-3, Dialing Code, and dmm\_record\_id. The table has a purple border around its data area. On the left, there's a sidebar with icons for Home, Search, Refresh, and other data management functions. At the bottom right, it says 'Total records: 24'.

Rank	Name	Alpha-2	Alpha-3	Dialing Code	dmm_record_id
1	Afghanistan	AF	AFG	93	6cfb8991-d465-4198-b20e-
2	Albania	AL	ALB	355	bd388533-aed8-4a3f-8688-
3	Algeria				6-a6c5-
4	American Samoa				3-9c63-
5	Andorra				e-8c0c-
6	Angola				0-a136-
7	Anguilla				.d-9f1c-
8	Antarctica	AQ	ATA	672	948c444b-980f-45eb-9245-
9	Antigua and Barbuda	AG	ATG	1-268	631c06fb-4dfe-4447-93fe-
10	Argentina	AR	ARG	54	97632a6a-8924-460f-9237-
11	Armenia	AM	ARM	374	f89e0bcc-e21e-499b-8c7f-
12	Aruba	AW		297	84def0f5-89af-44ef-8dd4-
13	Australia	AU		61	68248c6e-861b-498c-a659-
14	Austria	AT		43	49267942-018f-4a2a-955a-

ONE Data provides extensive filtering choices, making it simple for you to locate, analyze, and modify the accurate records.

# Creating a ONE Data Table

You can **create a ONE Data table through:**

- Creating an empty table
- Importing from catalog items
- Uploading a CSV file.

The screenshot shows the Ataccama ONE Data Tables interface. On the left is a sidebar with icons for Home, Search, Catalog, Data, Protection, Analytics, Configuration, and Notifications (with 25 notifications). The main area is titled "ONE Data Tables" and displays a list of existing tables. Each table entry includes a checkbox, the table name, a "Terms" section, an "Anomalies" section, and columns for "Overall Quo", "Rows", "Last Update", and "Owner". Below the table list is a search bar and a "Hidden columns" dropdown. A large purple rounded rectangle highlights the top-right corner of the interface, containing three options: "Create table" (Start fresh with a new empty table), "From Catalog Item" (Load all or only invalid records from a catalog item), and "From file" (Import data from a CSV or XLS files stored on your local drive). The "From file" option is specifically highlighted with a blue border.

# Creating an empty table

1. Click **Create** table in the top-right corner.
2. Change the table name.
3. Customize the table structure.
4. Once the table structure is ready, you can fill in your records.
5. Verify whether the stewardship is correctly set for this table. Unless modified, the table owner and roles are inherited from the data source.

The screenshot shows the Ataccama ONE Data Tables interface. At the top right, there is a 'Create table' button with a purple rectangular highlight around it. Below the button is a sub-menu with options: 'From Catalog item', 'From file', and 'Import data from a CSV or XLS files stored on your local drive'. The main area displays a list of existing tables, each with columns for Name, Terms, Anomalies, Overall Quality, # Attributes, # Records, and Origin. A pink arrow points from the bottom of the previous screenshot towards this interface.

The screenshot shows the Ataccama ONE Data Tables interface with a newly created table named 'Table ab59d3bd-691a-4140-af48-7856407a54c5'. The table has three columns: 'Attribute 1', 'Attribute 2', and 'Attribute 3', each with three rows of data. The data in 'Attribute 1' is '1', '2', and '3'. The data in 'Attribute 2' is empty. The data in 'Attribute 3' is '01b618f6-6c39-410f-aef9...', '74ec5856-c3c0-4e27-8842...', and '6ccaf844-8dc8-44e8-91e3...'. The interface includes tabs for Data, Overview, History, Lineage, Data Quality, Profile & DQ Insights, Relationships, and Data Export & Transformations. A pink arrow points from the bottom of the previous screenshot towards this interface.

# Importing Data from a Catalog Item (1/2)

1. To start your import, In ONE Data, select **Import > Catalog Item**.
2. Find and **select the catalog item** that you want to use.
  - You can **import** an entire catalog item with or without DQ results, only data, or deduplicated data.
  - Once data is loaded, you can enhance it by **adding, removing attributes** or **modifying the table values**.

ONE Data Tables

Create table

From Catalog Item

From file

Name	Terms	Anomalies	Overall Quality	# Attributes	# Records	Origin	
xy_US_Customers	-	-	-	10	one-data	DMM > default	
Table f5c3b765-3c3a-44ae-81be-f704ffdfab06	-	-	-	4	one-data	DMM > default	
Turkey city	-	-	-	2	167	one-data	DMM > default
Saudi Arabia city	-	-	-	2	87	one-data	DMM > default
Airports	International Airport Code	-	-	2	1,942	one-data	DMM > default
HTTP Request	-	-	-	2	9	one-data	DMM > default
R&D	First name   Surname   Email   +2 *	23%	23%	7	1,000	one-data	DMM > default
HTTP Response Code	-	-	-	2	48	one-data	DMM > default
Italy city	ITA City	>99%	>99%	2	144	one-data	DMM > default
Conference Attendees	First name   Surname   E-mail   +4 *	52%	52%	11	199	one-data	DMM > default
Indonesia city	-	-	-	2	230	one-data	DMM > default
Country	Country   ISO-2 Country Code   +1 *	66%	66%	5	240	one-data	DMM > default

