Data Sentinel 1.0.0 – Setup Guide

Prescience Data Sentinel is a data quality assessment and transformation tool with over 60+ prebuilt data quality and data transformation rules. The tool allows enterprises to pick from these existing rules or create their own new rules. It then provides a comprehensive analysis of data and automatically generates data quality scores and reports. It also has a built-in data transformation engine which helps in cleaning and preprocessing data for downstream applications and systems.

Data Sentinel supports ingestion of data from multiple data sources like **Redshift, Snowflake, Databricks, SQL databases, CSV, JSON, Parquet files** etc, thus enabling an end-to-end view of the enterprise and business unit data. It securely ingests enterprise data without copying it and can be deployed as a standalone solution or be integrated with your existing data pipelines.

Depending on the user profile, with the Data Sentinel tool you can

1. Create a new tenant
2. Create users with name, email address, roles, login credentials
3. Select master rules for data quality and data transformation
4. Create new rules
5. Create, view, and edit connections
6. Create, view, and edit output locations
7. Create, view, and edit datasets
8. Create, view, and edit rule suites which are a combination of different rules
9. Create, view, edit, show results and history of each Jobs
10. Schedule jobs to run at specific time intervals
11. View job history with data quality and technical details of each run
12. View data quality dashboards
13. Clean and Transform data
14. Send notifications for job success or failure based on threshold

This trial version of the tool can be used as many times as required, for a period of 45 days from the registration.

1. User Profiles

This tool has been developed for three user profiles – Super Admin, Tenant Admin, and Data Steward.

The Super Admin can

- View basic details such as available master rules
- View the count of connections, data sets, rule suites, jobs, historical jobs runs, etc.
- Create Tenant
- Create master rules
- Configure the overall Data Sentinel tool

The Tenant Admin can

- Access dashboards with data quality scores, data transformation reports, historic trends, data quality dimension scoring, source system performance etc.
- Create users
- Create & configure connections
- Create and edit data sets & define output locations
- Create and edit rule suites
- Create & schedule jobs
- View job history
- Send notifications for job success or failure based on threshold

The Data Steward can

- Access dashboards with data quality scores, data transformation reports, historic trends, data quality dimension scoring, data rows processed, run time, rule split etc.
- View connections
- Create and edit datasets
- Create and edit rule suites
- Create & schedule jobs
- View job history

2. Getting Started

The Super Admin profile is used to login for the first time and configure the company profile. Once this is done, users who are in the administrator and data steward roles can access the tool.

The following steps are required to configure the application before using it on your enterprise data.

2.1 Super Admin Overview

Below are the different steps for logging in and configuring the company profile.

2.1.1 Login using the Super Admin

The details for logging into the Data Sentinel tool are as follows.

Application URL: http://<server_IP_address>
Default Username: admin
Default Password: Dqadmin123$

Image 1: Login page

Login using the default login credentials for super admin, then navigate to admin icon on the right side of the page. Select the drop-down and click on **Change password**.

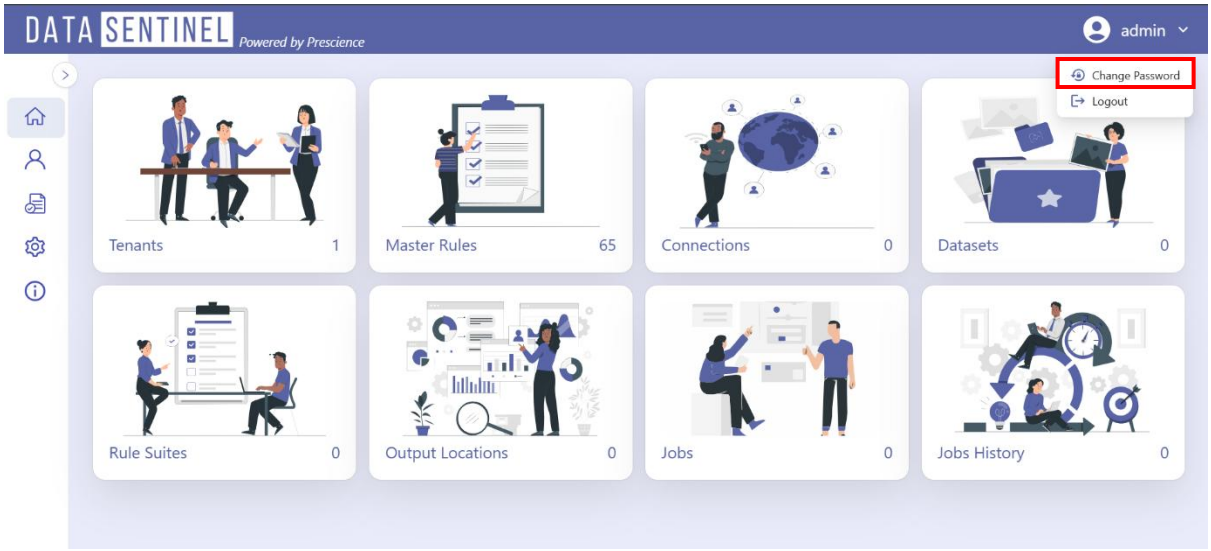

Image 2: Change password

Note: For security reasons, it is crucial to change the super admin password after the product deployment.

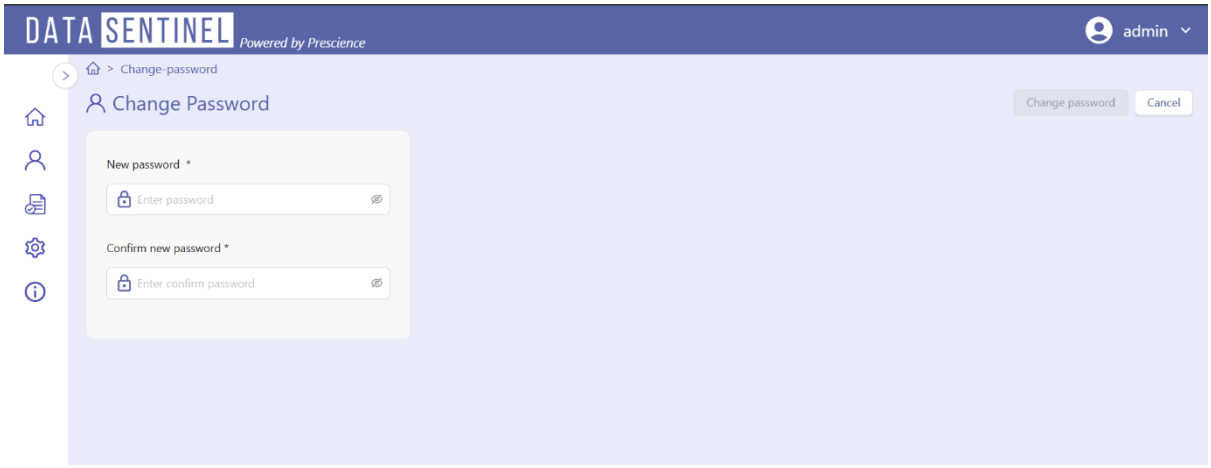

Image 3: Change Password page

2.1.2. Update Company logo

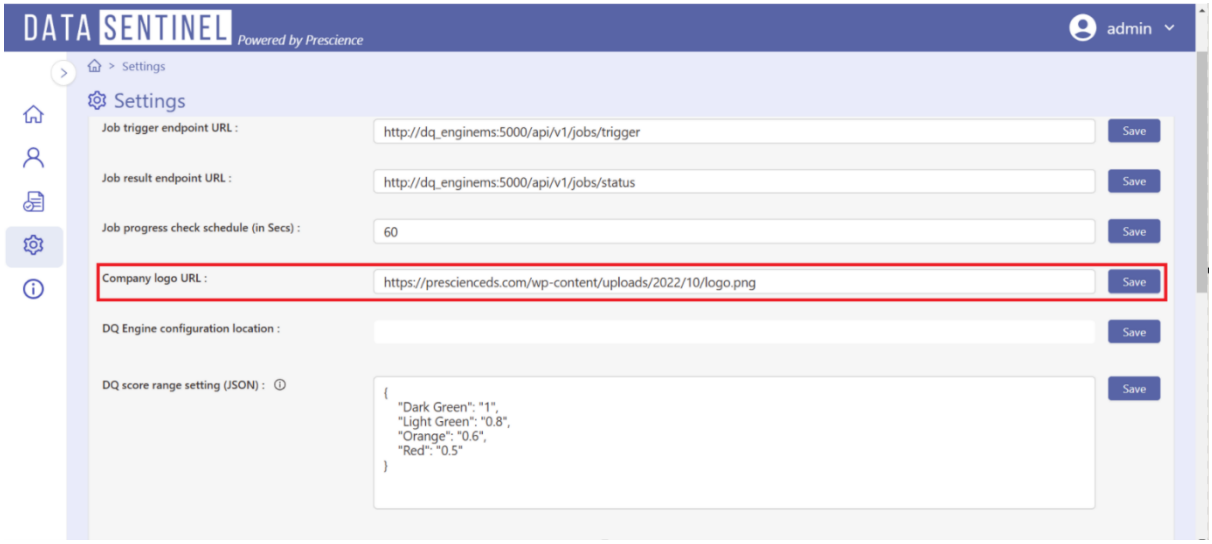

Image 4: Settings page to update company logo

To set up your company logo to display on the login page:

Access the Super Admin dashboard and proceed to the **Settings** section. Once there, find the field labelled **Company logo URL**. Update your company's logo URL into this field. Make sure to save the changes before logging out of the super admin account. After logout, your company logo will appear on the login page.

2.1.3. Create a Tenant

This is the company registration page to register your company. Enter the details of your company in the following fields.
Please remember subdomain, username, password for company/tenant admin to login to the application later.

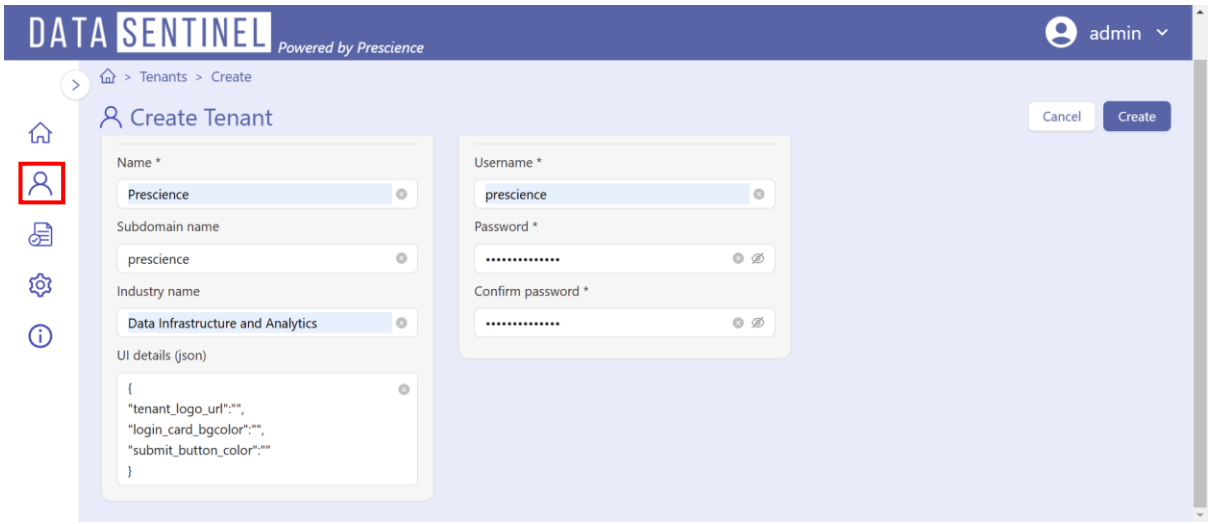

Image 5: Company/Tenant creation

2.1.4. Master rule

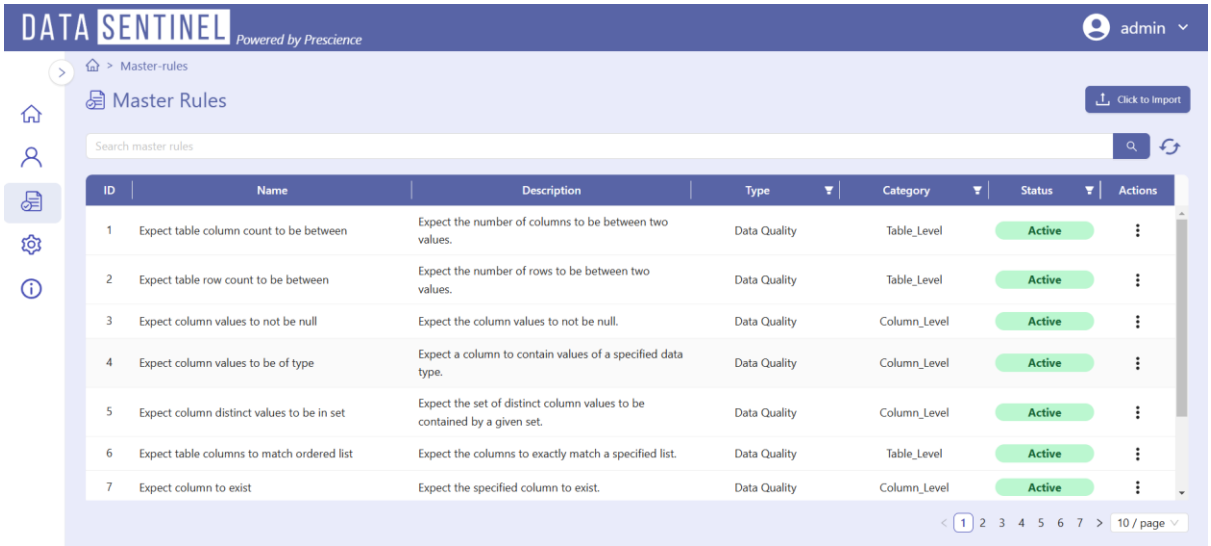

Image 6: Master rules page

Master rules are the set of predefined rules which can be applied on the dataset. They can be categorized in 2 types: **Data Quality** and **Data Transformation** rules. Data quality rules are applied to validate the data and check the data quality score. Data transformation rules are applied to clean and transform data.

