

Catalog & Glossary

v15.4.x



ataccama

Agenda

Catalog & Glossary

- ONE Web Application
- Knowledge Catalog
 - Sources
 - Profiling
 - Catalog Items
 - Reports
- Business Glossary
- Data Observability



Material GOALS

Who is this for:

- › This session is meant as an initial Gen 2 session for power users.

Prerequisites/what should be known:

- › No prerequisites are required.

Target group:

- › Power Users or Developers

Training Duration:

- › This session is estimated for ~3 hours duration. However, with many features and aspects being introduced, the duration might be prolonged if the participants will request additional time or initiate a discussion over these topics.

Applicable to Gen2 version:

- › v15.1.x

ONE Web Application



ataccama

Landing Page

Home is the default landing page - the initial starting point for all users, where you can:

- Manage and **track your tasks**, assign tasks to others, set due dates, and monitor progress.
- Chat with documentations with the help of Generative AI.

Access to landing pages is managed through a **View page access** level.

Viewers with roles that include this access level can view landing pages and their content but cannot make any changes.

The screenshot shows the Ataccama Home landing page. At the top, there's a navigation bar with a logo, user name 'Hi, Sanaz', a search bar, and a help icon. Below the bar, a purple sidebar on the left contains icons for Home, Catalog, Glossary, Data Quality, Observability, and other platform features. The main content area has three columns: 1) A central column with a 'Getting started with Ataccama' section and links to Catalog, Data Quality, and Observability guides. 2) A right column titled 'Explore the platform with our guides:' with links to documentation and community resources. 3) A bottom section titled 'Tasks' showing a list of 'Review Request' items. To the right of the tasks is a 'Chat with documentation' sidebar with AI-generated answers to questions like 'How to upload a file?' and 'What is the difference between rules on terms and rules mapped to Attributes?'. The overall interface is clean and modern, designed for data management and analysis.

ONE Web Application–Interface

Application group

- Home
- Global Search
- Data Observability
- Knowledge Catalog
- Business Glossary
- Data Quality
- Data Protection
- Data Stories
- ONE Data



Administration group

- Audit
- Global Settings
- Tasks & Workflows
- Processing Center
- Notification Center
- My Account



The screenshot displays the Ataccama web application's main interface. At the top, a navigation bar includes a user greeting "Hi, Sanaz", a search bar with placeholder text "Type at least 3 characters to find Catalog items, Terms and more", and a help icon. Below the bar, a purple header section titled "Viewing panel" contains links for "Getting started with Ataccama", "Explore the platform with our guides", and "Learn about the platform at your own pace". The main content area features a "Tasks" section with filters for "I'm author" and "I'm assignee", and a "Language" dropdown set to English. To the right, there's a "Chat with documentation" section with a list of questions and answers. The bottom right corner features the Ataccama logo.

ONE Home Page

1. Default widgets:

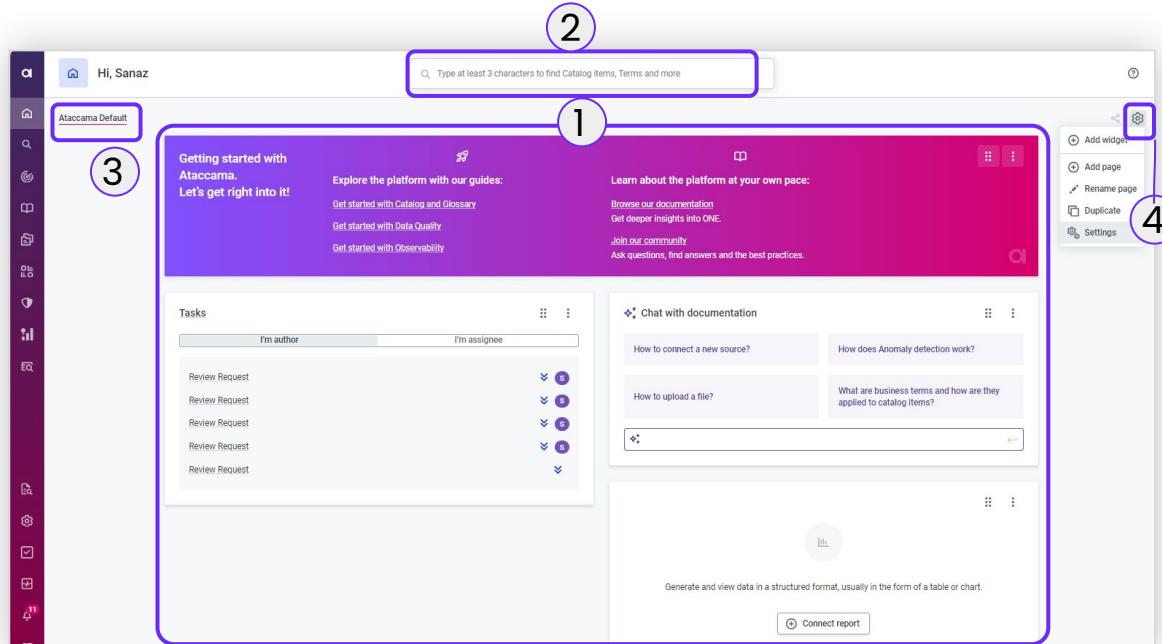
- Getting started with Ataccama
- Tasks
- Chat with Documentation
- Connect Report

2. **Global Search:** a full-text search which is enabled **on the whole application** and provides ability to locate the area for further search if it is needed

3. **Tab with Default page and secondary landing pages, if available**

4. **Settings menu**

- Add widget
- Add page
- Rename page
- Duplicate
- Settings



Topic Highlights

- One Web Application user interface consist of the **Navigation Panel** on the left and the **Viewing panel** on the right.
- The **Viewing panel** is the primary working space of the ONE Webapp.
- After selecting an item in the navigational panel or Hierarchy Side menu, the viewing panel will be the space where you can view details, edit and publish changes.
- By navigating through **Hierarchy Side menu** different items, **Viewing panel** options change accordingly.
- The navigation of the left-side panel and landing page can be changed to better suit your needs.

Memory Refresher #1 ONE Web Application



Match terms with their correct icon:

Business Glossary

Knowledge Catalog

Global Search

Data Quality

1.



2.



3.



4.



Global Search

Knowledge Catalog

Business Glossary

Data Quality

Knowledge Catalog

Data Catalog



ataccama

Data Catalog - Overview

Knowledge Catalog contains all the necessary tools for data management, including importing data from data sources, and viewing, discovering, and profiling data.

- **Data Catalog** is a part of Knowledge Catalog.
- Data catalog consists of Catalog Items, Sources, Term suggestions and Lineage import.
- **In Data Catalog you can:**
 - configure connections
 - manage automatically built metadata.
 - view and profile data entities.

You can view a summary of the latest lineage import and mapping on the **Lineage Import** tab.

The screenshot shows the Ataccama Knowledge Catalog interface. On the left is a sidebar with navigation links: Reports, Anomaly Overview, Data export projects, Master Data, Reference Data, Not Monitored, Sources, Term Suggestions, and Lineage Import. The main area is titled 'Knowledge Catalog' and 'Data Catalog'. It features a search bar and filters for Terms, Profile Date, Anomalies, and Stewardship. The 'Catalog Items' section displays a table with columns: Name, Description, Terms, Anomalies, Overall Quality, # Attributes, # Records, and Origin. The table lists items like 'products' and 'KS_Orders_transformed'. The 'Sources' section shows a table with similar columns for 'products', 'orders', and 'KS_Orders'. The 'Term Suggestions' section is shown as a table with columns: Name, Description, Status, and Last Seen. A callout box points to the 'Catalog Items' section with the text: 'Data assets along with their metadata can be found in Catalog Items section.' Another callout box points to the 'Sources' section with the text: 'Data sources can be defined and accessed in Sources section.' A third callout box points to the 'Term Suggestions' section with the text: 'View all information about pending term suggestions and approve or reject them directly from the Term Suggestions tab in ONE.'

Data Catalog - Sources

By using the Sources section under the Data Catalog section, a user can:

- Create and manage Data Sources.
- View information on the created sources.
- Run Data Discovery and Profiling for each selected source.
- Run the Documentation Flow for each selected source.

After a data source has been added, data and metadata from this source can be imported and subsequently cataloged and profiled, as well as monitored for data quality.

The screenshot shows the Ataccama Data Catalog interface. On the left, there is a sidebar with various icons and sections: Knowledge Catalog, Data Catalog (selected), Catalog Items, Reports, Anomaly Overview, Data Export Projects, Master Data, Reference Data, Not Monitored, Sources (highlighted with a purple box), Term Suggestions, Lineage Assets, and Lineage Import. The main area is titled 'Sources' and contains a search bar with placeholder text 'Type here to search full-text for Sources'. Below the search bar are several filter buttons: Source name, Connection type, Catalog items count, DO state, DO monitoring, DO issue count, Term, and Stewardship. A table below these filters lists data sources. The table has columns for Name, Description, and Stewardship. Each row includes a checkbox, the source name, a description, and a stewardship role indicator. The rows listed are: Analytics Snowflake DWH (Snowflake DWH Analytics, Data Office), BI DWH (Single source of Sales and Marketing BI reports, Procurement), Data Stories (Connection to Ataccama ONE Data Stories, Data Office), Databricks (Databricks Data Source, Data Office), Finance DO (Snowflake Data Observability content (Finance), Finance), Global CRM (Global CRM data, Data Office), and hr_data (Snowflake HR Data Content, HR).

Name	Description	Stewardship
Analytics Snowflake DWH	Snowflake DWH Analytics	Data Office
BI DWH	Single source of Sales and Marketing BI reports	Procurement
Data Stories	Connection to Ataccama ONE Data Stories	Data Office
Databricks	Databricks Data Source	Data Office
Finance DO	Snowflake Data Observability content (Finance)	Finance
Global CRM	Global CRM data	Data Office
hr_data	Snowflake HR Data Content	HR

Data Catalog - Catalog Items

In Data Catalog, a user can:

- View Catalog Items and their metadata
- Check profile and DQ insights of each item
- Check and evaluate data quality per item
- Resolve Anomalies
- Add rule directly to attributes of an item

Name	Description	Terms	Anomalies	Overall Quality	# Attributes	# Records	Origin	Location	Stewardship
products	xx_Scale	-	-	-	9	110	training	xx_training > ... > public	
KS_Orders_transformed	-	-	-	-	9	326	manually created	ks_training > Transformed_VCI	
customers	Country First name Surname +4	23%	23%	23%	15	122	postgres	ks_training > ... > public	
products	xx_Product Line xx_Scale	-	-	-	9	110	postgres	ks_training > ... > public	
orders	-	-	-	-	8	326	postgres	ks_training > ... > public	
KS_Orders_Cancelled	-	-	-	-	2	-	postgres	ks_training > SQLItems	
ks_customers_phone	-	-	-	-	3	-	postgres	ks_training > ... > public	
ks_product_info	-	-	-	-	5	-	one-data	one-data > DMM > default	Data 0

Note that in order to generate Catalog Items, Data Sources should be created and profiled.

Generate catalog item descriptions using generative AI where catalog items descriptions are missing or could use improving, save time .

Data Catalog - Catalog Items

The information you can see on the Data Catalog tab depends on your global role and access level, and whether access was shared with you.

Reports: allows to manage in ONE all BI reports on catalog items that are build in BI tool.