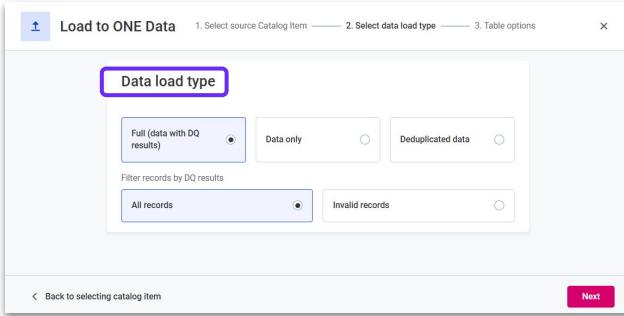
Load to ONE Data

1. Select source Catalog Item    2. Select data load type    3. Table options

Name	Terms	Overall Quality
orders	-	-
customers	Country   First name   Surname   +4 *	23%
products	K\$ Scale   K\$ Product Line	100%
ks_Orders_Cancelled	-	-
Table 4e037114-338f-4f9a-afc1-204bb6c24ef3	-	-
KS_Orders_transformed	-	-
ks_customers_phone	-	-
ks_product_info	-	-

# Importing Data from a Catalog Item (2/2)

3. Once you select the catalog item, **select the export type**.



## Export Type:

- a. Full (data with DQ results)

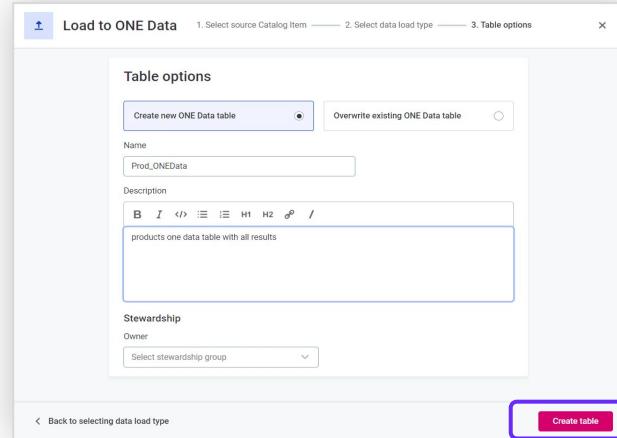
## Filter Records by DQ Results:

- All Records
- Invalid Records

- b. Data only
- c. Deduplicated data

4. Click **Next** and choose one of the two options:

- Create new ONE Data table:
  - Enter a **unique name** and optionally a description.
- Overwrite existing ONE Data table:
  - In Target table, **find and select** which table you want to overwrite.
  - Then '**Create Table**'.



# Configure Visibility of attributes and DQ results

- ONE Data lets you hide attributes and data quality results from view on the **Data** tab.
- To configure this, select which attributes are hidden or shown from the visibility menu.
- You also have the option to hide all **Technical attributes** and **Data quality results**.
- To make all previously hidden attributes appear, select **Make all visible**.

The screenshot shows the ONE Data Tables interface for the 'Country' table. The sidebar on the left has a dark purple header with icons for Home, Search, Filter, Relationships, and Data Export & Transformations. Below this is a red 'Data' button, followed by 'Overview', 'History', 'Lineage', 'Data Quality', 'Profile & DQ insights', 'Relationships', and 'Data Export & Transformations'. The main area is titled 'Country' and shows a table with columns: 'Name' (with filters for 'Country', 'ISO-2 Country Code', 'ISO-3 Country Code', and 'Dialing Code'), 'Alpha-2', 'Alpha-3', and 'Dialing Code'. The table lists 12 countries with their codes and populations. At the bottom left is a 'BV' button with a plus sign and 'Add record'. At the bottom right is a note 'Total records: 240'. On the far right, there's a vertical search bar with a magnifying glass icon, containing options like 'Data quality results', 'Technical attributes', and 'Make all visible'. The 'Make all visible' option is highlighted with a purple box. Another purple box highlights the visibility menu icon in the sidebar.

	Name	Alpha-2	Alpha-3	Dialing Code
1	Afghanistan	AF	AFG	93
2	Albania	AL	ALB	355
3	Algeria	DZ	DZA	213
4	American Samoa	AS	ASM	1-684
5	Andorra	AD	AND	376
6	Angola	AO	AGO	244
7	Anguilla	AI	AIA	1-264
8	Antarctica	AQ	ATA	672
9	Antigua and Barbuda	AG	ATG	1-268
10	Argentina	AR	ARG	54
11	Armenia	AM	ARM	374
12	Aruba	AW	ABW	297

# Data Deduplication (1/2)

- You can **deduplicate** your datasets to easily create **managed reference data**.
- Once your reference data is ready, use it in **DQ rules** to continuously improve the quality of your data.

## Creating a reference table

1. If you selected Deduplicated data in the previous step, you can now choose which attributes the new ONE Data table should contain. (at least one attribute to proceed, however, you can edit at the following step), then click Next.
2. Select the attribute or a combination of attributes based on which the data will be deduplicated. If needed, select Add or remove attributes to modify the selection then click Next.