2.2 Tenant Admin & Data Steward Overview

Below is a walkthrough of the different functionality that is available for the tenant admin and data steward profiles.

2.2.1. Logging In to Company

The URL for tenant admin login: http://<server_IP_address/?client={subdomain}
For e.g. http://10.20.30.1/?client=prescience

You will now be redirected to a login page.
Use the tenant admin credentials as created above to login to the tenant admin.

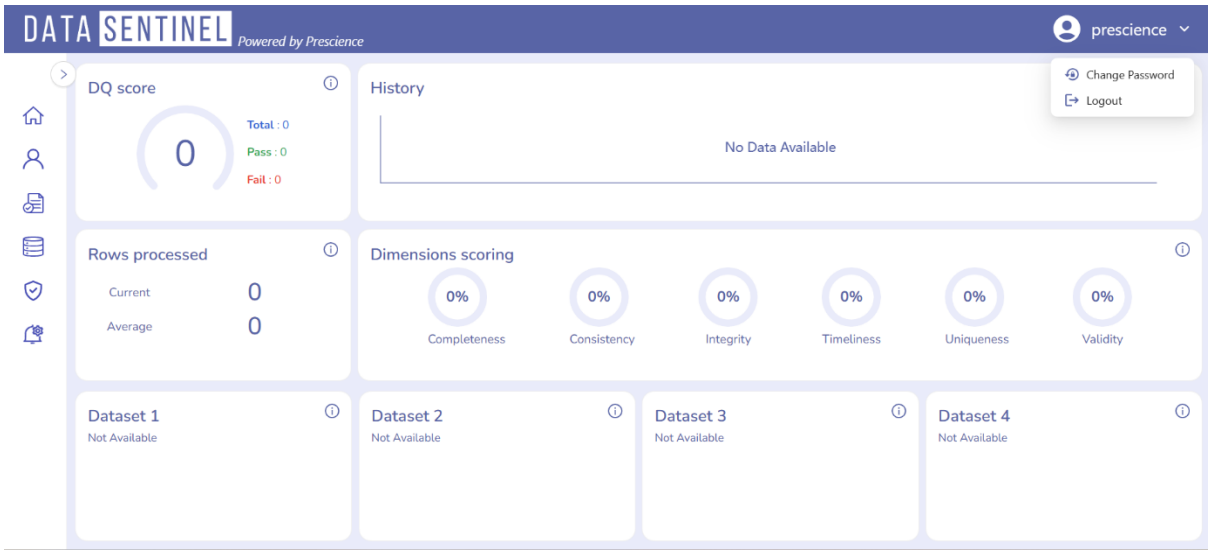

Image 7: Tenant admin dashboard

2.2.2. Create New Users

Go to the Create User tab and enter all the required details.

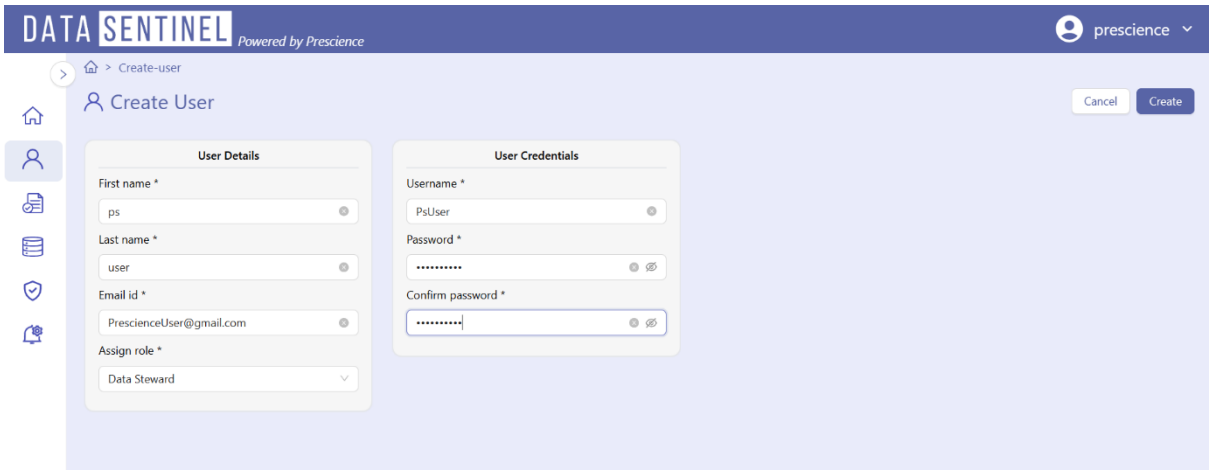

Image 8: Creating new users

Tenant admin can further create data steward or other tenant admin users.

Logging in to newly created tenant users

The URL for tenant user's login: http://<server_IP_address>/?client={subdomain} (use the same subdomain name that you used while creating the tenant admin)

For e.g. http://10.20.30.1/?client=prescience

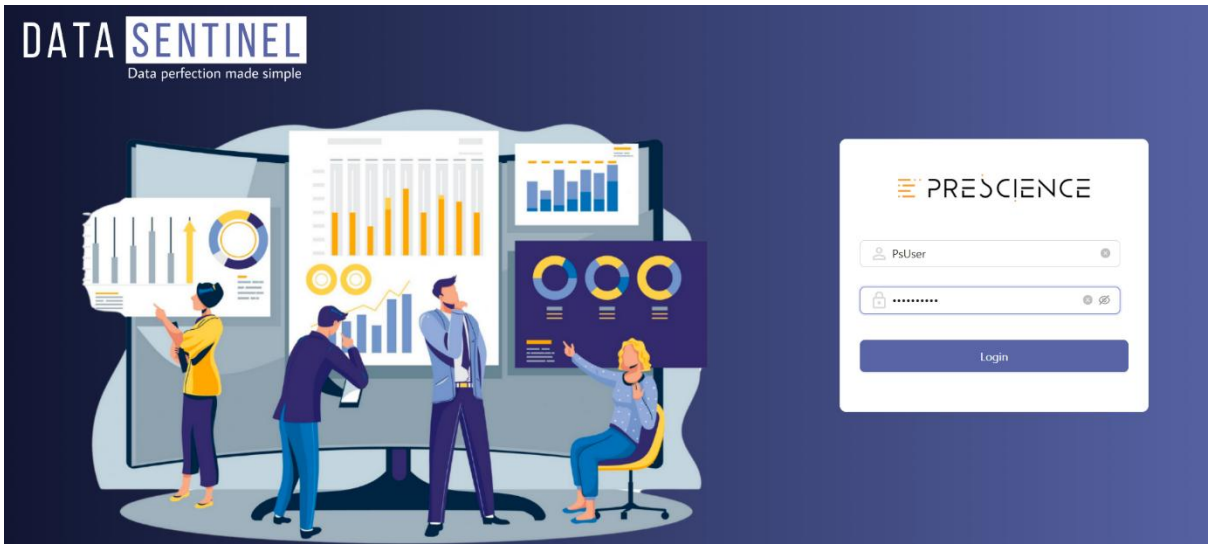

Image 9: Tenant user login page

Login using the tenant user(s) username and password.

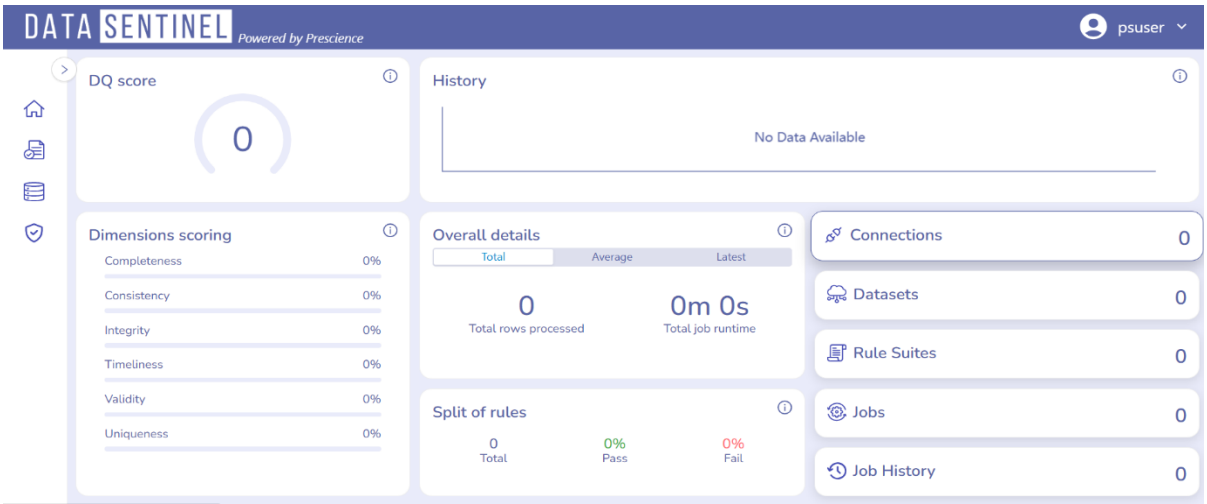

Image 10: Tenant user dashboard

2.2.3. Create New Connection

Go to the Connections tab and enter all the required details. The supported data sources are:

- SQL Databases (PostgreSQL, MS SQL, MySQL, Redshift, Snowflake, Databricks)
- AWS S3 Bucket
- Azure Blob Storage