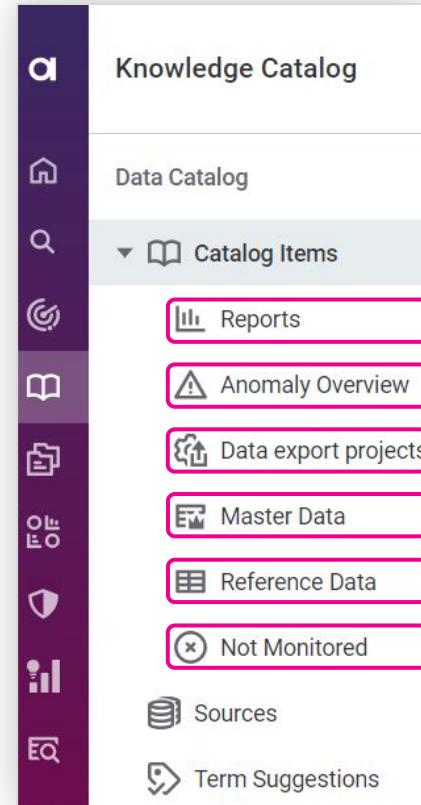
Anomaly Overview: a list of all catalog items with anomalies.

Data Export Projects: all data export projects created in ONE.

Master Data: a list of all catalog items contained within MDM data sources.

Reference Data: a list of all catalog items contained within RDM data sources.

Not Monitored: a list of all catalog items that are not currently included in any monitoring projects.



Topic Highlights

- The **Data Catalog** is built automatically via the Data Cataloging user flow, but it is also possible to manage the metadata manually.
- The Data Catalog can store **schemas, tables, metadata models, flat files, or custom data sets**.
- **On the Catalog Items tab**, users can create and manage catalog items, check profiles, DQ insights, evaluate data quality, resolve anomalies, and add rules directly to attributes.
- **On the Sources tab**, users can create and manage data sources, view connected sources, run discovery and profiling, and run documentation flow.

Memory Refresher #2 Data Catalog



Data Catalog Sources



Data Source

- The source Overview page displays the list of profiled Catalog Items, tables, etc.
- To Edit a data source, click on the 3 dots menu at the top right corner of each source:**
 - To delete the data source, click on Instant delete.
 - To schedule documentation, select Schedule option and configuring it.
- The History tab provides a list of the previous versions.

The screenshot shows the 'Sources' section of the Ataccama interface. A specific data source named 'ks_training' is selected. The main panel displays various metrics and configuration options. A callout bubble points to the 'Folders' section, which contains the value '1'. Another callout bubble points to the 'Catalog Items' section, which contains the value '4'. A third callout bubble points to the 'Workspace' section, which contains the value 'SQL_Items'. A context menu is open on the right side of the screen, listing options like 'Create SQL Catalog Item', 'Schedule', 'Edit', 'Request access', 'Instant delete', 'DQ Settings', and 'Edit page template'. A note in the bottom right corner suggests assigning stewardship to the asset.

Catalog Items	Tables	DSL query Catalog items	Database locations
4	3	1	1

Folders
1

Catalog Items include tables, views SQL and Virtual Catalog Items

Sources
ks_training

Description

Catalog Items

Tables

DSL query Catalog items

Database locations

Folders

1

Summary

Deployment

+ Add

AI Term suggestions enabled

Locations

Add Location

postgres

Workspace

Add Folder

SQL_Items

Stewardship

Edit

Create SQL Catalog Item

Schedule

Edit

Request access

Instant delete

DQ Settings

Edit page template

Who is responsible for this asset?
Assigning stewardship is recommended.

Term occurrences

Folders are created to store SQL & Virtual Catalog Items

How to create a Data Source?

(1 / 3)

1. Go to **Data Catalog > Sources**.
2. Click the **Create** button and specify a **Name** and Description (optional).

The image shows two screenshots of the Ataccama Data Catalog interface. On the left, the 'Sources' list is displayed with a search bar and various filters. A purple arrow points from the 'Create' button in the top right of this screen to the 'Create Source' dialog on the right. The 'Create Source' dialog is titled 'Create Source' and contains sections for 'General information' and 'Deployment'. The 'Name' field is highlighted with a purple border and contains the value 'xx_training'. A large purple arrow points from the 'Create' button in the main list to this dialog.

Sources

Type here to search full-text for Sources

Source name Connection type

Catalog items count DO state

DO monitoring DO issue count Term

Stewardship

Standard view Hidden columns

Name	Description
by_training	
ks_training	
one-data	ONE Data source which allows the user to work with d...

Create

Create Source

General information

A business perspective of the data source. It can refer to business purpose, department, location OR to a particular user group. It should capture the organization of data sources in your company.

Name

xx_training

Description

Deployment + Add

Stewardship

Save and publish

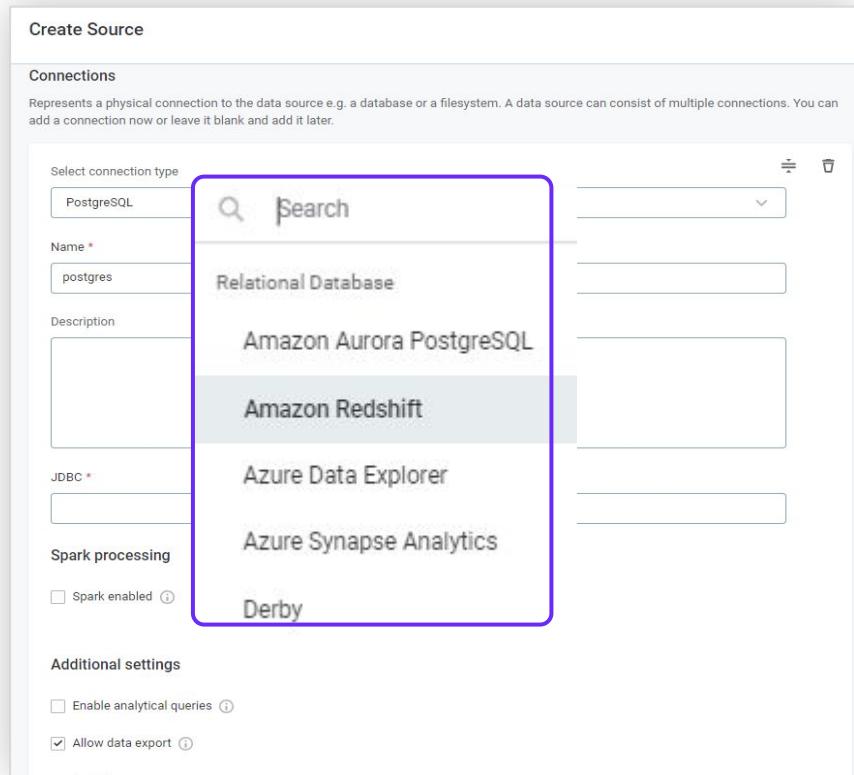
How to create a Data Source? (2 / 3)

3. Select **connection type** to the data source e.g., a database or a filesystem.

Multiple connections can be specified (grouped together) for one data source and they can be of **different types** (Amazon S3, Oracle, etc.).

It is possible to **disable Data Export** for each data source **connection**. Disabling Data Export on the data source does not restrict exporting to ONE Data, only to other databases.

Analytical queries must be enabled in order to **Create Visualization** in **Data Stories** from catalog items.



How to create a Data Source? (3 /3)

- Depending on the type of connection fill in available fields.
- **Save and Publish.**

For every connection multiple credentials can be defined.

Credentials can be **Set as default**. In that case these credentials are used for profiling or DQ evaluation.

Before saving the connection, validity of the connection can be tested using the **Test Connection** button.

The screenshot shows the 'Credentials' configuration screen. It includes fields for 'Credential type' (set to 'Username and password'), 'Name' (empty), 'Description' (empty), 'Username' (set to 'admin'), 'Password' (set to '.....'), a 'Test Connection' button, and a 'Set as default' toggle switch.

Credentials

Credential type

Username and password

Name

Description

Username *

admin

Password *

.....

Test Connection

Set as default

Supported Data Sources in the ONE

Supported databases:

- Amazon Aurora PostgreSQL
- Amazon Redshift
- Azure Data Explorer
- Azure Synapse Analytics
- Derby
- MariaDB
- MS SQL(jTDS)
- MSSQL Server

Ataccama Platform Integration:

- MDM
- ONE Data

- MySQL Server with no locks.
- Oracle
- PostgreSQL
- Snowflake
- Sybase
- Teradata
- ... Any JDBC compliant DB driver

Reporting tools:

- Data Stories
- Power BI
- Power BI Report Server
- Tableau

Cloud Services

- Amazon S3
- Azure Data Lake Storage Gen2
- Dropbox
- Filesystem
- Google Cloud Storage
- Google Drive
- ONEDrive Storage

Other Sources:

- Salesforce
- SAP RFC

Topic Highlights

- **Multiple connections** can be specified for one data source.
- Connections can be of **different types** (Amazon S3, Oracle, etc.)
- Set credentials as default for profiling or DQ evaluation.
- For every connection **multiple credentials** can be defined.
- Validity of the connection can be tested using the **Test Connection** button.
- Delete a data source instantly by clicking on **Instant delete**.

Memory Refresher #3 Sources



Data Discovery & Profiling



ataccama

Profiling Scenarios



- There is a new data source in place.



- Deeper insight into current data sources is needed before changes are proposed.



- Root-cause / pattern analysis of data issues is required.

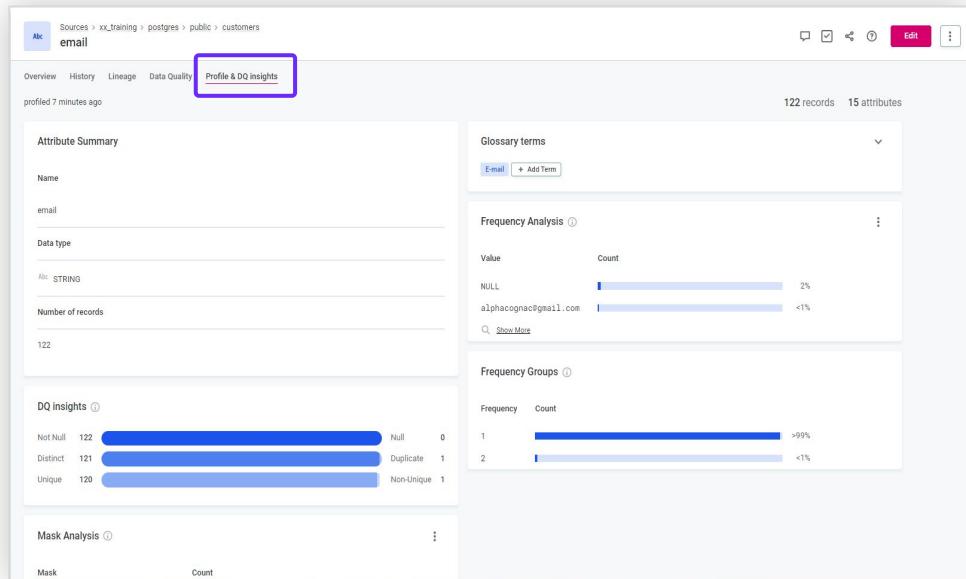


- Single-purpose evaluation of particular data is required.