The screenshot shows the 'Export to ONE Data' interface with the 'Deduplicate data' step selected. The table lists attributes with their data types and names. The first row for 'job\_title' has a checked checkbox and is highlighted in blue. The second row for 'country' has an unchecked checkbox. To the right of the table is a button labeled 'Add or remove attributes' with a purple border.

	Data type	Name	Overall quality	Terms	Evaluated
<input checked="" type="checkbox"/>	Abc	job_title	-		-
<input type="checkbox"/>	Abc	country	-		-

3. Enter a unique name for the table.
4. Select Create table. Depending on the size of your catalog item, it might take a few minutes To continue working with the platform in the meantime, select Run in background.

# Data Deduplication (2/2)

You can also deduplicate data (create a reference table) from the following locations:

- From the ONE Data table detail screen:  
On the **Data** tab, select one or more attributes, then choose Create **reference** table in the banner that appears.

The screenshot shows the ONE Data Tables R&D interface. On the left, there's a sidebar with 'Data', 'Overview', 'History', 'Lineage', 'Data Quality', 'Profile & DQ insights', 'Relationships', and 'Data Export Projects'. The main area has tabs for 'Data' (selected), 'Overview', 'History', 'Lineage', 'Data Quality', 'Profile & DQ insights', 'Relationships', and 'Data Export Projects'. A 'Filter' section is at the top. Below it is a table with columns: First Name, Last Name, Surname, Email, Phone, Age, and Credit limit. The table contains several rows of data. At the bottom of the table, there's a banner with the text 'This is a Data Sample' and 'Displayed data is live. The data displayed is a sample of the first 50 lines of the d...'. Below the banner, there are three columns: 'first\_name', 'last\_name', and 'email'. Each column has a 'Create reference table' button. The 'first\_name' column also has a 'Create new lookup' button.

The screenshot shows the Conference Attendees catalog item detail screen. The top navigation bar includes 'Sources > one-data > DMM > default' and tabs for 'Lineage', 'Data Quality', 'Profile & DQ insights', 'Overview', 'History', 'Data' (selected), and 'Relationships'. Below the navigation is a section titled 'This is a Data Sample' with the subtext 'Displayed data is live. The data displayed is a sample of the first 50 lines of the d...'. At the bottom, there are three columns: 'first\_name', 'last\_name', and 'email'. Each column has a 'Create reference table' button. The 'first\_name' column also has a 'Create new lookup' button.

- From the catalog item Overview, Profile & DQ Insights, or Data tabs: Expand the three dots menu for a particular attribute and select Create reference table.

- From the catalog item detail screen: Expand the three dots menu for a catalog item and select Load to ONE Data, then follow the steps described in this article.

# Importing Data from a File (1/2)

## 1. To start your import from a file, select **Import > Catalog Item**.

- You can import only one file at a time.
- Supported formats are : CSV, TSV, XLS, and XLSX.
- The maximum upload size is limited as follows:
  - Microsoft Excel files (XLS or XLSX): 50 MB.
  - CSV or TSV files: 2 GB.
- Before loading any data, a validation check is executed to prevent the import of files with incorrect formatting.

The screenshot shows the Ataccama ONE Data Tables interface. On the left is a sidebar with various icons. The main area displays a list of tables under 'Standard view'. One table, 'International Airport Code', is selected, showing its schema with columns: Name, Terms, Anomalies, Overall Qua, First name, Surname, E-mail, and a status bar indicating 23% completion. A modal window titled 'Create table' is open, with the 'From file' option highlighted by a purple rectangle. The 'From Catalog Item' and 'From scratch' options are also visible.



The screenshot shows the 'File import' dialog. It has two tabs: 'Source file' (selected) and 'Create table'. Under 'Source file', there is a dashed box for dragging files and a 'Browse files...' button, which is also highlighted with a purple rectangle. Below the file selection area, a message states: 'Drag & drop your files. Only CSV, TSV, XLS and XLSX files are supported.'

# Importing Data from a File (2/2)

Once the file selected ,verify data correctness in the file preview

Double-click the column header to edit values, then confirm changes.

The screenshot shows the 'File import' interface. On the left, there's a 'Source file preview' window displaying a table with columns: customernumber, customername, contactfirstname, and contactlastname. The preview shows several rows of data from an Excel file named 'orders\_transformed.xls'. On the right, there's a 'Source file' configuration panel. It includes a 'Select sheet' dropdown set to 'Sheet0', an 'Import options' section with a checked 'Use first row as header' checkbox, and a 'File settings' section where the file type is set to 'XLS'. A 'Continue' button is at the bottom right of the configuration panel.

For Excel files with multiple sheets, choose the desired sheet in "Select sheet"; otherwise, the first sheet is selected.

By default, all columns are selected for import. To select specific columns, use the shortcut or click "Select specific columns to import" and make your selection.