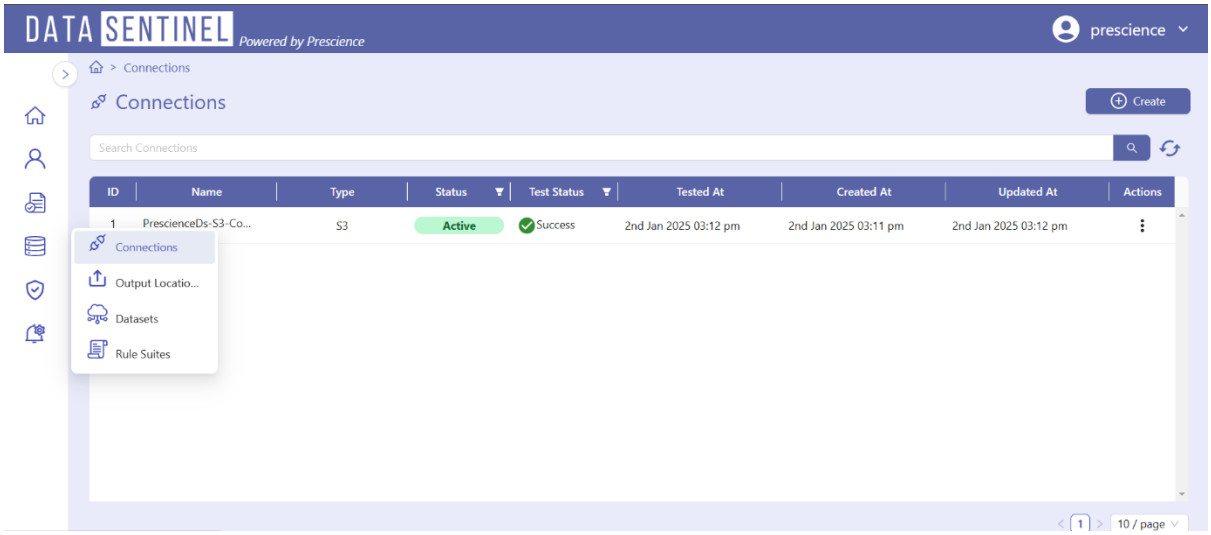

Image 11: Connection page

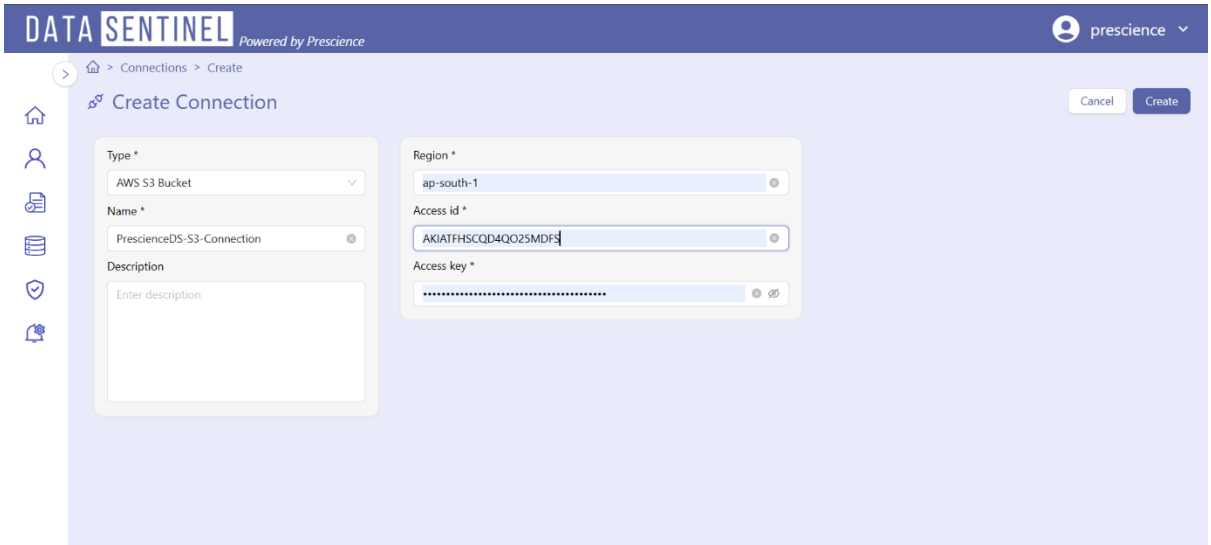

Image 12: Configuring connection details

Test Connection: After establishing the connection, click on the three dots located on the right side of the connection and select **Test** to validate the configurations provided in your connection settings.

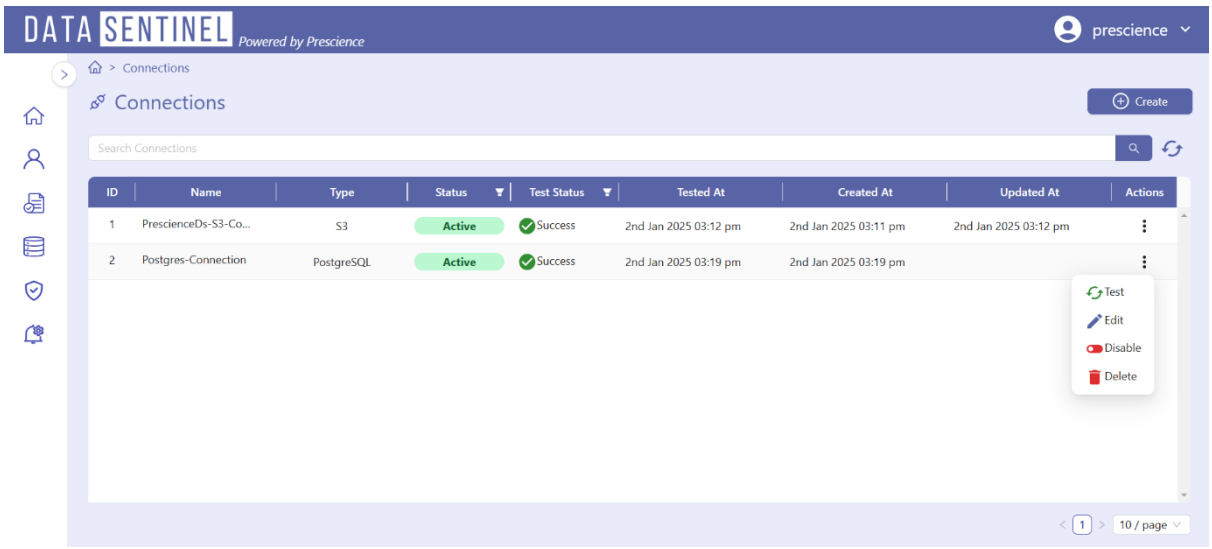

Image 13: Test connection

2.2.4. Create Output Location

1. Navigate to AWS Console.
2. Navigate to S3 service.
3. Create an AWS S3 bucket with a valid name for setting up the output location for storing the result generated by the job.

Go to the Output Locations tab and enter all the required details.

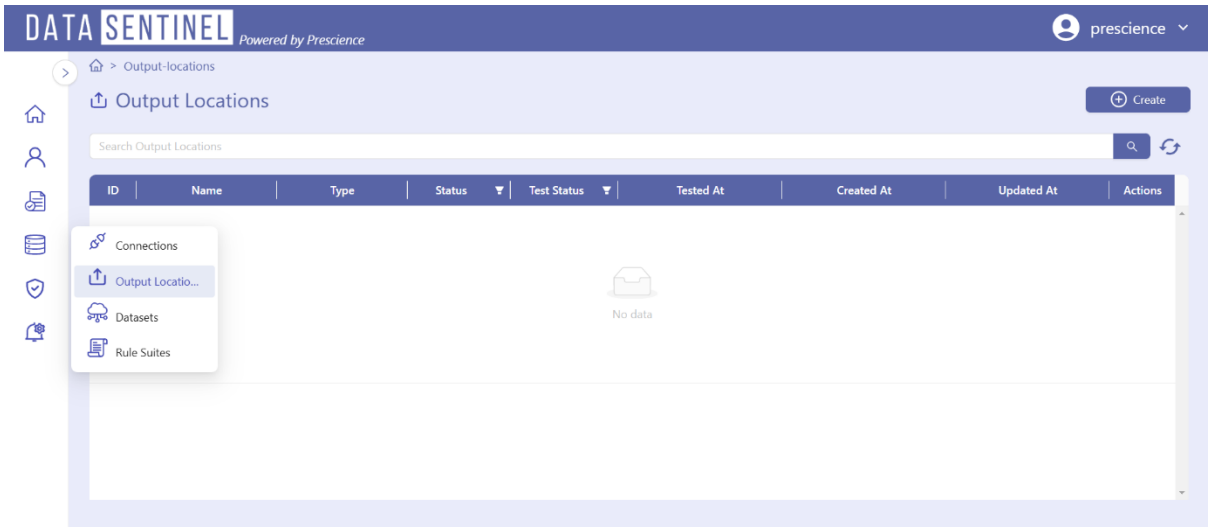

Image 14: Output location page

On the right side of the page, click the **Create** button. When creating the connection, it will attempt to connect to the specified bucket. If the credentials are correct, the connection will be established; otherwise, it will fail.

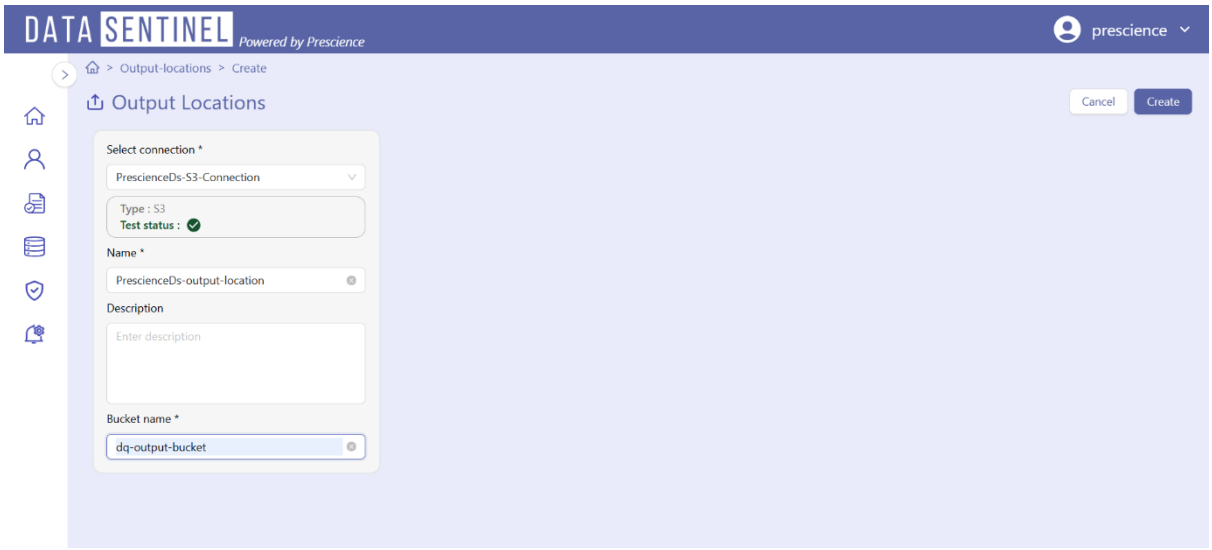

Image 15: Configuring output locations

After creating the output location, click on the three dots located on the right side of the output location and select **Test** to verify the configurations provided in your output location.

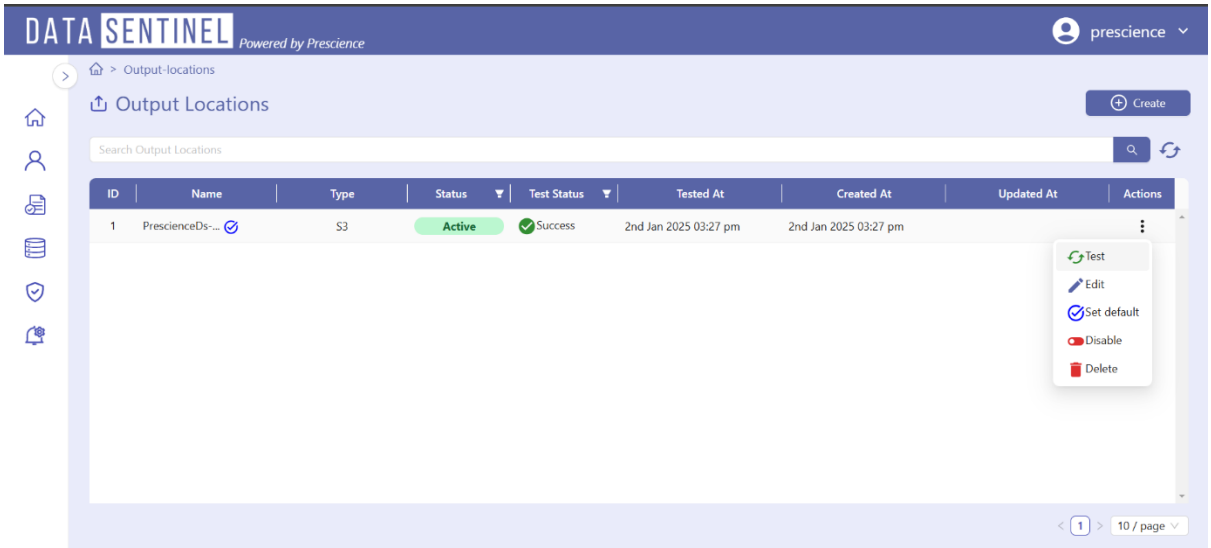

Image 16: Testing output location

2.2.5. Create Dataset

Go to the Create Dataset tab and enter all the required details.