Profiling

Profiling is:

- Process of **creating catalog items** in the ONE.
- Tool for **analyzing**, understanding, and **discovering** patterns in data.
- The **necessary step** to use other features of the application (for instance, creating lookups).



Manual Profiling

To profile a source, go to the **Source** **>Connection> connection browser** and select the desired assets.

Manual profiling options:

- **Import to catalog** – metadata import only; attributes and their data structure will be available whereas data quality, profiling and DQ insights will not.
- **Discover** – profile a data sample; in this option all profiling insights will be provided based on limited number of records.
- **Profile** – full profiling.

The screenshot shows the Ataccama Connection Browser interface for a 'postgres' connection. The 'public' schema is selected. A table named 'customers' is selected, indicated by a checked checkbox in the left column. A context menu is open over the 'customers' row, with the 'Profile' option highlighted and surrounded by a red box. The menu also includes 'Import to catalog' and 'Discover' options. At the bottom of the browser, there are buttons for 'Clear selection' (with '1 item selected'), 'Profile' (highlighted with a red box), and a dropdown menu.

Name	Type
addresses	TABLE
addresses_2	TABLE
banking_information	TABLE
banking_information_certification	TABLE
city_an	TABLE
contact_kartik	TABLE
customer_enriched	TABLE
customer_information	TABLE
customer_information_certification	TABLE
<input checked="" type="checkbox"/> customers	TABLE

Automatic Data Discovery & Profiling

To profile a source, go to the **Source**, open the **drop down menu** on the top right hand side and select the desired option.

Automatic profiling options for sources:

- **Import** - imports the metadata of all entities in the data source without accessing the data.
- **Discover** - consists of profiling a small sample of records from the data set.
- **Document** - imports the metadata of entities, efficiently profile and validate the DQ.

The screenshot shows the Ataccama Knowledge Catalog interface. On the left is a sidebar with various icons and a 'Data Catalog' section containing 'Catalog Items', 'Reports', 'Anomaly Overview', 'Data export projects...', 'Master Data', 'Reference Data', 'Not Monitored', 'Sources', 'Term Suggestions', and 'Lineage Import'. The main area has a header 'Sources xx_training'. Below it is a table with one row: 'Catalog Items' (1), 'Tables' (1), and 'Database locations' (1). A 'Summary' section follows, then a 'Deployment' section with a '+ Add' button. Under 'Locations', there is a 'postgres' entry with an 'Add Location' button. The 'Workspace' section shows a folder icon with the message 'No folders have been added yet.' and an 'Add Folder' button. On the right, there is a 'Document' button highlighted with a red box. A tooltip for 'Document' says: 'Import metadata, run quick data discovery and efficiently profile and validate data quality of all relevant assets.' Other options shown are 'Import' (import all metadata) and 'Discover' (quick data discovery).

Monitoring progress



The **Processing Center** is used to monitor and manage profiling as well as other latest and scheduled jobs.



The **Notification Center** provides real-time updates on the running jobs and other changes noticed in different parts of the application.

The diagram illustrates the monitoring process. On the left, a screenshot of the 'Latest jobs' section of the Processing Center shows a list of completed jobs. A callout box with a purple border and white text says: 'Click on each process line to access details'. An arrow points from this screen to the right, leading to a detailed view of a specific job. The right-hand screen shows the 'Base jobs' page for a job named 'customers (Sample profiling)'. The 'General information' section includes fields like Type (PROFILING), Execution type (MANUAL), and various timestamps. A large callout box with a purple border and white text labeled 'Details of the job' covers the right side of the detailed view, indicating where to click for more information.

Latest jobs Scheduled jobs 5

3 days ago DQ Evaluation of Catalog Item 'products' Finished

3 days ago Anomaly detection of products Finished

3 days ago Profiling of products (Full Profile & DQ Evaluation) Finished

3 days ago Metadata part of training_materials Open Processing Center

Click on each process line to access details

Base jobs customers (Sample profiling)

Overview History Access

General information

Type PROFILING

Execution type MANUAL

Created at June 15, 2022 1:21:29 AM

Started at June 15, 2022 1:21:29 AM

Finished at June 15, 2022 1:21:36 AM

Linked entity node path /sources/locations/catalogitems

Linked entity id 15bbce34-0000-7000-0000-0000003019b9

Details of the job

Anomaly Detection

Anomaly detection discovers any **inconsistencies** or **irregularities** in the **data**.

- **Anomaly detection is applied in the following cases:**
 - **After profiling the data:** Anomaly detection is automatically started following each profiling run.
 - **In monitoring projects,** as part of the data quality (DQ) evaluation.
- **Anomaly detection relies on one of the two models:**
 - **Time-independent model:** This model is used **by default**.
 - **Time-dependent model**, based on the time series analysis.
- You can select **time-independent** or **time-dependent anomaly detection** to be used during **profiling**.
- The model is equipped to detect **unexpected values of different kinds** such as unexpected nulls, negatives, positives or zeros and changes from established trends.
- Time-dependent anomaly detection requires a **minimum of 6 runs**.

The screenshot shows the Ataccama Profiling interface. On the left is a sidebar with icons for Global Settings, Application Settings, Profiling, Retention Settings, and Catalog Item retention. The main area is titled '00000000-0000-0000-000000000007 Full Profile & DQ Evaluation'. It contains sections for 'General information' (Name: 'Full Profile & DQ Evaluation', Description: 'Profile the full dataset and DQ evaluation'), 'Partitioned' (checkbox checked), 'DQ eval enabled' (checkbox checked), and 'Detect anomalies' (checkbox checked). A note below states: 'Anomaly detection will help you identify any potential corruptions in your data by flagging attributes with, e.g., unexpected length (string) or null count (numeric). This detection will occur each time you profile your dataset.' At the bottom, there are dropdown menus for 'Anomaly detection sensitivity' (set to 'Medium') and 'Anomaly detection model' (set to 'Time Independent'). The 'Anomaly detection model' dropdown is highlighted with a purple rounded rectangle.

Profiling settings (1/2)

To view the profiling settings, go to Global Settings > Profiling:

- **Partitioned** (Boolean): If true means partitions are supported.
- **DQ eval enabled** (Boolean): If true will mean DQ evaluation is supported.
- **Detect Anomalies** (Boolean): If true will mean anomaly detection is enabled.
- **Anomaly detection sensitivity**: Very high, High, Medium, Low, and Very low.
- **Limit number** of profiled rows (count or percentage).

The screenshot shows the Ataccama Global Settings interface. On the left is a sidebar with icons for Import and Export, Email Templates, MS Teams, Slack, Documentation flow, Graph visualization, Profiling (which is selected and highlighted in grey), Relationship Types, and Search configurations. The main area is titled "Profiling" and contains a search bar, a "Standard view" dropdown, and a "Hidden columns" dropdown. Below these are several configuration rows:

Name	Type
Sample Profile & DQ Evaluation	Sample profiling
Full Profile & DQ Evaluation	Full profiling
Full profiling	Full profiling
Sample profiling	Sample size 1% or at most 10 000 records

The screenshot shows the "General information" tab of a Profiling settings dialog box. It includes fields for Type (set to FULL), Anomaly detection enabled (checked), Partitioned (checked), Sampling limit count (set to "-"), Sampling limit percentage (set to "-"), DQ eval enabled (unchecked), Anomaly detection sensitivity (set to Medium), and Anomaly detection model (set to Time dependent). A purple callout bubble points to the "DQ eval enabled" field with the text: "Parameters are different for **custom_sample** and **custom_full profiling** configurations".

Profiling settings (2/2)

Creating a New Profiling Option:

- Click **Create** in the upper-right corner of the page.
- Fill in **Name** and **Description**.
- Select the Type: **CUSTOM_FULL** or **CUSTOM_SAMPLE**
- If you selected **CUSTOM_FULL**, define: Partitioned, DQ eval enabled, Detect anomalies, Anomaly detection sensitivity, Anomaly detection model.
- If you selected **CUSTOM_SAMPLE**, define: Partitioned, DQ eval enabled, Limit number of profiled rows.
- Click **save** and **publish** the changes.

The screenshot displays two parallel configuration panels for creating a new profiling option. Both panels have a header 'Create Profiling' with a 'Save' button.

Left Panel (CUSTOM_SAMPLE Configuration):

- Type ***: CUSTOM_SAMPLE
- Partitioned**: Unchecked
- DQ eval enabled**: Unchecked
- Limit number of profiled rows**:
 - % or at most: 20%
 - records: 100000

Right Panel (CUSTOM_FULL Configuration):

- Type ***: CUSTOM_FULL
- Partitioned**: Unchecked
- DQ eval enabled**: Unchecked
- Detect anomalies**: Unchecked
- Anomaly detection will help you identify any potential corruptions in your data by flagging attributes with, e.g., unexpected length (string) or null count (numeric). This detection will occur each time you profile your dataset.**
- Anomaly detection sensitivity ***: Medium
- Anomaly detection model ***: Time independent

Topic Highlights

- Tool for analyzing, understanding, and discovering patterns in data.
- **Manual Profiling options:**
 - Import to Catalog
 - Discover
 - Profile
- **Automatic Data Discovery & Profiling options:**
 - Import
 - Discover
 - Document

Memory Refresher #4 Data Discovery & Profiling



Data Catalog Catalog Items



ataccama

Catalog Items

- Catalog Items are generally **created** by **profiling** Data Sources.
- Catalog Items contain **data** assets and their **metadata**.

Catalog Items include information such as:

- number of records
- assigned terms
- profiling results and DQ results
- data lineage
- relationships

The screenshot shows the Ataccama Knowledge Catalog interface. On the left is a sidebar with various navigation options like Data Catalog, Catalog Items, Reports, Anomaly Overview, Data export projects, Master Data, Reference Data, Not Monitored, Sources, Term Suggestions, and Lineage Import. The main area is titled 'Sources > ks.training > postgres > public > customers'. It has tabs for Overview, History, Data, Data Structure, Lineage, Data Quality, Profile & DQ Insights, Relationships, Data Export & Transformations, and Sample Profiling. The Overview tab is selected. It displays a table with columns for Attribute, Type, Description, and Actions. Attributes listed include customernumber, customername, contactfirstname, contactlastname, phone, addressline1, addressline2, city, state, postocode, and email. Below the table is a 'Data Quality' section showing Overall Quality (23%), Validity (23%), Completeness (40%), and Accuracy (27%). A red circle highlights the Overall Quality score. Annotations with arrows point from callout boxes to specific parts of the interface: one arrow points from the word 'Terms' to the 'customername' attribute entry; another arrow points from the word 'Number of records and attributes' to the 'customernumber' attribute entry; and a third arrow points from the text 'Overall Data Quality results and per attribute' to the 'Data Quality' section.