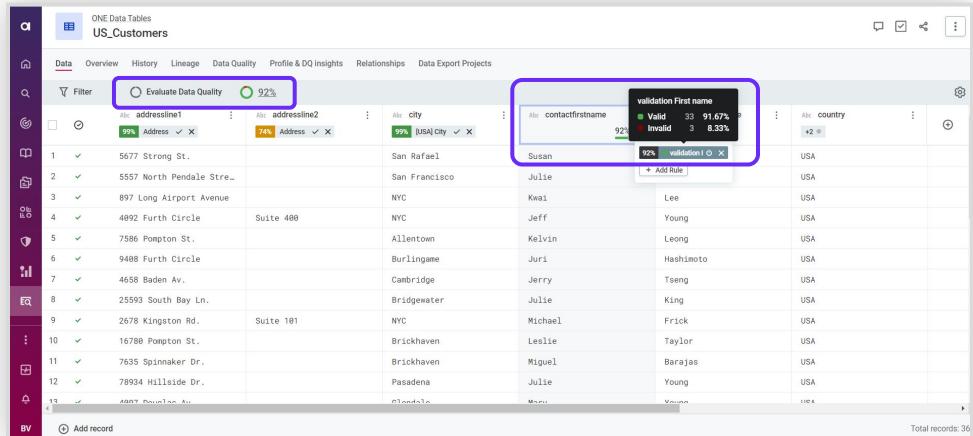
Choose "Use first row as headers."

Click on the File Settings to modify file format, value and decimal separators, date and date-time formats, and file encoding as needed.

Once you're happy with your choices, select Continue.

# DQ Evaluation in ONE Data (1/2)

- Once a table created in ONE Data, you can verify data validity by executing **Full Profiling & DQ Evaluation**.
- This feature is accessible through the Knowledge Catalog or the Profile & DQ insights tab if the table hasn't been profiled yet.
- It conducts a thorough **analysis of the entire table**, making it ideal for initial processing.



# DQ Evaluation in ONE Data (2/2)

- Business terms relevant to the attribute are displayed in the attribute header, and DQ results are visible on both attribute and row headers.
- To interpret DQ evaluation results, consider the following metrics:**
  - Overall quality
  - Attribute quality
  - Invalid records
- Once initial DQ results are available, real-time validation occurs as records are edited.
- This validation is reflected in DQ results at the cell and row levels, along with DQ filtering outcomes.
- DQ results at the attribute level and the overall **Data quality remain unchanged until another DQ evaluation is performed.**

Number of records in the table that passed all applied DQ rules.

Move your cursor over the warning to identify the DQ rules that the record did not successfully pass.

Validity: validation First name - IS\_INVALID

	addressline1	city	contactfirstname	contactlastname	country
13	4897 Douglas Av.	Glendale	Mary	Young	USA
14	361 Furth Circle	San Diego	Valarie		USA
15	8489 Strong St.	Las Vegas	Jean	King	USA
16	149 Spinnaker Dr.	New Haven	Keith	Franco	USA
17	39323 Spinnaker Dr.	Cambridge	Marta	Hernandez	USA
18	3758 North Pendale Stre...	White Plains	Steve	Frick	USA
19	4575 Hillside Dr.	New Bedford	Wing	Huang	USA
20	7734 Strong St.	San Francisco	Julie	Brown	USA
21	7476 Moss Rd.	Newark	William	Brown	USA
22	782 First Street	Philadelphia	Francisca	Cervantes	USA
23	6047 Douglas Av.	Los Angeles	Rian	Chandler	USA

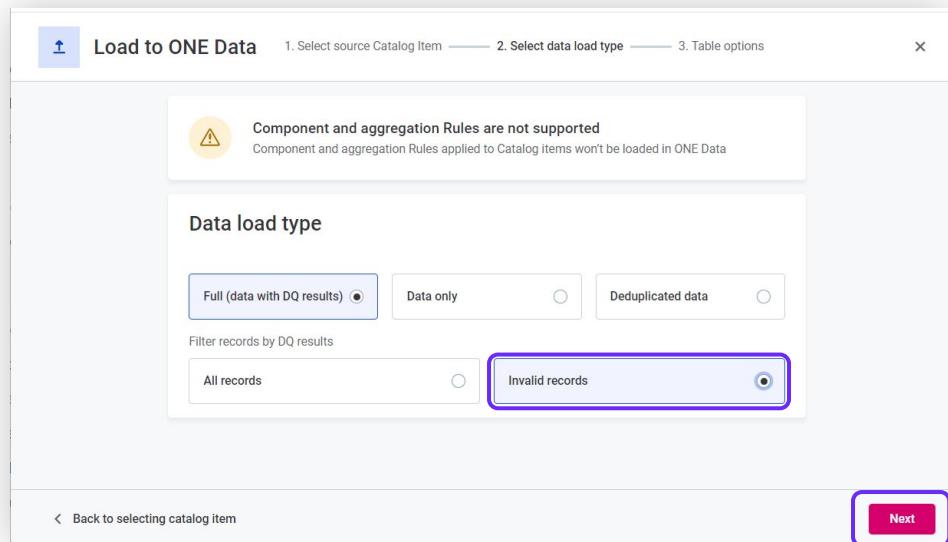
ataccama

# Data Remediation with ONE Data (1/4)

## Step 1: Import invalid records

With ONE Data, you can **import all invalid records** from the Data Catalog or the Data Observability module to ONE Data, where you can then filter results by DQ, edit records, and make other necessary changes before exporting the data back to the data source.