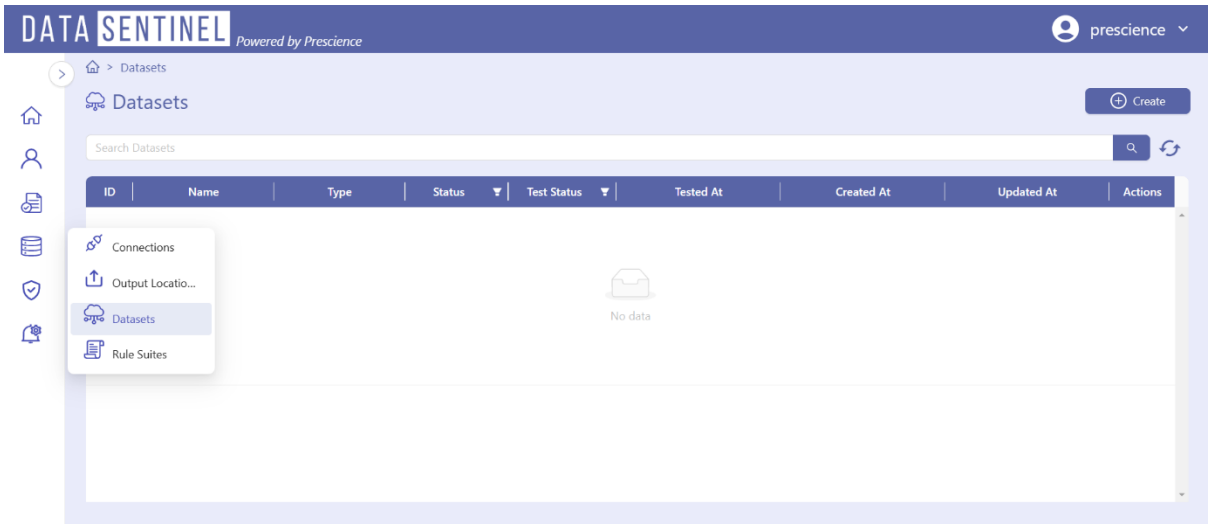

Image 17: Datasets page

Select an existing connection from the list to create a source dataset.

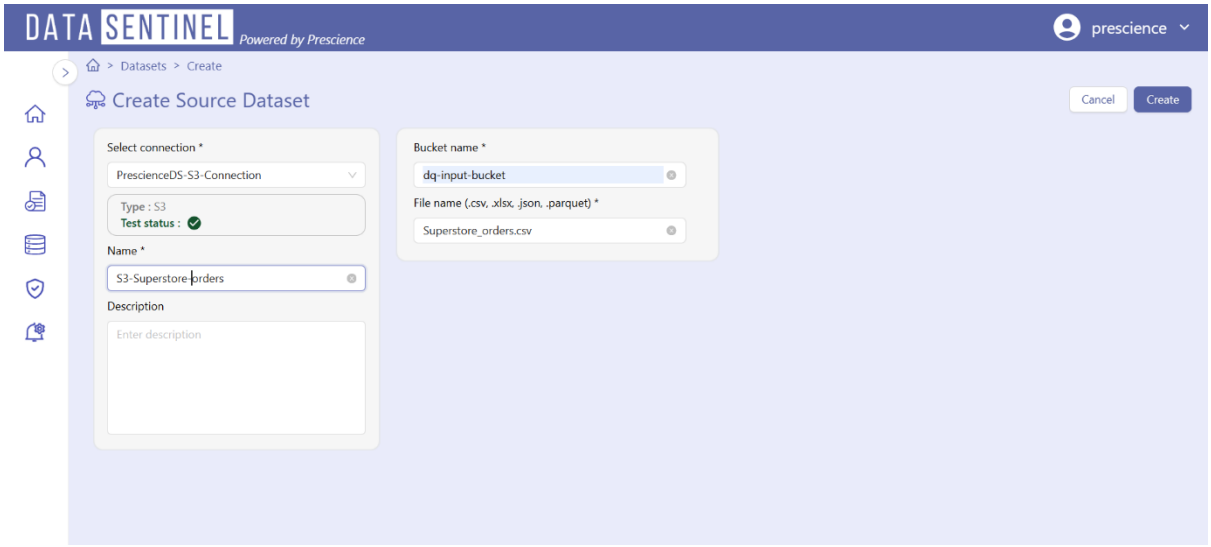

Image 18: Configuring datasets

If the existing connection is SQL database, customized SQL operations can be performed using the SQL Query option.

Note: Only **SELECT** SQL operations are allowed.

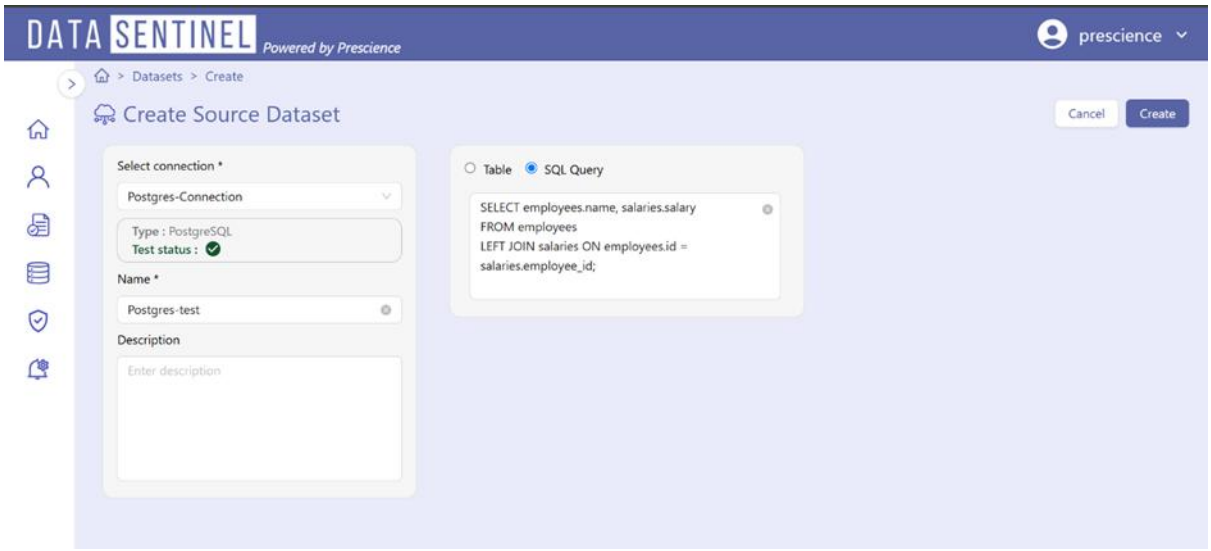

Image 19: Configuring datasets

Once the dataset is created, you can explore the data profile by clicking on the three dots in the "Actions" column. This will allow you to access the data profile by clicking on **Show Data Profile**, where you can examine various properties essential for data analysis.

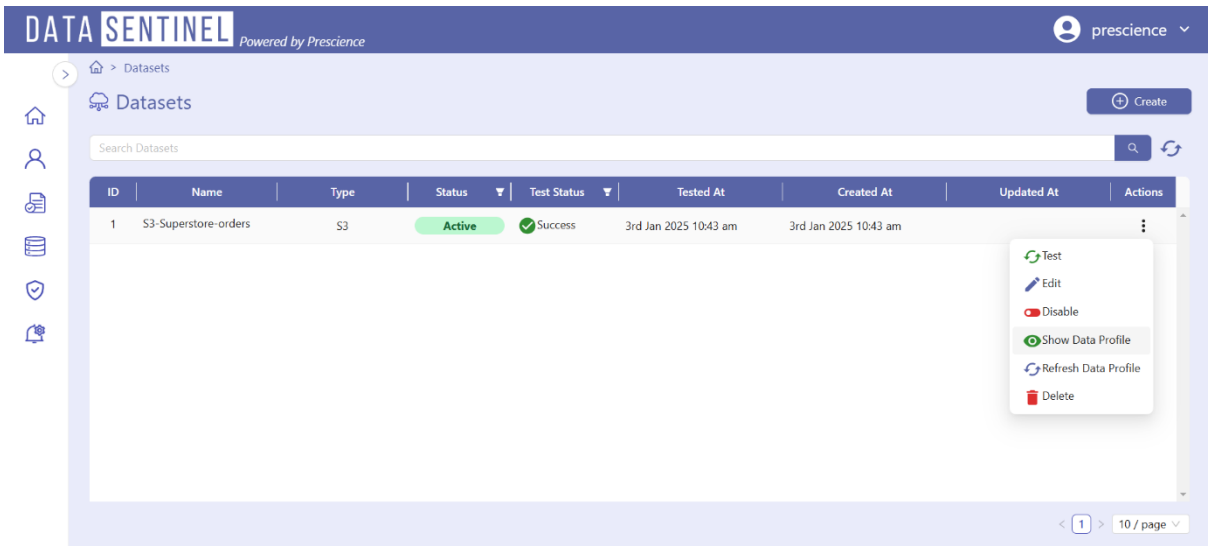

Image 20: Checking Data profile

Below are some of the properties you can review to better understand the dataset and facilitate analysis.

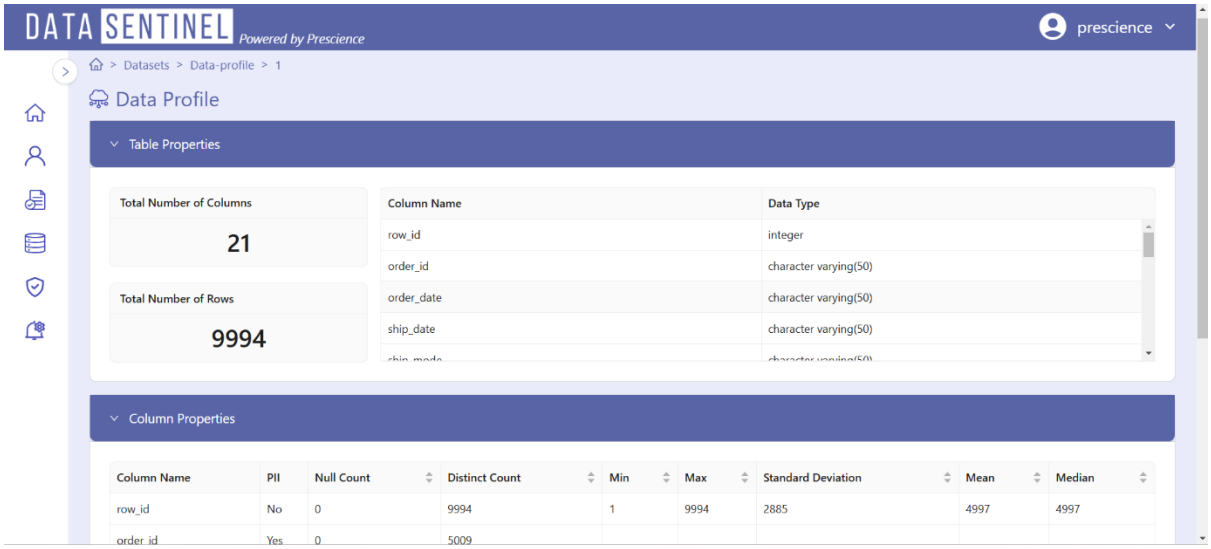

Image 21: Example data profile

If you make changes to the dataset manually, simply click **Refresh Data Profile** present under three dots in action column of the modified dataset to update profile data.

2.2.6. Create a Rule Suite

A rule suite is a combination of multiple rules in one suite. Create a rule suite from the given set of prebuilt rules.

Go to the **Rule Suites** tab and click the **Create** button to open a dropdown menu with options for **Data Quality** and **Data Transformation.** Select **Data Quality** to proceed.