Terms

Overall Data Quality results and per attribute

Number of records and attributes

Catalog Items: Profiling & DQ Insights - Item Level

Profiled attributes, including their data types, assigned terms and any detected anomaly

The screenshot shows the Ataccama Data Catalog interface with the following details:

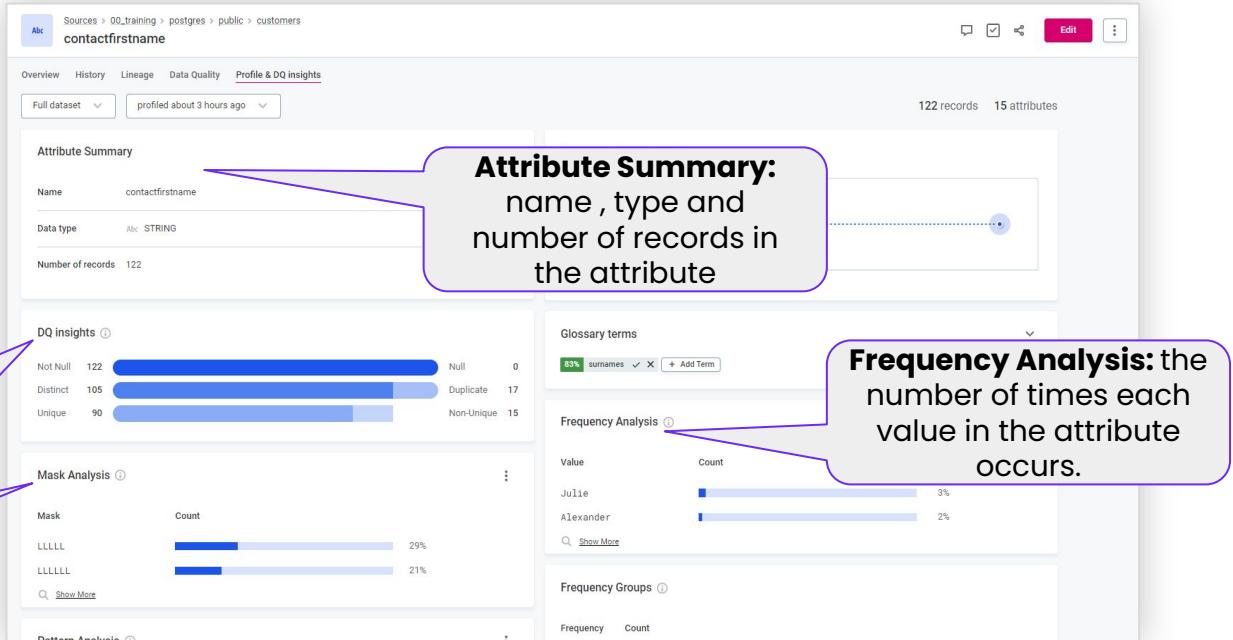
- Top Navigation:** postgres > public
- Tab Selection:** Profile & DQ insights (highlighted)
- Dataset Information:** profiled 4 min ago, 122 records, 15 attributes
- Left Sidebar:** Catalog Items, Sources, Term Suggestions, Lineage Import.
- Central Content:**
 - Profiled Attributes Table:** Shows columns: Name, Anomalies, Glossary Terms. Examples include customernumber (Anomaly icon), customername (Abc), contactlastname (Abc, Surname), contactfirstname (Abc, First name), phone (Abc, [North America] Phone Number), addressline1 (Abc, Street with number), addressline2 (Abc, Street with number), city (Abc), state (Abc, [USA] State).
 - DD Insights Table:** Shows top values and masks for various attributes. For example, for 'customernumber':
 - Top 3 values: NULL (0%), 183 (1%), 112 (1%)
 - Masks: LLLLLL LLLLLLLL, LLLLLL LLLL
 - Other details: Show More +78
 - Other Attribute Examples:** contactfirstname (Young, Brown, Frick), phone (6175558555, 1 (82) 5554 67, 1 (071) 23 67 2555), addressline1 (Jakarta, 1 rue Alsace-Lorraine, 1-6-20 Dojima), addressline2 (Null, Suite 101, 1 Garden Road), city (Madrid, NYC, Auckland), state (Null, CA).
- Bottom Right Text:** Some sample records give insight as to what is in the data source that has been profiled.

Results of the profiling can be found on the Profile & DQ Insights tab.

Catalog Items: Profiling & DQ Insights - Attribute Level (1/2)

Basic Analysis includes:

- number of records
- data type
- count of NULL
- distinct
- unique values



Attribute Summary:
name , type and
number of records in
the attribute

DQ Insights: will notify
you if any potential
anomalies have been
identified

Mask analysis: the
structure of the data.

Frequency Analysis: the
number of times each
value in the attribute
occurs.

For different data types profiling
analytics are different.

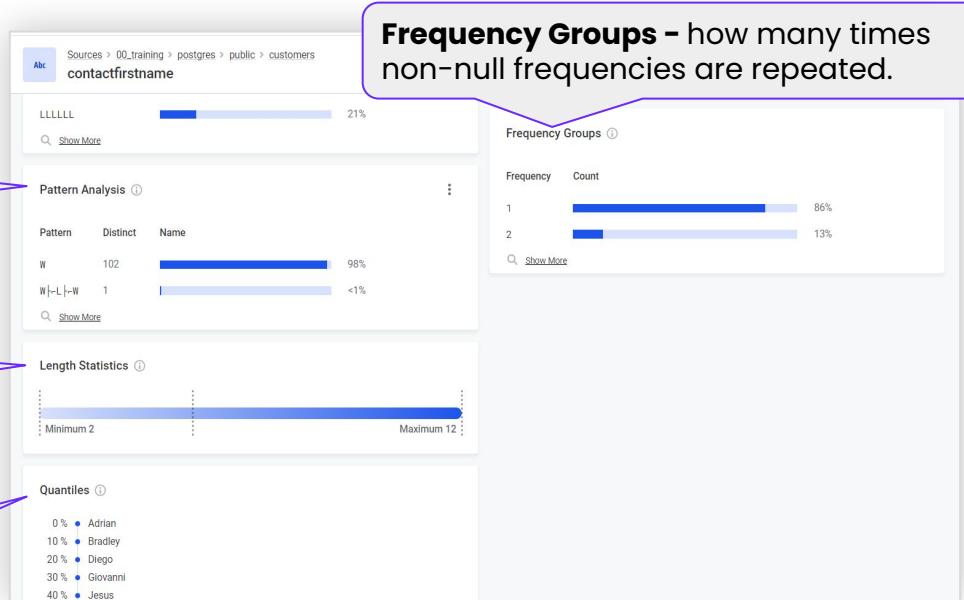
Catalog Items: Profiling & DQ Insights - Attribute Level (2/2)

For different data types profiling analytics are different.

Pattern analysis - how many records have the same pattern and how many of them are distinct.

Length Statistics - The range of records length and minimum and maximum lengths

Quantiles - how many times non-null frequencies are repeated.

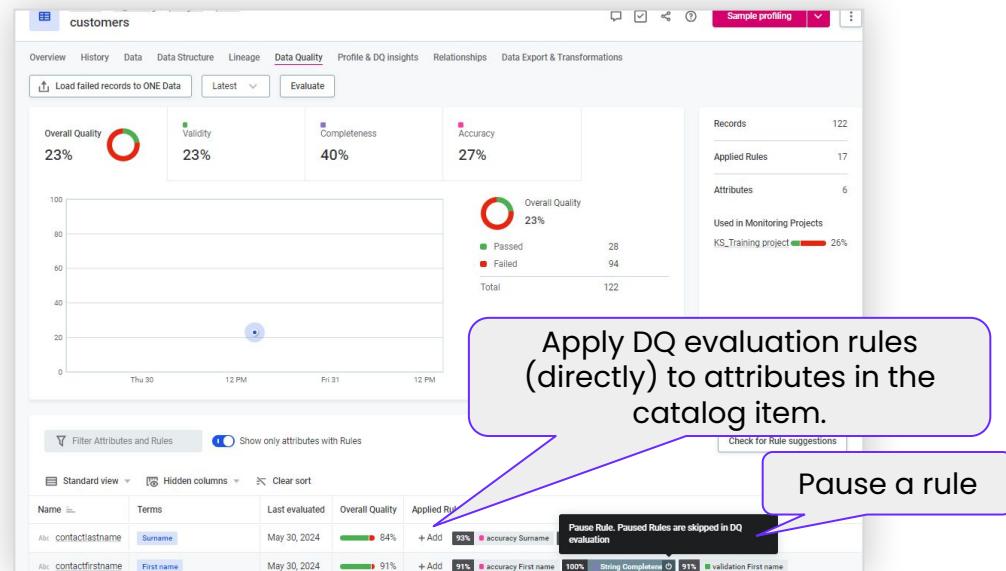


Catalog Item - Data Quality

In order to see the Data Quality results of a catalog item, data quality rule(s) should be added directly or indirectly (through terms) to its attribute(s).

To directly add the rules to an attribute:

- Go to the **Data Quality** tab.
- Locate the attribute's row in the table below the data quality results graph.
- In the Rules column, select (plus) **Add DQ Checks**.
- Choose an existing **rule** from the options available.
- Select **Assign Rule**.
- **Publish** the changes.



- You can **pause** and **reactivate** both terms-assigned and manually-assigned rules as needed.
- When rules are paused, they won't undergo evaluation, and their outcomes won't count to the overall data quality.

SQL Catalog items

Catalog Items can be created using SQL queries for the following cases:

- Combine multiple tables into a single Catalog Item.
- Apply transformations to an existing table before creating a Catalog Item.
- Perform various operations applicable by SQL queries over JDBC or Metastore connection types.

The screenshot shows the Databricks Data Catalog interface. On the left is a sidebar with icons for Knowledge Catalog, Data Catalog, Catalog Items, Reports, Anomaly Overview, Data export projects, Master Data, Reference Data, Not Monitored, Sources, Term Suggestions, and Lineage Import. The main area has tabs for Overview, Data Observability, Connections, Catalog Items (which is selected), History, and Relationships. A search bar is at the top. Below it is a table with columns: Name, Terms, Anomalies, Overall Quality, # Attributes, # Records, and Origin. The table lists several catalog items: 'orders' (postgres), 'customers' (postgres, with a dropdown for Country, First name, Surname, and +4 other columns, showing 23% overall quality), 'products' (postgres, with a dropdown for KS_Product Line and xx_Scale, showing 0% overall quality), 'KS_Orders Cancelled' (postgres, highlighted with a purple border), 'KS_Orders_transformed' (manually created), and 'ks_customers_phone' (postgres). Each item has a three-dot menu icon on the far right.

Name	Terms	Anomalies	Overall Quality	# Attributes	# Records	Origin
orders	-	-	-	8	326	postgres
customers	Country, First name, Surname, +4	23%	15	122	postgres	
products	KS_Product Line, xx_Scale	0%	9	110	postgres	
KS_Orders Cancelled	-	-	-	2	-	postgres
KS_Orders_transformed	-	-	-	9	326	manually created
ks_customers_phone	-	-	-	3	-	postgres

How to create SQL Catalog items?(1/2)

SQL catalog items can be created from two different sections:

- a. Catalog Items
- b. Sources

1. Navigate to **Catalog Items** or select a desired **Source**.
2. Click the three dots menu and select the '**Create SQL Catalog Item**' option.

The screenshot shows the Ataccama Knowledge Catalog interface. On the left is a sidebar with icons for Home, Data Catalog, Catalog Items (which is selected and highlighted in blue), Reports, Anomaly Overview, Data export projects, and Master Data. The main area has a title bar with 'Knowledge Catalog' and 'Catalog Items'. Below the title bar are tabs for 'Published', 'Unpublished', and 'All'. A search bar says 'Type here to search full-text for Catalog items'. Below the search bar are several filter buttons: Terms, Data Quality, Data Source, Location, Number of Attributes, Number of Records, Profile Date, Anomalies, Suggested terms, and Stewardship. At the bottom are buttons for 'Standard view' and 'Hidden columns'.