- a. In the three dots menu, select **Load to ONE Data** and then as **Data load type** choose **Full (data with DQ results)** and **Invalid records**.



# Data Remediation with ONE Data (2/4)

## Step 2: Fix invalid records

- a. Once the import is complete, find the newly imported table within ONE Data.
- b. Apply a DQ Filter and examine the filtered records to identify issues that can be addressed immediately
- c. Review the DQ Rules for the attributes and identify invalid records, whether they are null or contain incorrect values.
- d. Fix the incorrect values(you can use **Bulk Edit** option); the changes made are automatically saved.
  - When the DQ warning is eliminated, the list of filtered records is promptly updated.
  - To remediate the null records, you might need additional data collection or verification.

The screenshot shows the Ataccama ONE Data Tables interface for the table 'xx\_customers\_invalid\_records'. The table has columns: reenum, creditlimit, src\_sys\_id, and email. Rows 1 through 6 show various email addresses, some of which are flagged as invalid (indicated by a red exclamation mark icon). A purple box highlights the 'email' column. Below the table, a 'Filter records' dialog is open. The 'WHERE' clause is set to 'All DQ rules'. A sub-dialog for 'Filter DQ rules' is also open, with a purple box highlighting the 'Filter by DQ score' checkbox. The dialog lists 'DQ dimensions' (Select All, Validity, Completeness) and 'DQ rules' (validation E-mail, String Completeness). A purple arrow points from the text 'DQ results persist until recalculated.' to the 'Filter by DQ score' checkbox. The bottom right corner of the slide features the Ataccama logo.

DQ results persist until recalculated.

# Data Remediation with ONE Data (3/4)

## Step 3: Examine rule implementation

- Identify potential issues
- Examine flagged records
- Explore rule implementation
- Decide on corrective action
- Removal of business term

The screenshot shows the Ataccama ONE Data platform interface. On the left, a data table titled "xx\_customers\_invalid\_records" is displayed, filtered by "contactlastname". A context menu is open over a row, listing validation rules: "validation Surname", "String Completeness", and "accuracy Surname". To the right, a "Surname" data quality card provides an overall score of 84% with details for Validity (84%), Completeness (100%), and Accuracy (93%). Below the card, a "Schema" section shows 366 records, 9 applied rules, and 3 catalog items.

Customer Name	Contact Last Name	Contact First Name	Phone
Mini Wheels Co.	Murphy		(1) 356-5555
Porto Imports Co.	de Castro		20.16.1555
Daedalus Design Imports	Ramé		(1) 42.34.2555
La Corne D'abondance, C...	Bertrand		(171) 555-2282
UK Collectables, Ltd.	Devon		(514) 555-8954
Quibec Home Shopping N...	Frennizière		+33 1 46 62 7555
Lyon Souvenirs	Da Silva		(684) 555-4555
Royal Canadian Collecta...	Lincoln		02 9938 8555
Anna's Decorations, Ltd	O'Hara		+61 2 9495 8555
Souveniers And Things C...	Huxley		+31 28 491 9555
Schuylar Imports	Schuyler		(02) 5554 67
Petit Auto	Dewey		
Corporate Gift Ideas Co...	Wain		

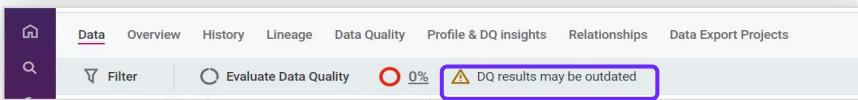
# Data Remediation with ONE Data (4/4)

## Step 4: Update DQ results

- a. Once the rule implementation is examined, publish the changes, and run DQ evaluation again.
- b. Check remaining invalid values and update them.
- c. After the update, run DQ evaluation again and ensure the attribute no longer has any invalid records.
- If a value in the table is updated or DQ results haven't been refreshed, the application displays a message '**DQ results may be outdated.**'
  - We recommend periodically running DQ evaluation to ensure you stay on top of any DQ issues.

## Step 5: Export results from ONE Data

- a. Export data to origin source using the Export feature.
- b. Open the table from **Knowledge Catalog** or **ONE Data**.
- c. In the three dots menu, select **Export > Export to database**.
- d. Choose attributes for export.
- e. Specify the target table for record export.
- f. Decide whether to overwrite the existing table or create a new one.



The Ataccama ONE Data Tables interface displays the 'xx\_customers\_invalid\_records' table. The 'Evaluate Data Quality' button shows 0% completion. A purple box highlights the 'Export' button in the top right corner of the table's detail view.

# Validate Data using ONE Data Tables

- In addition to using DQ rules to evaluate the quality of your ONE Data tables, you can also set DQ standards for other data within the platform.

**Compared to lookup items, an alternative approach for managing reference data in ONE, ONE Data tables provide several key advantages that align with the evolving needs of your business:**

- Data can be directly modified within ONE Data:**  
You can add new attributes, edit incorrect values, or remove duplicate data without leaving the platform.
- Data is updated dynamically:** If a rule uses a lookup item, a new version of the lookup must be uploaded manually after any update. With ONE Data tables, the reference data in your rule is updated in real time, as soon as a change is made.