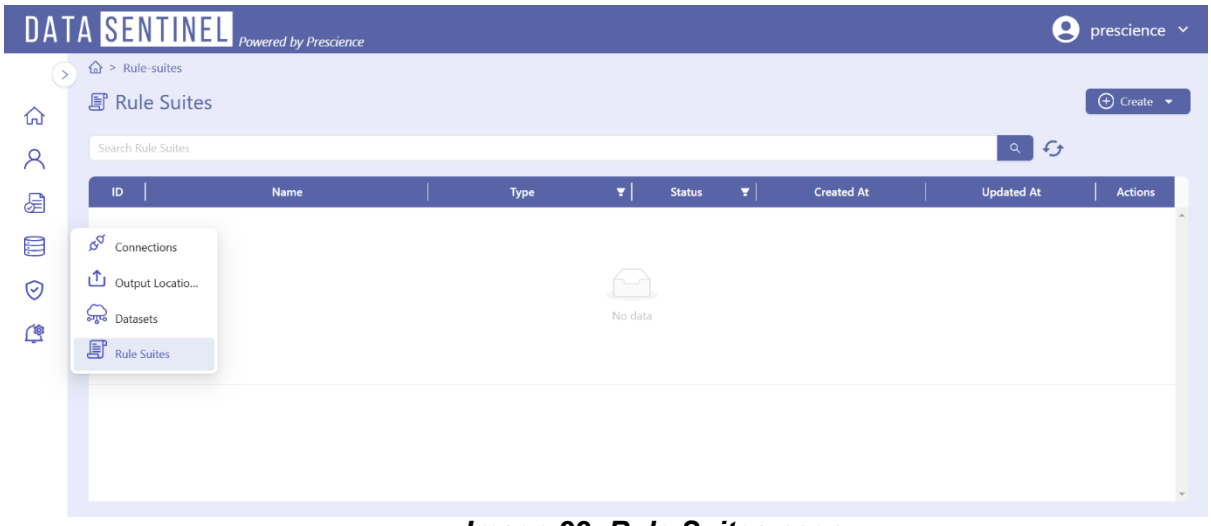

Image 22: Rule Suites page

Select the rules to be added for your dataset from the dropdown.

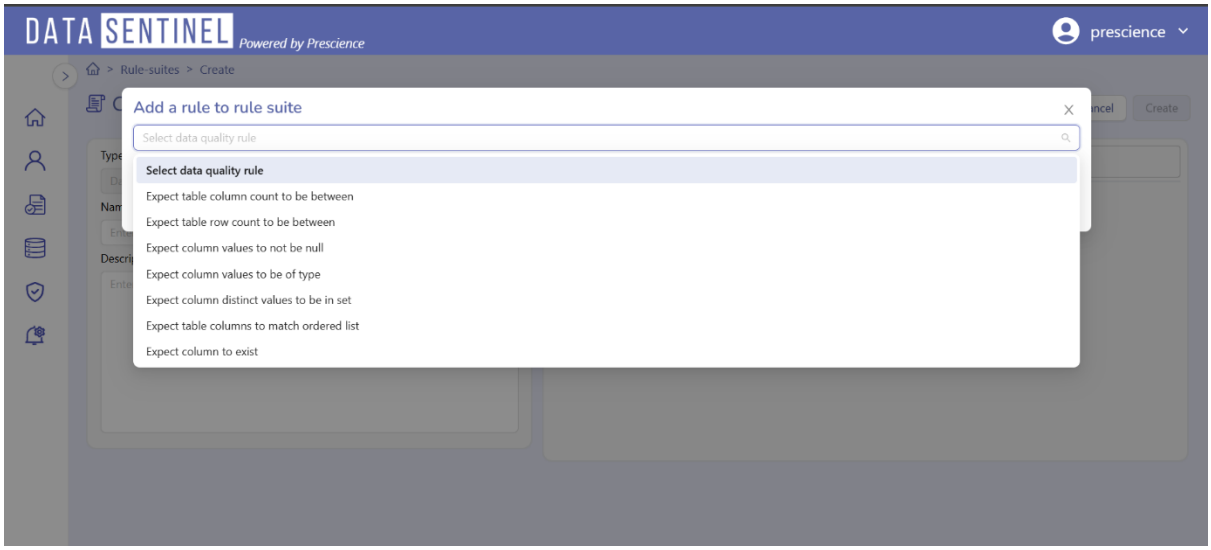

Image 23: Selecting rule from dropdown menu

Please make sure you replace **#SET_VALUE** with correct value in the rule JSON.
This rule suite can be created based on the user requirements for their dataset.

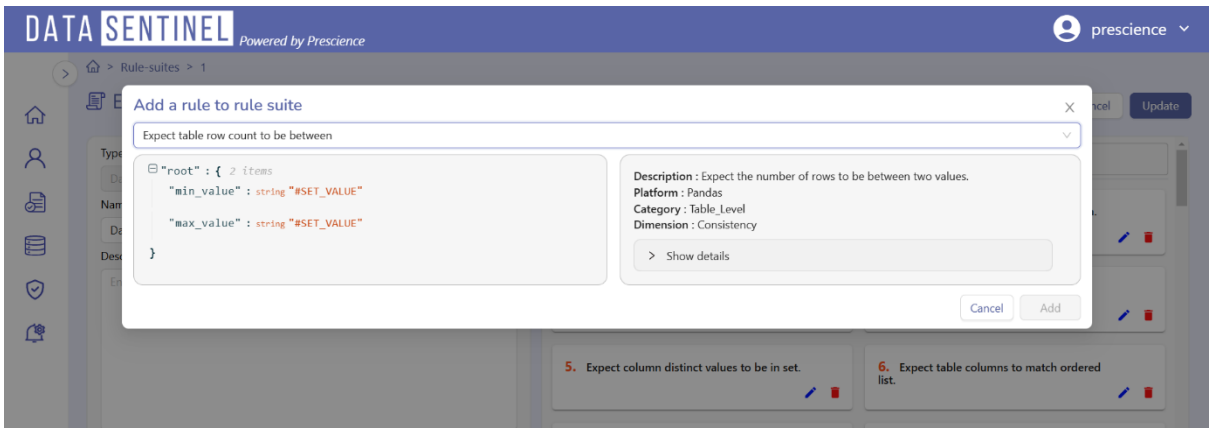

Image 24: Modifying values in the rule

Click the **Add** button to include a rule in your rule suite.

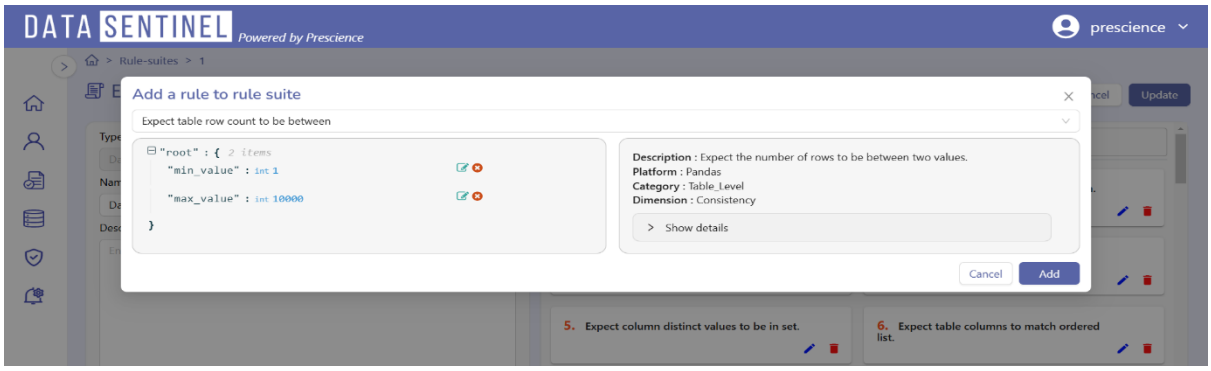

Image 25: Adding rule to rule suite

Click the **Create** button to save the rules to the rule suite list.

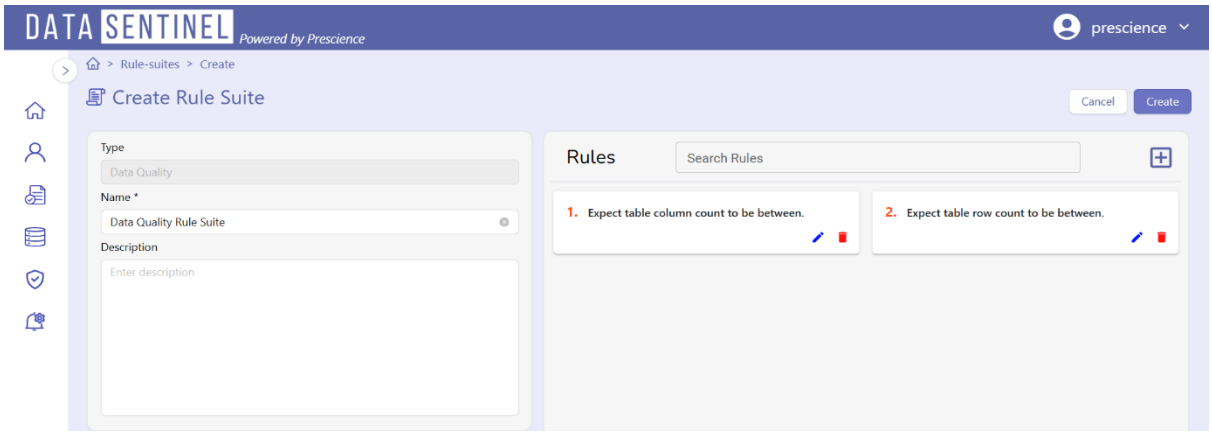

Image 26: Creating rule suite

Data quality rule suite is created.

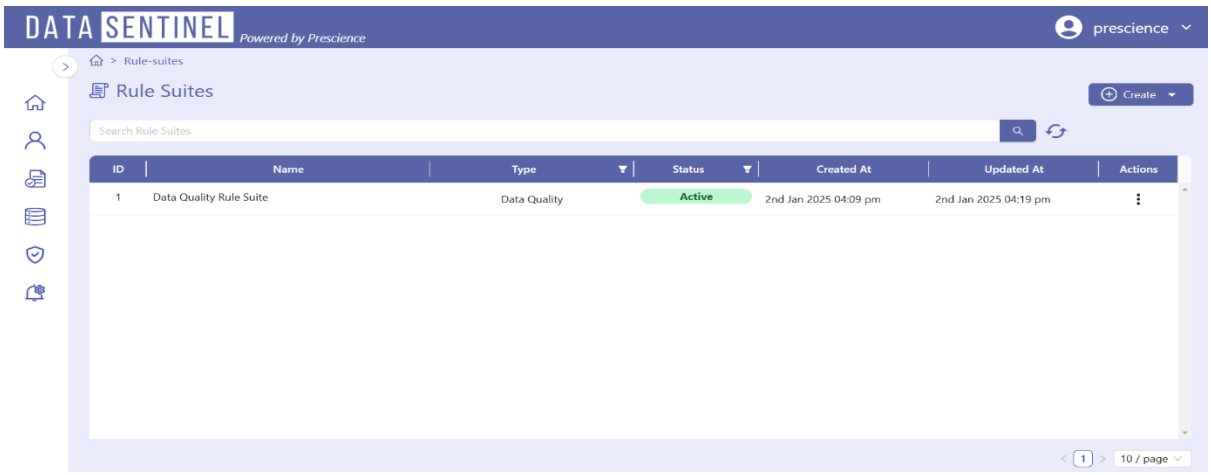

Image 27: Rule suites page

2.2.7. Create Data Quality Job

Go to the **Jobs** tab and click the **Create** button to open a dropdown menu with options for **Data Quality** and **Data Transformation.** Select **Data Quality** to proceed.

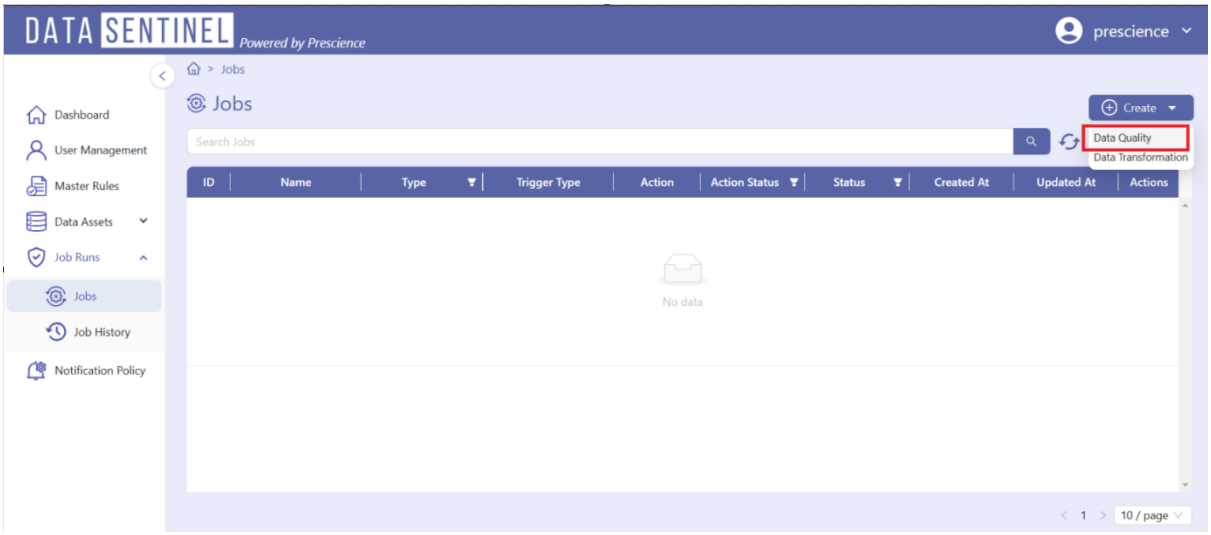

Image 28: Jobs page

Select a source dataset from the dropdown menu and enter the job properties, such as the job name, description, trigger type. Specify the trigger type whether the job will run manually or be scheduled.