The screenshot shows the Ataccama Knowledge Catalog interface with 'Sources' selected in the sidebar. The main area has a title bar with 'Knowledge Catalog' and 'Sources xx_training'. Below the title bar are tabs for 'Overview', 'Data Observability', 'Connections', 'Catalog Items', 'History', and 'Relationships'. The 'Overview' tab is active. There is a 'Description' field containing a single dash. Below the description is a table with four columns: Catalog Items, Tables, Database locations, and Schema. Each column contains the value '1'. On the right side, there is a three-dot menu with options: Document, Create SQL Catalog Item (which is highlighted with a purple box), Request access, Instant delete, DQ Settings, and Edit page template.

How to create SQL Catalog items? (2/2)

The image consists of three side-by-side screenshots illustrating the workflow for creating an SQL Catalog item:

- Select Source:** A screenshot of the "Create SQL Catalog Item" interface, step 1. It shows a list of available sources: "postgres-testdata-svc", "invest", "crm", "life", and "mailchimp". The "postgres-testdata-svc" option is selected. The interface includes tabs for "Select source", "Transform using SQL", and "Review catalog item".
- Transform via SQL:** A screenshot of the "Create SQL Catalog Item" interface, step 2. It shows an AI prompt suggesting a query to find customers with a credit limit above 100000. The user has typed the following SQL query:

```
1 SELECT customername, email, creditlimit
2 FROM public.customers
3 WHERE country = 'USA' AND creditlimit > 100000
4 ORDER BY creditlimit DESC;
```

A "Run query" button is present. Below the query, a preview table shows customer data for two rows: "Atelier graphique" and "Signal Gift Stores". The "customername" column is highlighted, and a blue arrow points from the "customername" column to the "contactlastname" column, indicating a transformation mapping.
- Review & Manage:** A screenshot of the "Knowledge Catalog" interface, showing the "Sources" tab for "Postgres Source". It displays metrics like 50 Catalog Items, 37 Tables, 11 Views, 1 DDL query catalog items, and 1 Virtual catalog items. The "Workspace" section shows a folder named "SQL_Cls".

3. Search and choose the **source connection** from the available list.

4. Apply **SQL query** to create the desired catalog item.

- You can use **generative AI** to assist in writing your SQL query.
- Query can be run immediately to check the results.

5. Define/select a folder to store the **SQL CI**.

SQL CIs are visible in the **Workspace** section of the Source.

Topic Highlights

- Detailed profiling and DQ insights are available at item and attribute level for catalog items.
- DQ evaluation and Profiling Schedule is available in catalog items.
- DQ Rules can directly be added to attributes in catalog items.
- Catalog Items can be created by profiling data sources or by applying SQL queries.

Memory Refresher #5 Catalog Items



Reports



Reports

- With **Reports**, you can manage **all BI reports** on catalog items that you build in your BI tool.
- The connection between ONE and your BI tool enables you to:**
 - Display, store and explore** data reports in one place.
 - Visualize** data in several ways.
 - Schedule** periodical **report** updates to get the most up-to-date information on your data.

The screenshot shows the Ataccama Knowledge Catalog interface. On the left, there's a sidebar with 'Knowledge Catalog' and 'Reports' selected. The main area is titled 'Reports' and shows a grid of report cards. One card is highlighted with a yellow border. The cards include: 'College' (Tableau, Samples, Regional), 'Commission Model' (Tableau, Samples, Superstore), 'Company employee overview' (Data Stories), and 'Customer Profitability Sample' (PowerBI, Test Workspace). There are also sections for 'Anomaly Overview', 'Data Export Projects', 'Domain Suggestions', 'Master Data', 'Reference Data', and 'Not Monitored'. A search bar at the top right says 'Type here to search full-text for Report Catalog Items'.

Currently, ataccama support dynamic interactions on reports with the following BI tools:

- Tableau
- PowerBI
- Data Stories

Manage Reports

In ONE:

- You **can** preview reports, adjust the visuals, and configure metadata changes.
- You **cannot** change the report content or structure; this can only be done from the BI tool where this report was built.

To view Reports go to Data Catalog > Catalog Items > Reports:

- Search for reports by their status, applied terms, other parameters, or using full-text search.
- Preview reports, change the visuals, and configure metadata changes in ONE.

The screenshot shows the Ataccama ONE interface with the 'Reports' catalog selected. The left sidebar includes sections for Knowledge Catalog, Data Catalog (Catalog Items, Reports, Anomaly Overview, Data export projects, Master Data, Reference Data, Not Monitored, Sources, Term Suggestions, Lineage Import), and a search bar. The main area displays a grid of report cards. One card is visible with the title 'Company employee overview' and a small icon. Below the card, it says 'DASHBOARD' and 'Analysis Year to Date Summary'. At the bottom, it indicates the source is 'Power BI > Operational Reporting Library (demo data)'. Other cards are partially visible on the right.

Topic Highlights

- With reports, you can manage BI reports on catalog items that you build in your BI tool, such as: **Tableau, PowerBI, Data Stories**
- You can schedule periodical report updates in ONE.
- The report content and structure can be changed only from the BI tool where the report was built.

Memory Refresher #6 Reports



Workshop #1

Catalog & Profiling



Business Glossary



Glossary Terms

Glossary terms can be assigned to different assets in the ONE: Catalog Items, Item attributes, projects, rules, etc.

- Glossary terms assignment on attributes can be **manual** or **automatic**.
- **Terms can be assigned to catalog item attributes based on:**
 - Metadata based rules
 - Data based rules
 - Data and Metadata based rules
- Glossary terms can have hierarchy of relationships.

The screenshot shows the Ataccama Business glossary interface. On the left, a sidebar lists categories like Terms, Business Terms, Key performance i..., Security Terms, Data Domains, Enterprise, Finance, Human Resources, Information Technol..., Manufacturing, Marketing, Procurement, Sales, Supply Chain, and Term Reports. The 'Business Terms' category is highlighted with a purple box. The main panel displays the 'Personal Data' asset details. The 'Overview' tab is selected, showing the following information:

- General information: Abbreviation PD, Term definition source Ataccama Default Business Terms.
- Description: Business definition Personal information might be subject to GDPR, CCPA or other privacy and data protection regulations.
- Occurrence Statistics: Catalog items 0, Attributes 0, Sources 0, Locations 0.
- Relations: Contributes to Personal Identifiable Information (Date of Birth, Special Category Data, Contacts, Surname, Full name, Criminal conviction data, Address, Credit card number, Personal ID), Parent of Date of Birth, Special Category Data, Contacts, Surname, Full name, Criminal conviction data, Address, Credit card number, Personal ID.

Glossary Terms

At term level, the following tabs are available:

- **Overview tab** offers general information about the term together with defined Relations, Stakeholders, assignment suggestions, Occurrence Statistics etc.
- **Occurrence tab** provides the list of Catalog items, attributes and other entities where the term is assigned.
- **Data Quality tab** offers an overview of the quality of the data with the term assigned.
- **History tab** provides a view of previous versions. It allows you to track changes made to the table over time.
- **Settings tab** allows for configuration of term detection and following data quality evaluation.
- **Relationships tab** shows how the terms are related to each other.

The screenshot displays the 'Personal Data' term details in the Ataccama Terms interface. The top navigation bar includes 'Terms', 'Personal Data', and tabs for 'Overview', 'Occurrence', 'Data Quality', 'History', 'Settings', and 'Relationships'. The 'Overview' tab is selected and highlighted with a purple border. The main content area is divided into sections: 'General information' (Abbreviation: PD, Term definition source: Ataccama Default Business Terms), 'Description' (Business definition: Personal information might be subject to GDPR, CCPA or other privacy and data protection regulations), 'Occurrence Statistics' (Catalog items: 0, Attributes: 0, Sources: 0, Locations: 0; Folders: 0, Monitoring projects: 0, Rules: 0, Components: 0), 'Reconciliation projects' (0), and 'Data Quality' (represented by three dots). To the right, there are sections for 'Stewardship' (Owner: Data Office, Data Owner: john.taylor, Data Steward: [redacted], Data Consumer: paul.james, rachel.adams), 'Relations' (Contributes to: Personal identifiable information, Parent of: Date of Birth, Special Category Data, Contacts, Surname, Full name, Criminal conviction data, Address, Credit card number, Personal ID, First name, [CAN] ZIP code, Middle name, Academic title, Gender), and 'Synonyms' (represented by a small icon).

Glossary Terms: Business Hierarchy

- Glossary terms can be stored in an unlimited hierarchical structure (**multiple hierarchies are supported**).
- Relations of terms can be defined on the Overview tab.
- Wide **range of relations** between entities supported: **parent/child, copy of, derived in/from, etc.**
- It is possible to define **new Relations in the Global Settings section** – Relationship types.

Name	Type	Abbreviation	Overall Quality
> Personal Firewalls	Security Term	-	-
> Personal Identifiable Information	Business Term	PII	-
Personal Data	Business Term	PD	-
> Date of Birth	Business Term	DOB	-
> Address	Business Term	ADDR	-
> Surname	Business Term	SUR	84%
> Full name	Business Term	FLN	-
> Contacts	Business Term	CON	-
> Personal ID	Business Term	PID	-
> Gender	Business Term	GEN	-
> Criminal conviction data	Business Term	CCD	-

Relationship Types

Relations

Select Type

Contributes to: Personal Identifiable Information

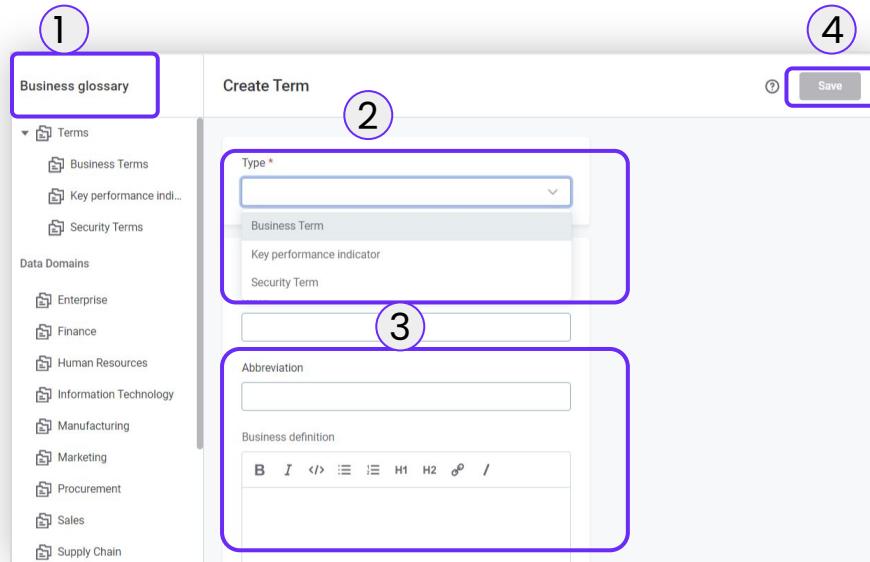
Parent of: Date of Birth, Special Category Data, Contacts, Surname, Full name, Criminal conviction data, Address, Credit card number, Personal ID, First name, [CAN] ZIP code, Middle name, Academic title, Gender

How to Create a term?