- To incorporate ONE Data tables into your DQ and detection rules, use the condition "**is from catalog item**" (or "**is not from catalog item**") in the rule logic.

## How a ONE Data table can be used in rules?

- Create reference data**
- Create business term**
- Create rule**
- Enable DQ evaluation on term**
- Add term and run DQ evaluation**

# Topic Highlights

- ONE Data allows you to build trustworthy and governed data products by creating, configuring, and maintaining reference data.
- You can fix data errors directly in ONE Data and leverage its features for data processing, DQ evaluation, collaboration, and reporting.
- Unlike catalog items, ONE Data tables represent actual data that can be modified directly.
- ONE Data tables can be created by starting from an empty table, importing from catalog items, or uploading a CSV file.
- Data can be imported from catalog items, with options for full data, data only, or deduplicated data, and you can create or overwrite tables.
- Deduplication of datasets allows you to manage reference data and create reference tables based on selected attributes.
- Once a ONE Data table is created, you can modify its structure or contents, check data quality, resolve issues, and collaborate with team members.
- Full Profiling & DQ Evaluation can be used to verify data validity and is accessible through the Knowledge Catalog or Profile & DQ Insights tab.
- Data remediation involves importing invalid records, fixing issues with DQ filters, examining rule implementation, updating DQ results, and exporting corrected data.
- To use ONE Data tables in DQ rules, apply conditions like "is from catalog item" in the rule logic.
- ONE Data tables can be utilized to create reference data, business terms, and rules, and to enable DQ evaluations on terms for effective data governance.

# Topic Highlights

- **ONE Data** allows you to build trustworthy and governed data products by creating, configuring, and maintaining reference data.
- **ONE Data tables** represent **actual data** that can be modified directly, allowing you to fix data errors and issues directly.
- **Tables** can be created by; creating empty tables, importing from catalog items, or uploading CSV files.
- ONE Data tables can be utilized to create reference data, business terms, and rules, and to enable DQ evaluations on terms for effective data governance.
- Full Profiling & DQ Evaluation is available to verify data validity and quality.
- **Data remediation** may involve importing invalid records, fixing issues with DQ filters, examining rule implementation, updating DQ results, and **exporting the corrected data**.

Name	Terms	Anomalies	Overall Quality	# Attributes	# Records	Origin	Location
xy.US_Customers	-	-	-	10	one-data		
Turkey city	-	-	-	2	167	one-data	
Saudi Arabia city	-	-	-	2	87	one-data	
Airports	International Airport Code	-	-	2	1,942	one-data	
HTTP Request	-	-	-	2	9	one-data	
R&D	First name Surname E-mail +2	23%	-	7	1,000	one-data	
HTTP Response Code	-	-	-	2	48	one-data	
Italy city	ITAI CITY	>99%	-	2	144	one-data	
Conference Attendees	First name Surname E-mail +4	52%	-	11	199	one-data	
Indonesia city	-	-	-	2	230	one-data	
Country	Country ISO-2 Country Code +1	66%	-	5	240	one-data	

# Memory Refresher #6 ONE Data



# Data Transformation Plans



# Data Transformation Plans

You can **Transform data in ONE Web** using Data Transformation Plans:

- Plans serve as **visual representations** of data processing job definitions.
- **Plans** can be **configured** to process input data according to the requirements.
- Data flows and output formats should be defined within the plan.
- Jobs are **executed** by running the configured plan.
- Each step in the plan **represents basic algorithms**.
- Data is processed **record by record** during the execution of the plan.

The screenshot shows the Ataccama Data Quality interface. On the left, a sidebar menu lists various components: Rules, Detection Rules, DQ Evaluation Rules, DQ Dimensions, Components (which is currently selected and highlighted with a purple border), Transformation Plans (also highlighted with a purple border), DQ Firewalls, Monitoring Projects, Reconciliation Projects, and Lookup Items. The main panel is titled "Transformation Plans". It features a search bar at the top. Below the search bar is a table with columns: Name, Description, Status, Input type, Output type, and Step. A single row is visible in the table, showing "xy\_customers\_transformation\_plan" with the description "create a plan that inputs the catalog item Cust...". The status is "No issues", Input type is "Catalog Item Input", Output type is "ONE Data writer", and Step is "5".