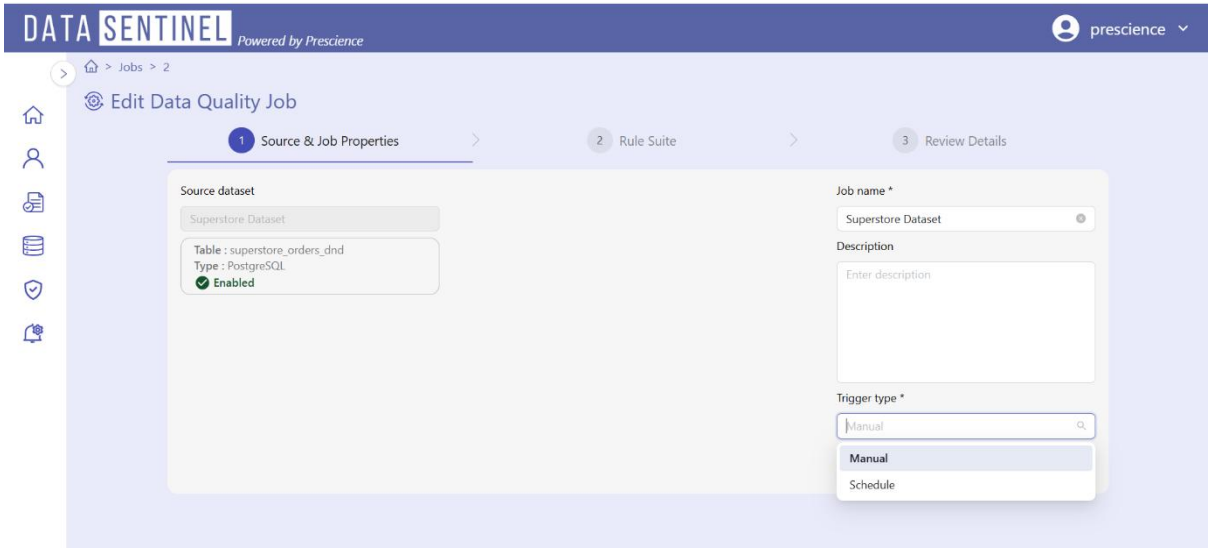

Image 29: Configuring new jobs in DQ

Manual - User needs to manually run the job

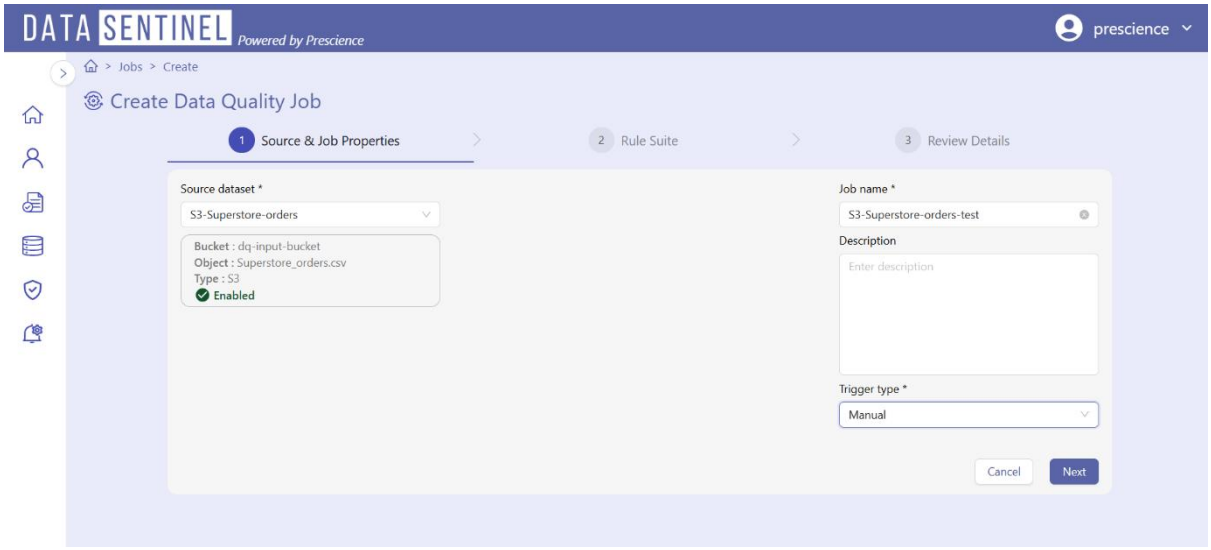

Image 30: Trigger type set to "Manual" in DQ

Schedule - User needs to schedule the job run for a certain time interval. An example is shown below.

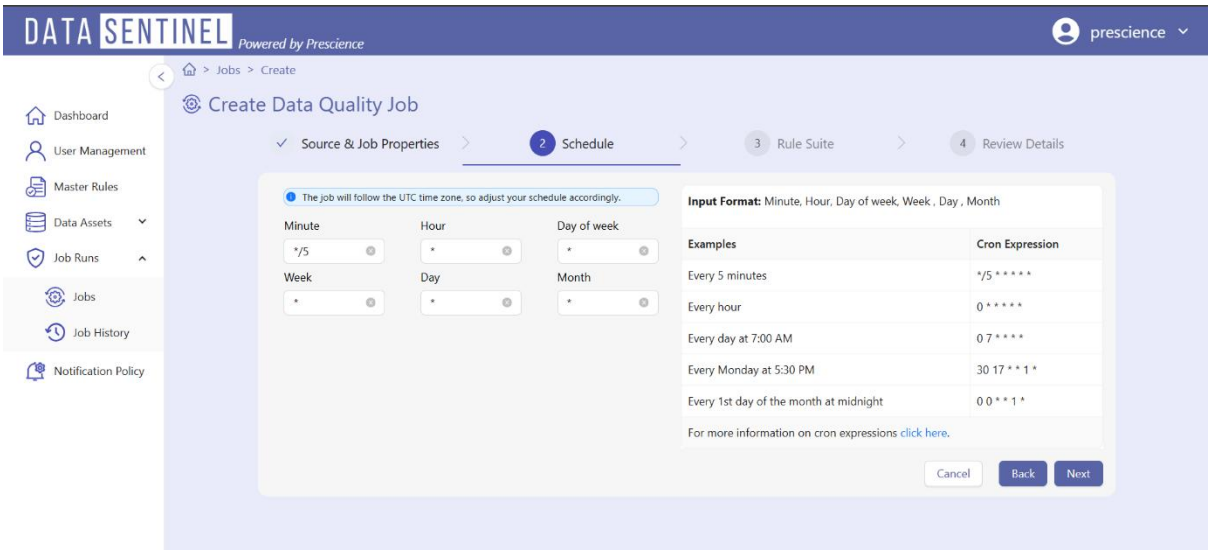

Image 31: Trigger type set to "Schedule" in DQ

Click Next to go to the next step.

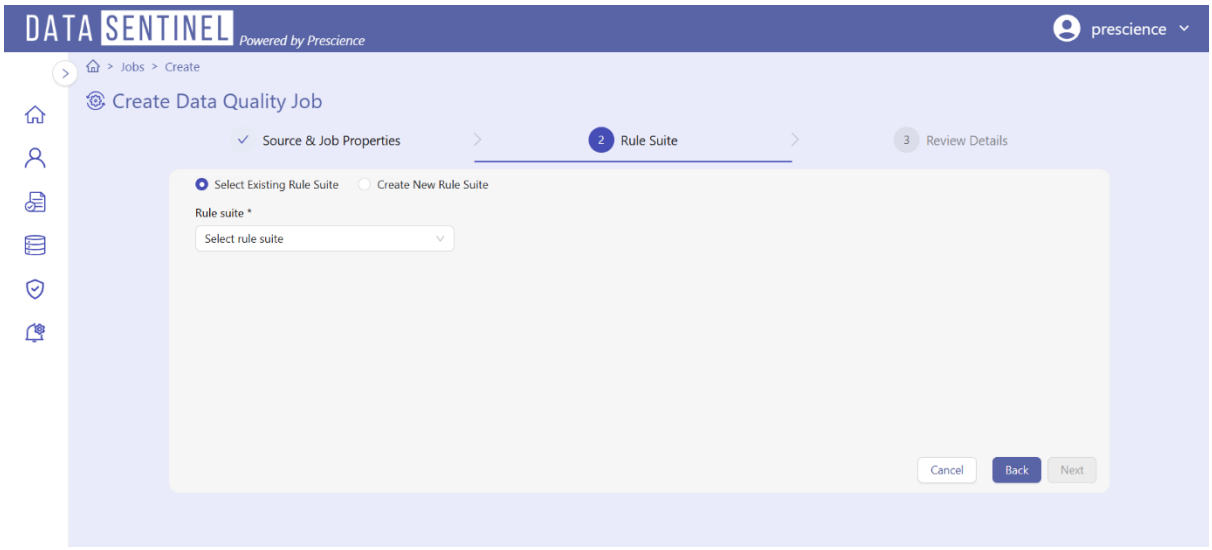

Image 32: Rule suite step in DQ

Rule suite - you can either select an existing rule suite or create a new one.

Existing Rule Suite

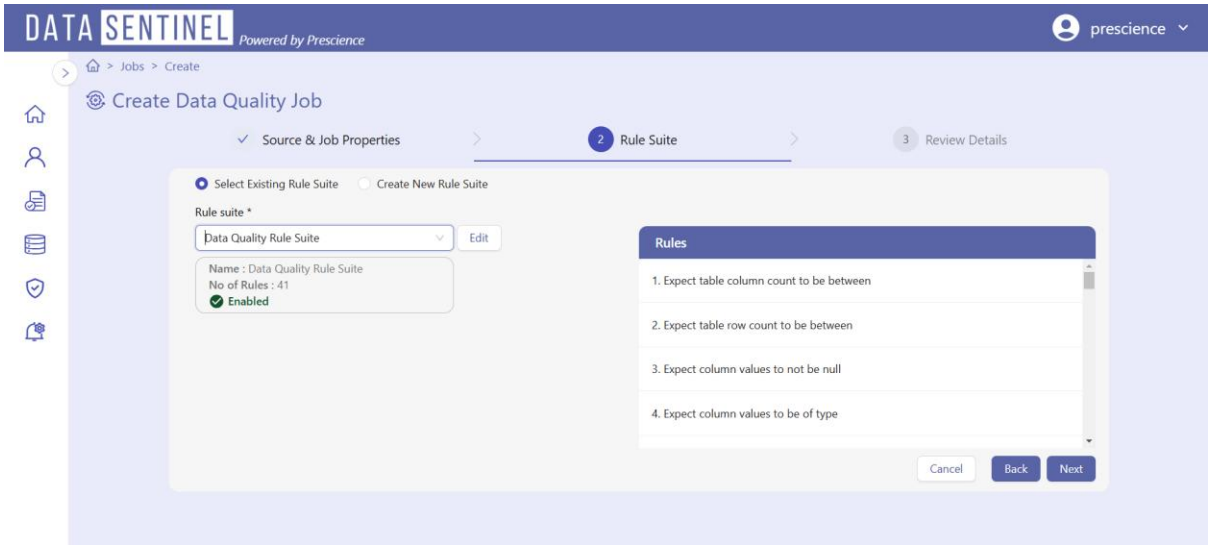

Image 33: Existing rule in DQ

Create new rule suite - Add the rules by clicking **+** button.