1. Go to **Business Glossary > Create**.
2. Fill in **Name** and select the **Type** for the new term.
3. Define other attributes like **Abbreviation**, **Business Definition**, **Owner**, etc. optionally.
4. Save the term.

Note: When the new term is Saved, complementary configurations can be done.

Note: In addition to the out of the box term types, you can create your custom terms.



Term Settings

The screenshot shows the 'Settings' tab for a term named 'xx_Scale'. The left sidebar includes 'Business glossary' and 'Data Domains' sections. The main area has tabs for 'Overview', 'Occurrence', 'Data Quality', 'History', 'Settings' (which is active), and 'Relationships'. The 'Settings' tab contains several sections:

- Detection on attributes:** Contains an 'AI detection' checkbox and a 'Detection Rules' section. The 'Detection Rules' section lists 'xx_Scale' (Based on data) and 'xx_Scale_attribute_name' (Based on metadata), with a threshold of 80%.
- Detection on tables:** Shows a table icon and a 'No conditions' button.
- Data Quality Evaluation:** Contains a checkbox for 'Apply the following rules automatically on all attributes with this term'. Below it are sections for 'Validity', 'Completeness' (with a 'String Completeness' sub-section), 'Uniqueness', 'Accuracy', and 'Timeliness', each with an '+ Add Rule' button.

Mapping terms to attributes can be done:

- manually for specific attribute.
- using Detection Rules.
- through Artificial Intelligence (AI), where the application suggests terms based on the previous mapping behaviors.

The Settings tab of a Term consists of the following configurable sections:

- 1 Detection on attributes
- 2 Conditions for detection on tables
- 3 DQ Evaluation

Term Settings: Detection on attributes

- In order **to detect a term automatically** at the attribute level, **Detection on Attributes** section of the Settings tab should be configured.
- **Detection of terms** on attributes can be achieved both through **AI** and through the **use of detection rules**.
- Appropriate **detection threshold** should be defined for detection rules based on data.

AI Detection:

- To **enable AI term detection** use the checkbox.
- If AI term detection is enabled, terms will be suggested based on attribute similarity and previously applied rules.

To add detection rules to attributes:

- Click **Add Rules** and choose from the list of available rules.
- If multiple detection rules have been added you can select the required operator: **-OR , AND**

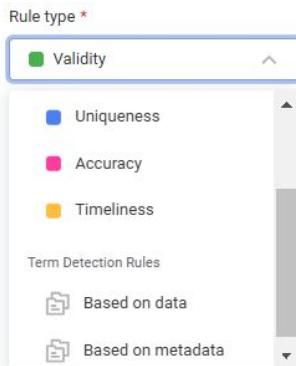
The screenshot shows the Ataccama Business glossary interface. On the left, there's a sidebar with 'Business glossary', 'Data Domains' (listing Enterprise, Finance, Human Resources, Information Technology, Manufacturing, Marketing, Procurement, Sales, Supply Chain), and 'Term Reports' (listing Mapped to Catalog, Suggested Mappings, Metadata Detection). The main area shows the 'xx_Scale' metadata model. The 'Settings' tab is active. Under 'Detection on attributes', there's a checkbox for 'AI detection' (checked) and a 'Detection Rules' section. The 'xx_Scale' rule is listed with a threshold of 80% and an 'OR' operator. Another rule, 'xx_Scale_attribute_name', is listed as 'Based on metadata'. Below this is a 'Detection on tables' section with 'No conditions' and an 'Add Condition' button.

Note: Before adding rules in the Settings tab, they should be created in Data Quality section of the application.

Rules

Rules are located under the **Data Quality** section.

- **There are two types of rules:**
 - **Detection rules** – for detecting business terms:
 - Based on **data**
 - Based on **metadata** (attribute, catalog item, connection, location name/description)
 - **DQ Evaluation rules**



A screenshot of the 'Create Rule' dialog in the Ataccama Data Quality interface. The 'General information' section includes fields for 'Name *' (set to 'Test_rule') and 'Rule type *' (set to 'Validity'). The 'Description' section contains a rich text editor with the text 'This is a test.' and 'Generated by AI: Review might be needed'. A blue callout bubble points from this text area to a larger callout box on the right.

Generative AI can help with grammar mistakes in the description.

Detection Rules

- Selecting Term Detection as the rule type as well as the type of detection is done while rule creation before saving.
- Rule logic configuration is located under the **Implementation** tab.
- The condition should be defined so that if satisfied, the term is detected.
- **There are two options for defining the condition:**
 - **Condition Builder** – offers predefined logical operations like 'is empty' or 'is from a reference data'.
 - **Advanced Expression** – offers writing complex condition.

The screenshot shows the Ataccama Rule Editor interface for creating a 'Term Detection Rule: Based on data'. The rule is titled 'Continent'. The 'Description' field contains the text 'Detection rule for continents.' The 'Inputs' section shows an 'ATTRIBUTES' dropdown set to 'Continent' and a 'value' input field. The 'VARIABLES' section has a '+ Add variable' button. The 'Rule Logic' tab is selected, showing a 'Condition' section with a 'WHEN' clause containing a complex logical expression involving 'abc', 'value', 'lowercase', 'is from list', and various continent names. Below the condition is a 'THEN' section with the text 'Detect Term Continent'. A purple box highlights the 'Condition' section. Two callout bubbles point to this area: one pointing to the 'Ask AI' button with the text 'you can use generative AI to generate the expression from a text prompt.', and another pointing to both the 'Condition' and 'THEN' sections with the text 'choose the option either condition builder or Advanced expression for defining the condition'.

Term Settings: Detection on tables

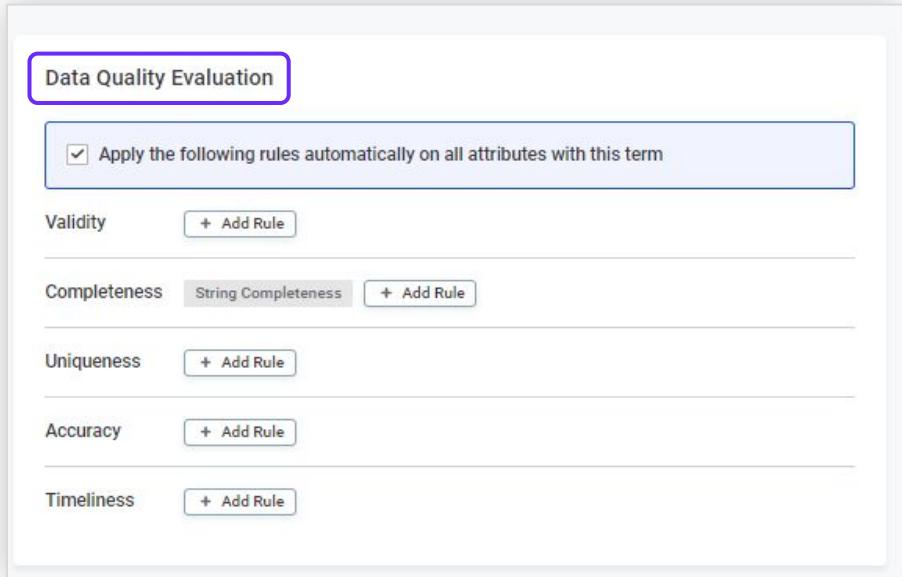
- Besides assigning terms to attributes, it is also possible to assign a term to a catalog item based on the terms assigned to its attributes.
- The condition for the detection can be either **AND** or **OR**.

The screenshot shows the 'Detection on tables' configuration screen. At the top, there are three buttons: 'Or', 'And', and '+ Add Condition'. The 'And' button is highlighted with a purple border. Below these buttons is a text input field containing the term name 'Detection on tables'. A descriptive text below the input field reads: 'Assign this term on a table if its attributes have the following terms'. The main configuration area contains a list of terms separated by 'and': '[CAN] Social Insurance Number' and '[USA] Social Security Number' and 'E-mail' and '[North America] Phone Number' and 'Phone number' and '[GBR] Phone Number' and 'Address' and 'Religious or philosophical beliefs' and 'Postal code' and 'Zip code' and '[GBR] ZIP Code' and '[CAN] ZIP code' and '[USA] Zip code' and 'Medical record number' and 'Health plan beneficiary number' and 'Health' and 'Date of Birth' and 'Place of Birth' and 'First name' and 'Surname'. There are two rows of '+ Add Term' buttons, each preceded by an 'Or' button and followed by a three-dot ellipsis. The entire configuration is enclosed in a large rectangular box with a purple border.

Term Settings: Data Quality Evaluation

- The data quality rules can be added to each term through the **Settings tab – Data Quality Evaluation section**.
- If a Catalog Item attribute has assigned term(s), its Data Quality will be available for Evaluation based on the DQ Evaluation rules added to the term(s).
- Overall Data Quality** of a Catalog Item will be evaluated based on the DQ results of its attributes.

Note: DQ Evaluation rules can also be directly added to catalog item attributes.



Business Glossary: Unmapping Terms

- Only users with editing permissions can remove assigned terms.
- **Removing an automatically assigned term** will turn off automatic detection of the term on that attribute or table.

The screenshot shows the Ataccama Business Glossary interface for the 'employees' table. The top navigation bar includes 'Sources > pg5_testdata > tcd > transactional_customer_data'. Below the navigation is a horizontal menu with tabs: Overview (which is selected), History, Data, Data Structure, Lineage, Data Quality, Profile & DQ insights, and Relationships. The main content area is titled 'Attributes' and contains a list of attributes with their current values and detection status. The attributes listed are:

Attribute	Value	Detection Status
employeenumber		Detected
lastname	Surnar	Detected
firstname	First name	Detected
extension		Detected
email	E-mail	Detected
officecode		Detected
reportsto		Detected
jobtitle		Detected

A context menu is open over the 'Surname' attribute, with a purple box highlighting the 'Delete' option. The menu contains the following text: 'Are you sure you want to delete Surname? Deleting detected term will turn off automatic detection of term "Surname" on this attribute'. At the bottom of the menu are two buttons: 'Delete' (highlighted with a purple box) and 'Cancel'.

Term Suggestions: Overview

Term Suggestions is proposing business terms based on the terms previously assigned to similar attributes.

There are two ways to access the Term Suggestions tab:

- Navigate to **Knowledge Catalog > Data Catalog > Term Suggestions**.
- Select any assigned term. In the sidebar that opens, select **Show all N suggestions**. This takes you to the Term Suggestions tab and automatically filters the results.

The screenshot shows the Ataccama Knowledge Catalog interface. On the left, a sidebar lists various catalog items: Data Catalog, Catalog Items (with a dropdown arrow), Reports, Anomaly Overview, Data export projects, Master Data, Reference Data, Not Monitored, Sources, and Term Suggestions. The 'Term Suggestions' item is highlighted with a purple box. The main area is titled 'Term Suggestions' and contains a search bar with placeholder text 'Type here to search full-text for Term suggestions'. Below the search bar are filter buttons for Confidence score, Catalog Item, Status (set to Any), Location, Attribute, and Suggested Term. A table displays suggested terms, with columns for Action (Rejected), Suggested Term (KS_Product Line or xx_Scale), Confidence (100%), Attribute (productline or productscale), and Location (xx_training > postgres > public > products). To the right of the table, there's a sidebar for 'Address' with sections for General information, Abbreviation (ADDR), and Term definition source (Ataccama Default Business Terms). At the bottom right, there's an 'Insights' section with a button labeled 'Show all 58 suggestions'.