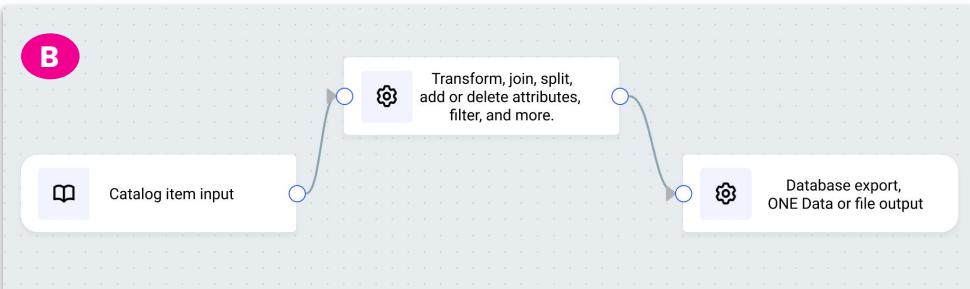
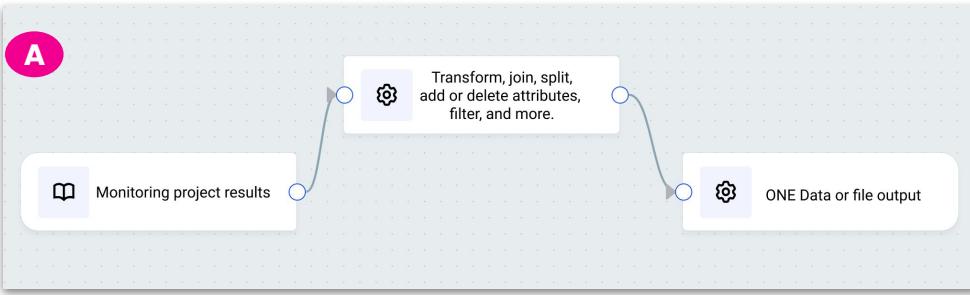
# Data Transformation Plans – Overview

**Transformation plans facilitates two key use cases:**

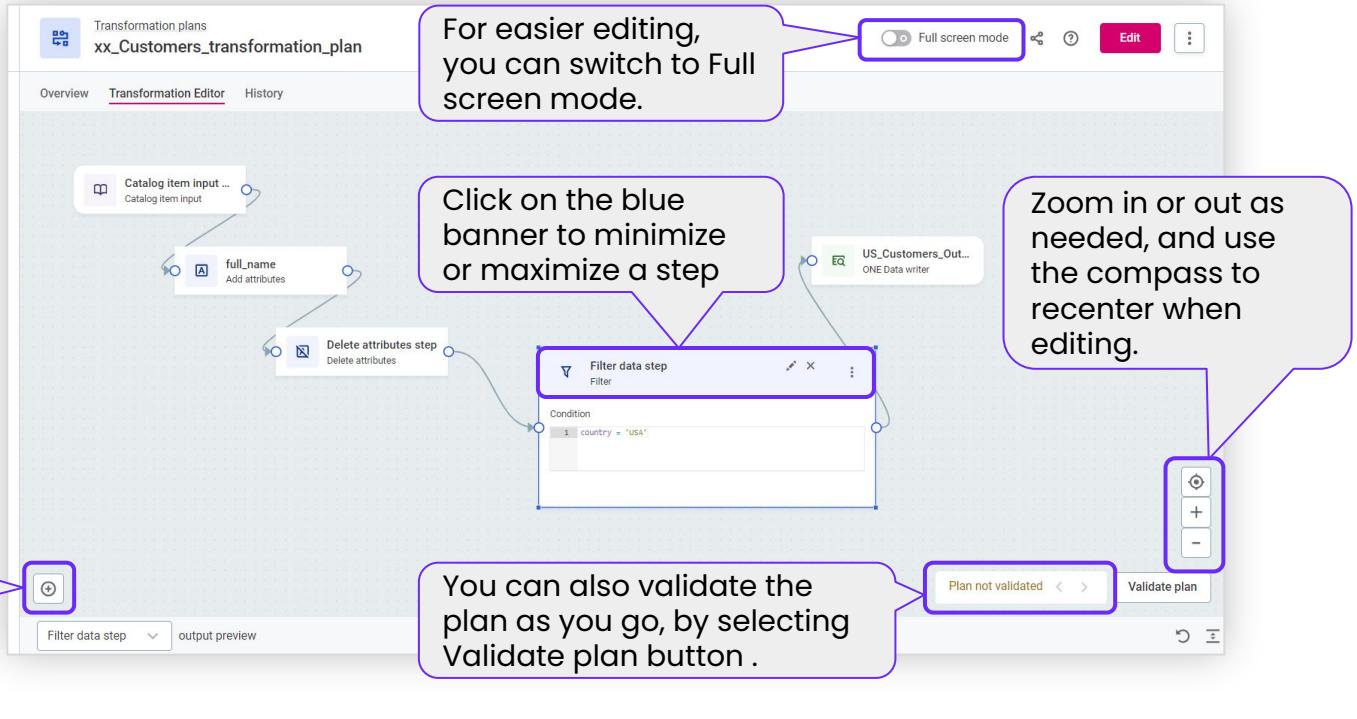
- A. Transformation of data quality **monitoring project** results. In this way, data transformation plans supersede **post processing** plans.
- B. **Standalone** data transformation of **catalog items**.

**Through the plans, you can:**

- Configure data input (catalog item or monitoring project results).
- Add various steps, including **filter**, **split**, **join**, **transform**, **add**, or **delete attributes**.
- Utilize the Embedded transformation plan step to embed plans within plans.
- Do post-transformation by **exporting** data to **ONE Data**, files to **ONE Object Storage**, or using it in **another plan**.
- Preview data and validate expressions in real-time during your work.