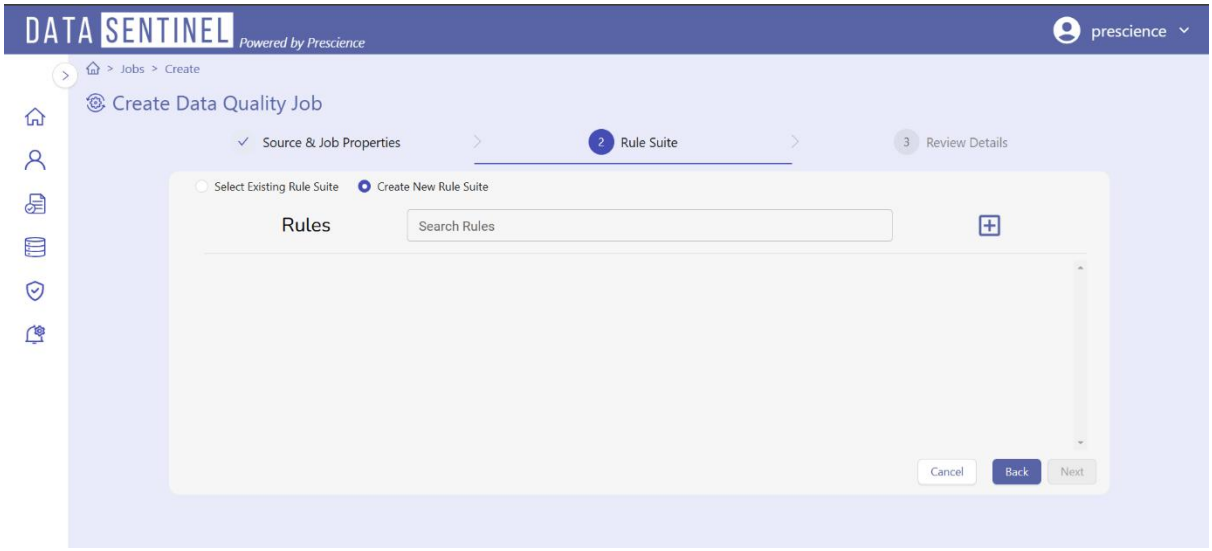

Image 34: Create rule suite in DQ

Select the data quality rules to be added for your dataset from the dropdown.

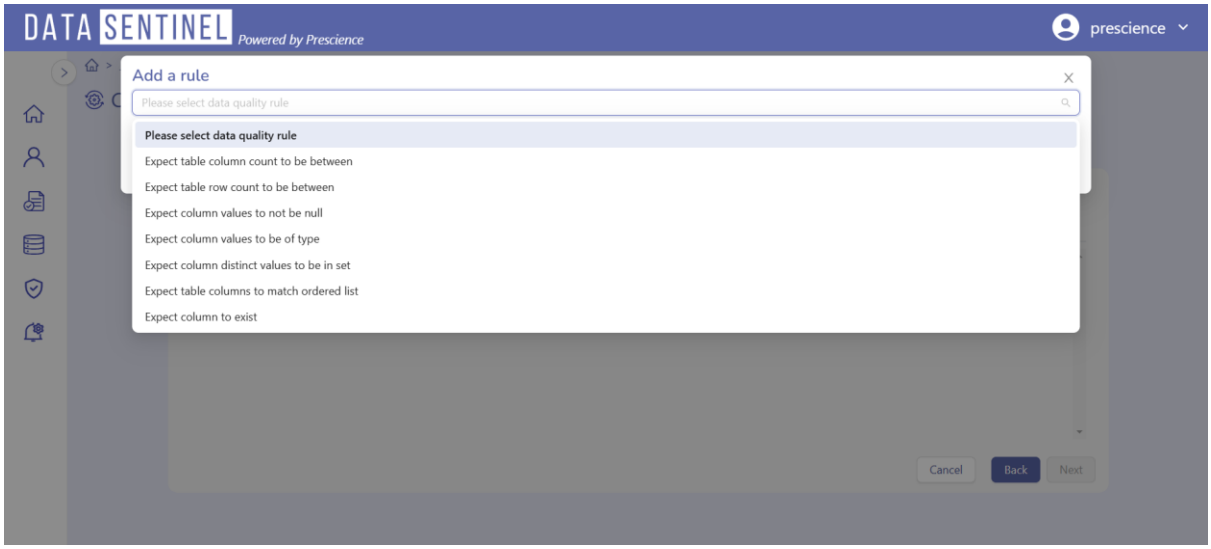

Image 35: Add a rule in DQ

Make sure you replace **#SET_VALUE** with column for which the rule needs to be applied.

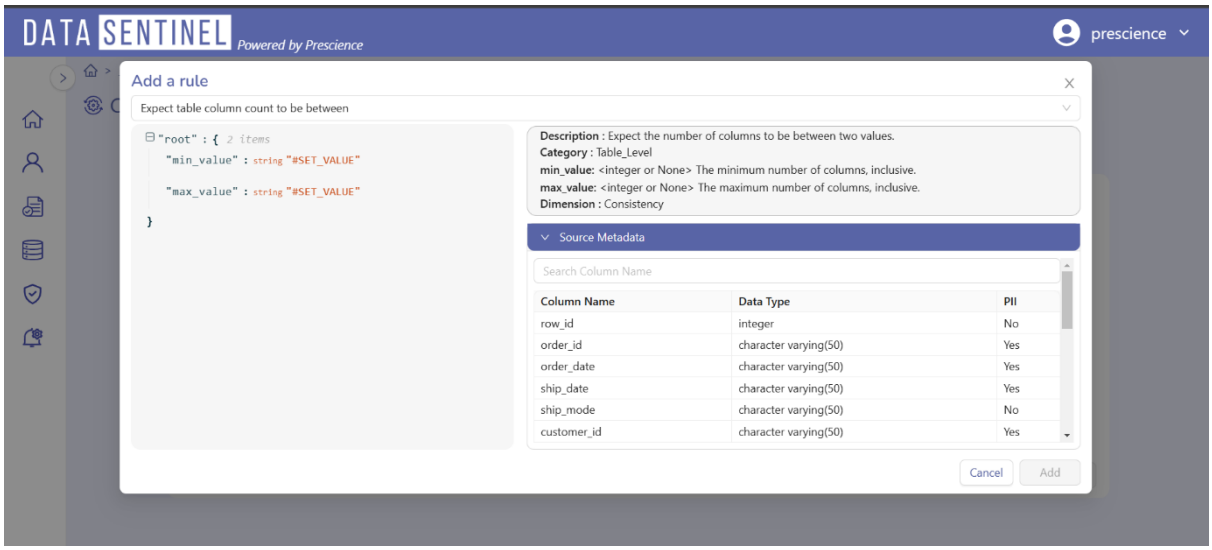

Image 36: Add rule details in DQ

The rule suite can be created based on the user requirements for the dataset.

After adding all rules. Select create button to save the rules.

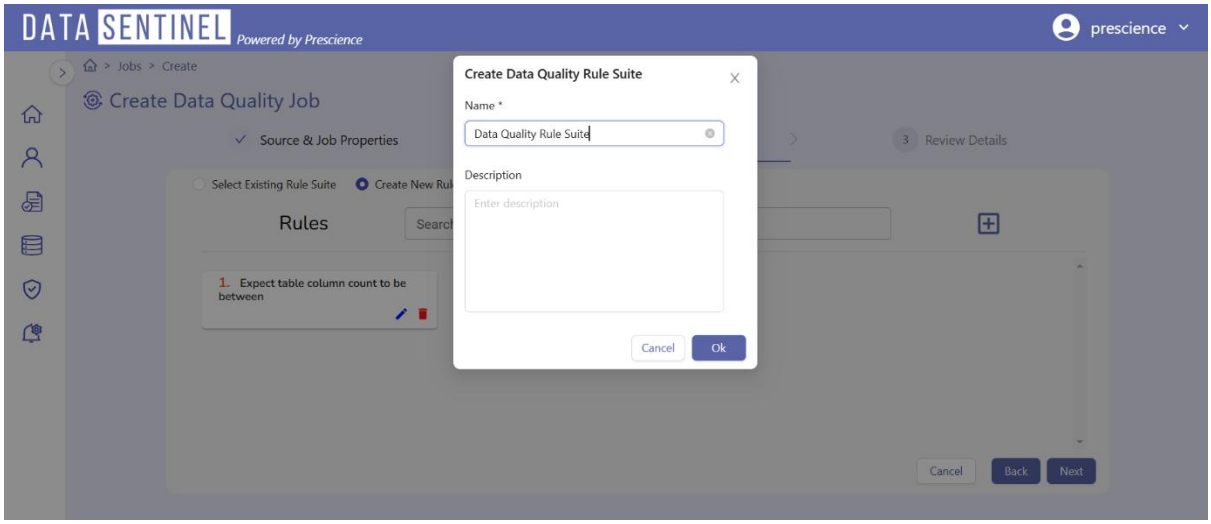

Image 37: Create data quality rule suite in DQ

Click the next button to proceed.

Now you can either select **Save** or **Save & Run** by selecting the submit button.

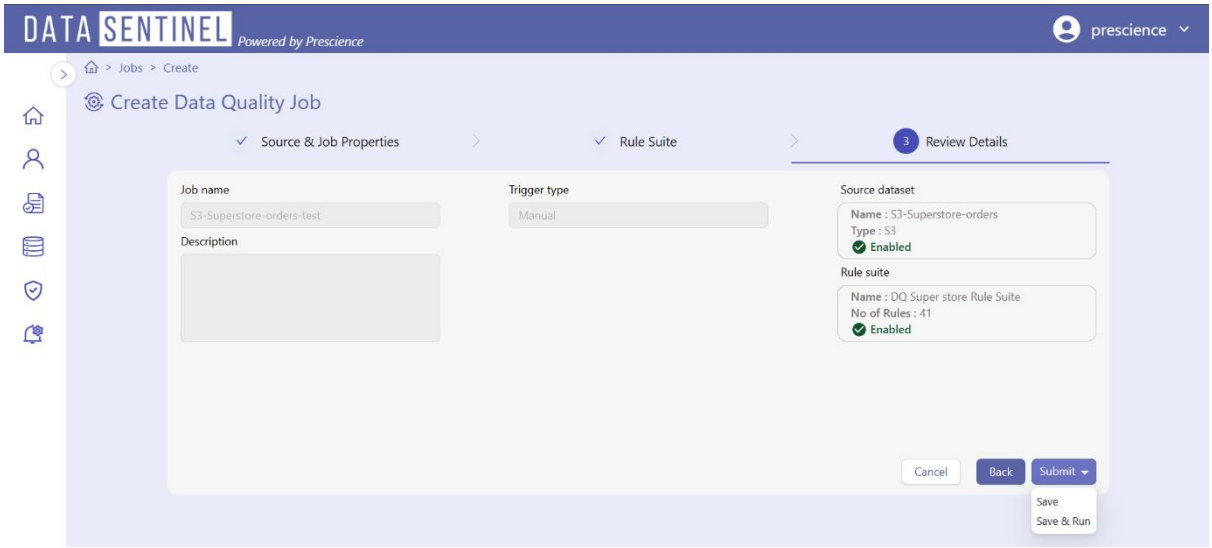

Image 38: Review details step in DQ

Save button clicked and Data Quality job will be created.

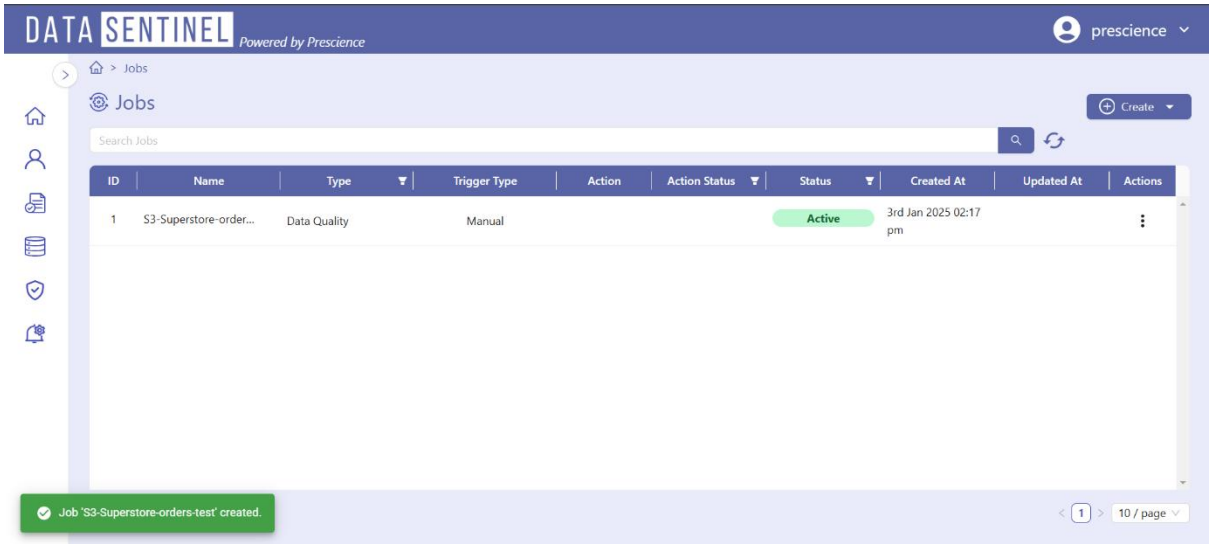

Image 39: Jobs page

To manually run the job, click the action icon on the selected job and choose **Run** from the dropdown menu.