Term Suggestions

The screenshot shows the Ataccama Data Catalog interface with several callout boxes explaining the Term Suggestions feature:

- The term proposed for the attribute.** Points to the "Suggested Term" column in the main table.
- The likelihood that the term matches the attribute.** Points to the "Confidence" column in the main table.
- The attribute that the term will be applied to.** Points to the "Attribute" column in the main table.
- You can approve or reject suggested terms here.** Points to the "Action" column in the main table, which contains checkboxes for accepting or rejecting suggestions.
- Attribute location in your sources.** Points to the "Location" column in the main table, which shows the data lineage for each attribute.

Data Catalog

- Catalog Items
- Reports
- Overview
- Import projects
- Lineage
- Not Monitored
- Sources
- Term Suggestions**
- Lineage Import

Type here to search full-text

Action Suggested Term Confidence Attribute Location

Action	Suggested Term	Confidence	Attribute	Location
<input checked="" type="checkbox"/> <input type="checkbox"/>	First name	100%	FIRST_NAME	Analytics Snowflake DWH > ... > ACTIVE_CUSTOMER_SOURCE
<input checked="" type="checkbox"/> <input type="checkbox"/>	[USA] City	100%	city	pgs_testdata > ... > customers
<input checked="" type="checkbox"/> <input type="checkbox"/>	First name	100%	FIRST_NAME	Oracle > ... > ACTIVE_CUSTOMER_SOURCE
<input checked="" type="checkbox"/> <input type="checkbox"/>	[USA] Zip code	100%	postalcode	pgs_testdata > ... > customers
<input checked="" type="checkbox"/> <input type="checkbox"/>	Address	100%	ADDRESS	Oracle > ... > EMPLOYEE_ALL
<input checked="" type="checkbox"/> <input type="checkbox"/>	Address	100%	ADDRESS	Oracle > ... > EMPLOYEE_CONTACTS
<input checked="" type="checkbox"/> <input type="checkbox"/>	Address	100%	addressline1	pgs_testdata > ... > offices_prime
<input checked="" type="checkbox"/> <input type="checkbox"/>	Customer code	100%	customernumber	pgs_testdata > ... > payments_prime
<input checked="" type="checkbox"/> <input type="checkbox"/>	Address	100%	addressline1	pgs_testdata > ... > customers
<input checked="" type="checkbox"/> <input type="checkbox"/>	Zip code	100%	POSTALZIP	Oracle > ... > EMPLOYEE_CONTACTS

Show 10 more 1–10 of 952 items

Term Detection: Configuration

Term detection and suggestion can be configured in Global settings:

- Term detection rules:** Metadata rules can be excluded from term detection to ensure better performance of ONE by activating the term detection toggle.
- Term suggestions:** The Latest reflected state shows the date and time of the current synchronization status.

The screenshot shows the Global Settings sidebar with various icons and sections. The 'Term Detection' section is highlighted with a purple border. Inside this section, there is a toggle switch labeled 'Limit detection based on metadata to improve performance'. Below the toggle, two explanatory text blocks are visible: one about excluding 'Or' logic and another about combined 'And' logic. The 'Term suggestions' section below contains a timestamp 'today at 11:10 PM' and a link to 'Read docs about Term suggestions'. At the bottom are 'Synchronize' and 'Available' buttons.

Global Settings

Maintenance Center

System changes

Metadata model

Database backups

Validations

Custom layouts

AI

Gen AI

Term Detection

Application Settings

Content Security P...

Term Detection

Term detection rules

Limit detection based on metadata to improve performance

Metadata rules will be excluded from detection that uses 'Or' logic in the Term settings.

Metadata rules combined with data rules using 'And' logic will still work.

Term suggestions

Terms suggestions are based on changes done by users or during profiling such as assigning of Terms. Changes can't be processed immediately, so very recent changes won't affect Term suggestion.

The most recent change fully reflected in Term suggestions was made: today at 11:10 PM

Synchronize the term suggestions again after every app recovery from a backup.

Read docs about Term suggestions

Synchronize Available

Topic Highlights

- Terms can be assigned to different assets in the ONE.
- Custom term types can be defined other than default term types.
- Glossary Terms have unlimited hierarchical structure and multiple hierarchies are supported.
- Approving or rejecting term suggestions directly, avoids the need to inspect each attribute individually.
- Approvals and rejections made in the Term Suggestions tab are automatically published.
- Terms can be manually or automatically assigned to catalog items and their attributes.
- Term settings through adding rules and conditions can automate the rules assignment to Cls and attributes.
- Detection and DQ Evaluation are the two types of rules in the ONE.
- Defining Condition for the rules can be done through the Condition Builder, Advanced Expression or AI.

Memory Refresher #7 Business Glossary



Workshop #2

Glossary



Data Observability



ataccama

Data Observability in Ataccama ONE

Data observability allows you to understand the state of your data based on the information you're collecting about that data, such as data quality issues, anomalies, or schema changes.

The key components of data observability are:

- Data quality monitoring
- AI-powered anomaly detection
- Data discovery
- Freshness

With these components in place you can better understand the nature of your data, including:

- Data quality information.
- Schema changes. (Virtual catalog items (VCI's) and files cannot undergo schema checks as schema checks are run on the database directly.)
- Anomalies
- Changes in business domains of data
- Data freshness
- Any other metadata coming from ETL processes

Data Observability in Ataccama ONE

The Data observability in ONE relies on the relationship between the data catalog, business glossary, and rules:

- Business terms are defined in the business glossary, data quality rules are applied to terms, and terms are applied to items in the catalog during data discovery.
- As the module contains a number of predefined rules and terms, the process is mainly automated.
- In addition to this process, AI detects anomalies and suggests business terms.

The following steps outline Data Observability works in the ONE:

- The discovery process analyzes the data and detects business terms within specific attributes.
- You select the business terms you want to monitor.
- Schedule data observability runs.
- Get notifications, improve AI, consume statistics on the data observability dashboard, analyze issues, and fix them.

Overview of the Observed Systems

The status of all observed systems can be presented after selecting Data Observability in the left navigation menu.

The screenshot shows the Ataccama Data Observability dashboard. At the top, it displays "Total" and "Observing 6 sources". A message indicates "4 sources with issues". Below this is a search bar and a filter bar with various dropdowns and buttons like "Source name", "Connection type", "Catalog items count", "DO state", "With issues", "Paused", "No issues", "DO monitoring", "DO issue count", "Term", "Stewardship", and "Clear filters". The main area shows a table of "Recent open issues" for six sources. Each row includes a system name, connection type, and date range. The first source has "No issues". The second source has "Detected issues" from 30/12/2023 to 01/01/2024, with "0" detected and "0" open/closed issues. The third source has "No issues". The fourth source has "No issues". The fifth source has "No issues". The sixth source has "No issues". To the right of the table are buttons for "Open dashboard" and "Edit configuration". On the far left is a sidebar with icons for Home, Search, Issues, Catalog, Monitoring, Configuration, and Help. A legend at the bottom right defines the colors: blue for "No issues", green for "With issues", yellow for "Paused", and red for "No issues".

You can filter the selection by source name, connection type, catalog item count, DO State, DO monitoring, DO Issue count, Term and stewardship.

Use the icons to switch between table view and list

You can filter the selection by selected time frame (such as day, week, Last 7 days, month)

To view results or edit the configuration, click on the system name or select the three dots menu.

Data Observability Dashboards

In Knowledge Catalog, go to Sources



Go to the Data Observability tab

After configuration, you can manually run the observability

You can pause observation at any point. This is particularly useful if you need to troubleshoot issues. When observability is paused scheduled monitoring will not be run.

Click on setting

Set up Data Observability

- **To use the Data Observability module for your data source:**
 1. Connect to a data source.
 2. Run data discovery.
 3. Configure data observability.
- If you are connecting to a **new source**, first you need to **run data discovery**(the import and initial analysis of the metadata in your connected system.)
- If you want to configure observability on an **already connected** source, select the source, navigate to the **Data Observability** tab and **Configure** observability.
- To **edit** an existing configuration, select the required source, navigate to the Data Observability tab and select the settings icon.

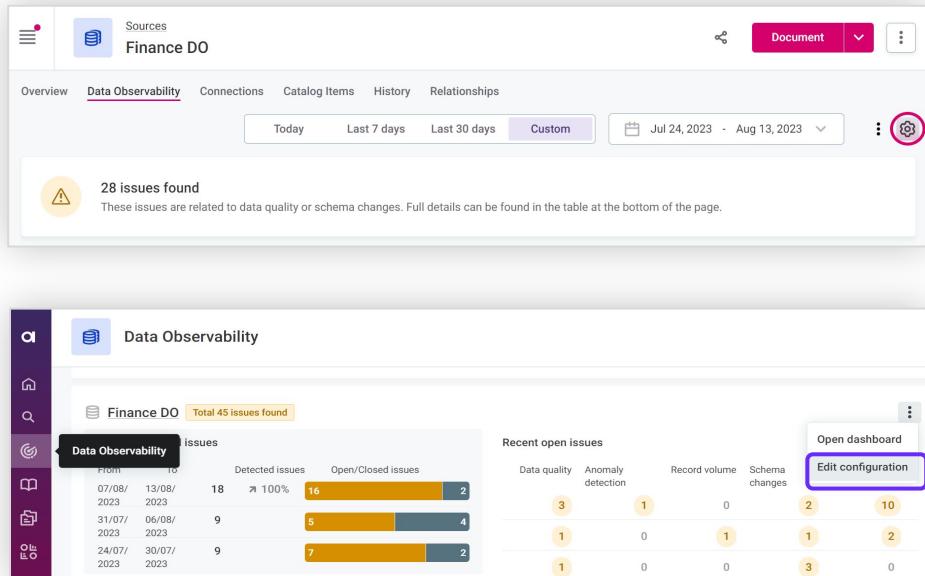
Data discovery

- It is necessary to run at least sample profiling on your source before configuring data observability.
- This can be done at the **source level**, using one of the **documentation flows** (Discover or Document), or at the **catalog item level**, using **sample** or **full profiling**.

The screenshot shows the Ataccama Data Catalog interface for a Snowflake source. The top navigation bar includes 'Sources' and a 'Snowflake' icon. Below the navigation, there are three tabs: 'Overview' (underlined), 'Data Observability', and 'Connections'. The 'Overview' tab displays a 'Description' section with the text 'Snowflake Data Source (Demo content)'. On the left, there are two summary boxes: 'Catalog Items' (3) and 'Tables' (3). On the right, there are two summary boxes: 'Data quality issues' (1) and 'Content issues' (1). A modal window is open over the interface, titled 'Document'. It contains three options: 'Import' (described as importing all metadata without accessing data), 'Discover' (described as a quick, sample-based discovery), and 'Document' (described as importing metadata, running quick data discovery, and profiling data quality). The 'Document' option is highlighted with a purple rounded rectangle.