# Data Transformation Plans: Transformation Editor

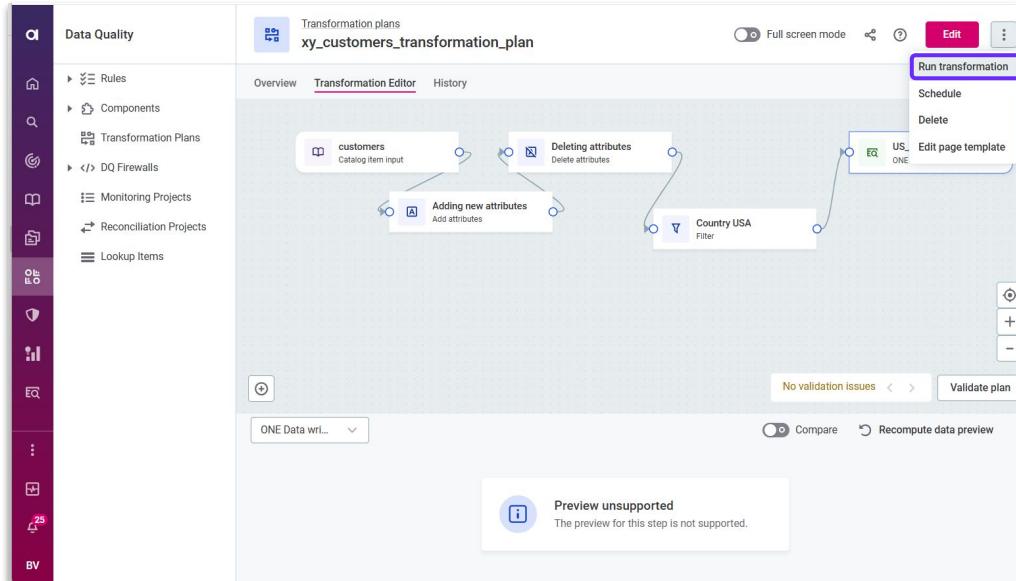


# How is data transformed?

To transform a catalog item using Transformation plans:

1. Select your input data.
2. Insert steps which can, for example, filter, transform or join the data.
3. Draw connections between steps to control the data flow.
4. Select your **data output**.
5. Run the transformation to trigger the flow of data through the plan.

- steps are similar to those in ONE Desktop, which leverage ONE These Expressions.
- Each step represents a data processing job.



# Available steps

- **Catalog Item input:** Reads data from the selected catalog item.
- **Input:** A generic input step which can be used to embed the plan into another.
- **ONE Data reader:** Reads data from the selected ONE Data catalog item.
- **File Output:** Writes data into a text file in ONE object storage.
- **ONE Data writer:** Writes data into new or existing ONE Data catalog items.
- **Output:** A generic output step which can be used to embed the plan into another.
- **Data Remediation output:** The data remediation output step should be used in data remediation plans.
- **Add attributes :** Adds new attributes.
- **Condition:** Splits the data flow into two streams based on the ONE expression condition.

# Available steps

- **Delete attributes**: Deletes selected attributes.
- **Filter**: Filters out all records which don't pass the ONE expression condition.
- **Embedded component**: Inserts a component from ONE Desktop.
- **Join**: Joins two separate data flows into a single data flow (vertical merge).
- **Embedded transformation plan**: Inserts an existing transformation plan created in ONE.
- **Split**: Splits the data flow into three streams based on the ONE expression condition.
- **Transform data**: Transforms attribute values using ONE expression language.
- **Union streams**: Merges input formats of two defined inputs into a single output format (horizontal merge). Data is read from the individual inputs (in\_a or in\_b) and written to the output. Therefore, in the output there are merged-format data records, where each row contains data read from either in\_a or in\_b input.
- **Monitoring project post processing input**: Inputs the data and DQ evaluation results from a monitoring project.

# Topic Highlights

- Use Data Transformation Plans to visualize and execute data processing jobs.
- Embed plans within other plans by using the Embedded transformation plan step.
- Export transformed data to ONE Data, save it as files, or use it in another plan.
- Preview data and validate expressions during the development process.
- Build plans with steps such as filtering, transforming, or joining data using ONE Expressions.

The screenshot shows the Ataccama Data Quality interface. On the left, a sidebar menu lists various data quality components: Rules (Detection Rules, DQ Evaluation Rules), DQ Dimensions, Components, Transformation Plans (which is selected and highlighted in blue), DQ Firewalls, Monitoring Projects, Reconciliation Projects, and Lookup Items. The main panel is titled "Transformation Plans" and displays a table of existing plans. The table has columns for Name, Description, Status, Input type, Output type, and Step. One row is visible, showing "xy\_customers\_transformation\_plan" with a description of "create a plan that inputs the catalog item Cust...", status "No issues", input type "Catalog item input", output type "ONE Data writer", and step "5". A "Create" button is located at the top right of the main panel.

Name	Description	Status	Input type	Output type	Step
xy_customers_transformation_plan	create a plan that inputs the catalog item Cust...	No issues	Catalog item input	ONE Data writer	5

# Memory Refresher #7

## Data Transformation Plans



# Lab Exercise #4

## Data Transformation Plans



# DQ Foundation

v15.4.x



ataccama