Configure Data Observability

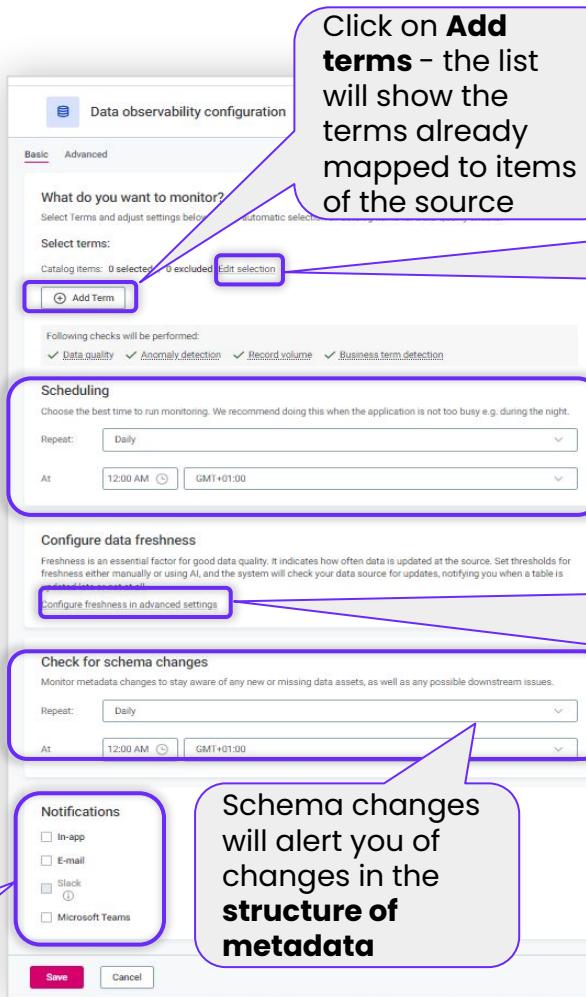
- Once you have an active connection to your data source and the data discovery process is finished, you can **configure** data observability.
- This involves selecting what you want to **observe**, how **regularly**, and how you want to be notified in case of issues.
- To **edit** an existing configuration, select the **settings** icon on the **Data Observability tab of the source** or navigate to **Data Observability in the navigation menu** and use the three dots menu for the relevant listing.
- To **create** a new data observability configuration, select the **settings** icon on the Data Observability tab of the source.
 - Basic configuration
 - Advanced configuration



Basic Configuration

Data Observability configuration enable users to select terms to be monitored and schedule monitoring:

- Selected **Glossary Terms** along with their added **rules** play a **key role** in Data Observability process.
- Select available terms associated with important items from the list.
- If no term gets added, observability will not proceed.
- Using basic configuration, you are alerted by default when new terms are detected in the source or when terms are newly suggested on observed catalog items.



Select how you would like to be alerted

Click on **Add terms** – the list will show the terms already mapped to items of the source

Click on **Edit selection** to exclude any of items from monitoring.

Select the **intervals, time, and time zone** at which you want the monitoring to be run

Freshness checks look at the metadata of a catalog item to see when it was last updated

Advanced Configuration (1/2)

Select the Advanced tab to go beyond the default settings:

- Use advanced configuration if you want more control over notifications, alert thresholds and anomaly detection sensitivity.
- Based on notification setting you will be alerted to changes in the data structure: Table (New, Missing), Column (New, Missing, Data type)

Get notified when **new terms are detected in your observed** system or when existing terms are newly added to attributes. You can also enable notifications for when terms are applied manually.

The screenshot shows the 'Data observability configuration' page with the 'Advanced' tab selected. In the 'What do you want to monitor?' section, 'AQL' is highlighted. Under 'Processing settings', 'Overall Data Quality' is checked. Two threshold options are shown: 'Manual threshold' (checkbox checked, threshold set to 10 %) and 'Adaptive threshold' (checkbox checked, sensitivity set to Medium). The 'Anomaly detection and record volume' and 'Business term detection' sections are also visible. The 'Scheduling' section at the bottom shows a daily schedule from 12:00 AM to GMT+01:00. Buttons for 'Save' and 'Cancel' are at the bottom.

Use **Ataccama Query Language** to define the data you want to observe.

You need to specify the required threshold (%); If this option is turned on, you will be alerted whenever changes in data quality exceed this threshold.

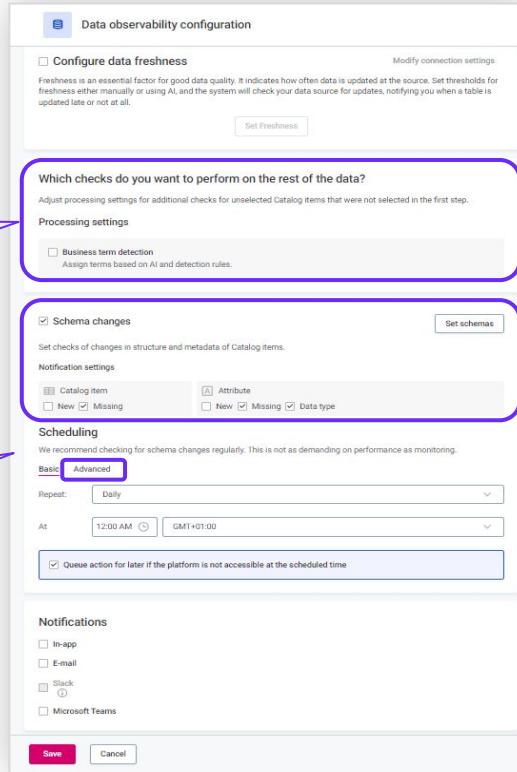
Select the required sensitivity:
High sensitivity -> more anomalous values -> may be **false positives**.
Low sensitivity -> less anomalous values -> may be **false negatives**.

Advanced Configuration (2/2)

Data without selected term:

Specify whether Business term detection is active. If this option is turned on, terms can be assigned when appropriate on any new catalog items.

Schedule monitoring using either basic or advanced settings



Be alerted of any changes in the structure of metadata, such as missing columns or changes in data type. VCIs and files are not included in schema change checks.

Sample Results: No Issues

The screenshot shows the Ataccama Data Observatory interface with several annotations:

- A purple callout points to the top center of the page with the text "No issues observed".
- A purple callout points to the top right corner with the text "Observing timeframe".
- A purple callout points to the bottom right corner with the text "The structure of the data source outlines the observing objects".

Observing Data Quality results
Checked every day at 00:00 GMT+10:00

Data Quality No issues
Anomaly Detection No issues
Record Volume No issues

Structure

Catalog items	158
Catalog items with terms	60
Connections	1

Schema Changes No issues

Data assets	+0
Attributes	+0
Data type change	0

Business Terms No issues

Terms	Total	New
Total	27	16
New	16	

Checked Catalog items: -

All Issues Data Quality Anomaly Detection Record Volume Schema Changes Terms **Structure**

Standard view Hidden columns

Name	Terms	Attributes	Anomalies	Overall Quality	VAL	COM	UNI	ACC	TML
customers	Personal Data, Surname, [USA] State +4	15	-	-	-	-	-	-	-
orders	no terms assigned yet	8	-	-	-	-	-	-	-
products	dwh_scale, scale, scale_d, PM_Scale	9	-	100%	N/A	100%	N/A	N/A	N/A
customer	Personal Data, Surname, [USA] State +7	9	-	-	-	-	-	-	-
customer_two	Personal Data, Surname, [USA] State +7	9	-	0%	0%	89%	N/A	18%	N/A
names	no terms assigned yet	2	-	-	-	-	-	-	-

Sample Results: Issues Found

The screenshot shows the Ataccama Data Observatory interface for the 'Finance DO' source. It highlights several key findings:

- Schema Changes:** A callout notes "Schema changes have been observed." pointing to a section showing 3 issues related to data assets, attributes, and data type changes.
- Validity, Anomaly Detection, and Record Volume:** A large callout notes "Thresholds were exceeded for **Validity, Anomaly Detection and Record Volume**. Alerts are sent as per configuration." pointing to sections for Data Quality (39%), Anomaly Detection (1 issue), and Record Volume (1 issue).
- Issue Types:** A callout notes "Issue Types are explained against each catalog item." pointing to a table where the first catalog item, 'LOAN_LAC', is listed with four 'Schema Change issue' entries.
- Validity:** A callout notes "Validity was exceeded on the catalog item." pointing to the 'Overall Quality Change' column for 'LOAN_LAC' which shows a red bar at 54.8%.
- Tasks:** A callout notes "Create associated **tasks** directly from the issues table by selecting the plus icon." pointing to a 'Task' column with a plus icon next to each row.

Data Quality results (3 issues):
Manual threshold: 3%
Adaptive threshold sensitivity: MEDIUM
Checked Catalog items: 3/15

Anomaly Detection: 1 issue
1 anomalous catalog item
AI sensitivity: MEDIUM
Checked Catalog items: 2/16

Record Volume: 1 issue
1 anomalous catalog item
AI sensitivity: MEDIUM
Checked Catalog items: 2/16

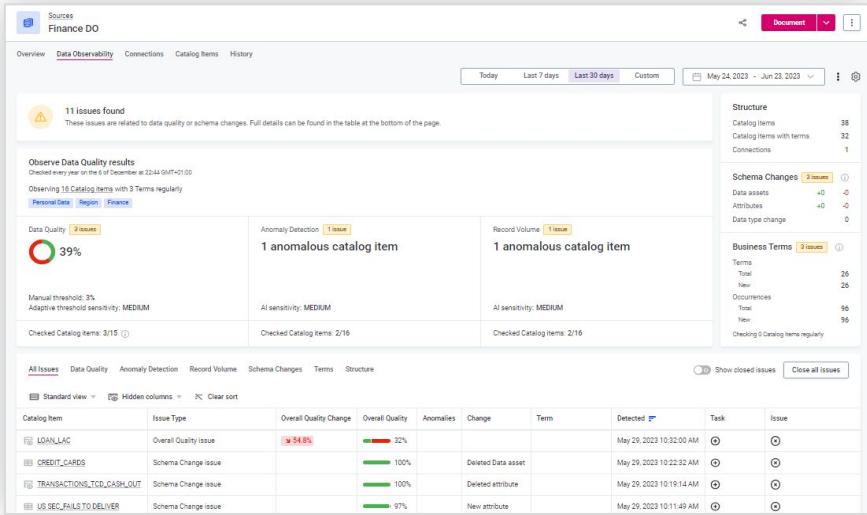
Schema Changes: 3 issues
Data assets: +0 -0
Attributes: +0 -0
Data type change: 0

Business Terms: 3 issues
Terms: Total New Occurrences Total

Catalog Item	Issue Type	Overall Quality Change	Overall Quality	Anomalies	Change	Term	Detected
LOAN_LAC	Overall Quality Issue	54.8%	32%				May 29, 2023 10:32:00 AM
CREDIT_CARDS	Schema Change issue	100%	100%		Deleted Data asset		May 29, 2023 10:22:32 AM
TRANSACTIONS_TCD_C	Schema Change issue	99%	99%		Deleted attribute		May 29, 2023 10:19:14 AM
US_SEC_FAILURES	Schema Change issue	97%	97%		New attribute		May 29, 2023 10:11:49 AM

Topic Highlights

- **Data Observability** relies on the **relationship** between the **data catalog**, **business glossary**, and **rules**.
- **Key components** of data observability are: DQ monitoring, AI-powered anomaly detection, Data discovery, Freshness evaluation.
- The selection of **glossary terms** to be mapped to the catalog items of each source is a mandatory step of Data Observability.
- Data Observability **Basic** configuration options contain the default settings.
- **Advanced** configuration provides more control over **alert thresholds** and anomaly **detection sensitivity**.



Memory Refresher #8

Data Observability



Catalog & Glossary

v15.4.x



ataccama