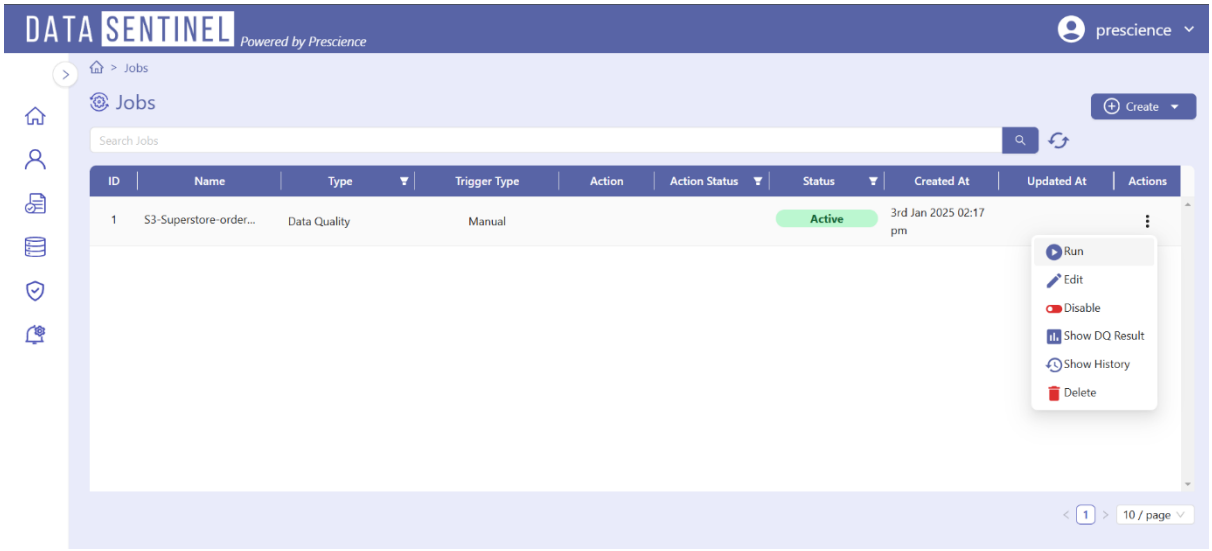

Image 40: Jobs page

Click on the **Show DQ Result** to view DQ Score.

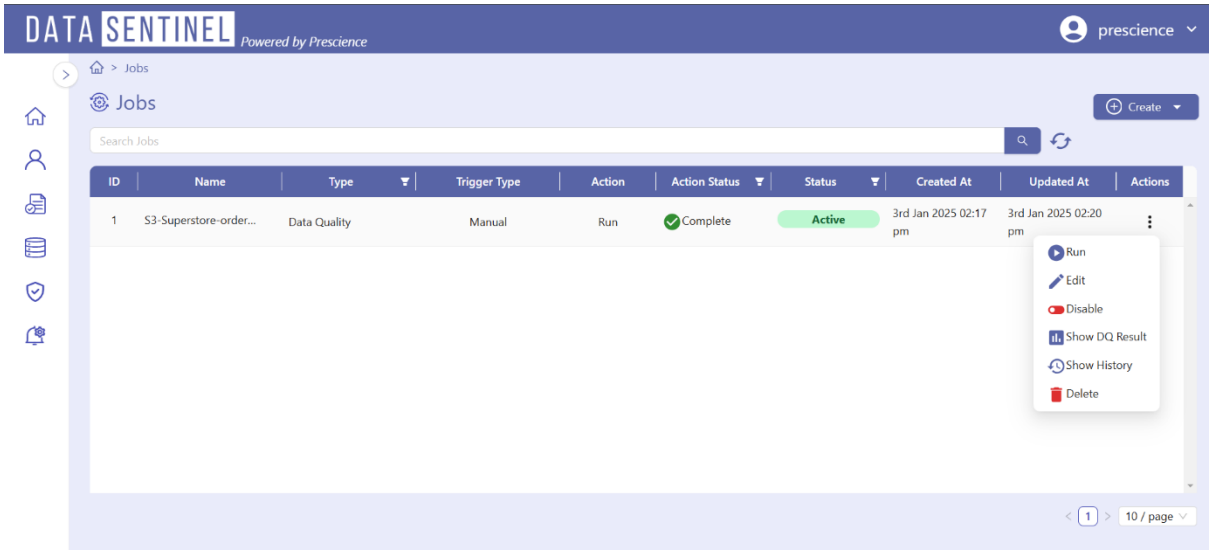

Image 41: Jobs page

Image 42: Show DQ result

Now click on the three dots and open **Show History**.

Now select show details to get a better understanding of where your data does not meet the required values.

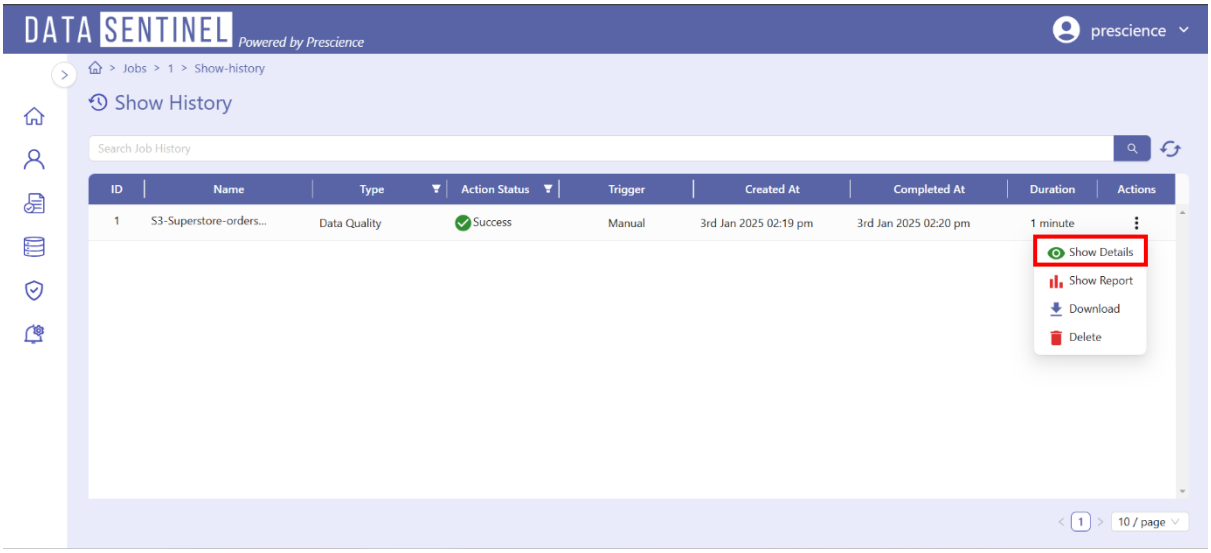

Image 43: Show History page

Go to the Jobs History tab to view all the required details of the existing jobs.

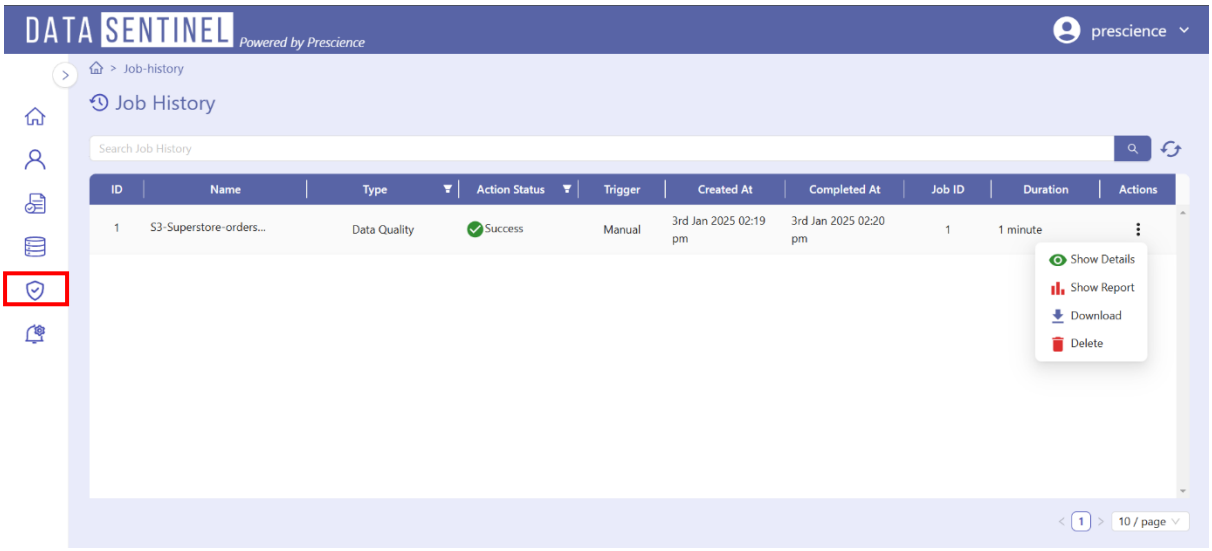

Image 44: Checking job history

Now you can explore various options to check the job run history and the data quality score for each run under the following options.

- **Show Details**
- **Show Report**
- **Download**

2.2.8. Create Data Transformation Job

Go to the **Jobs** section, click Select **Create** and choose **Data Transformation (DT)** option from the dropdown menu, as shown in the figure.

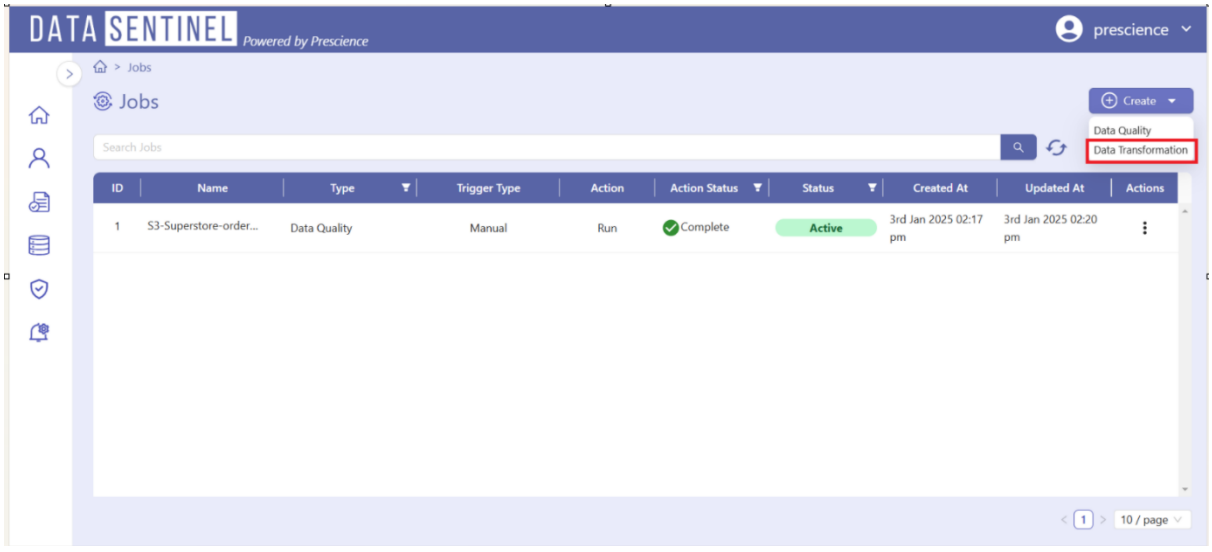

Image 45: Jobs page

Select a source dataset from the dropdown menu and enter the job properties, such as the job name, description, trigger type and continue on error. Specify the trigger type whether the job will run manually or be scheduled.

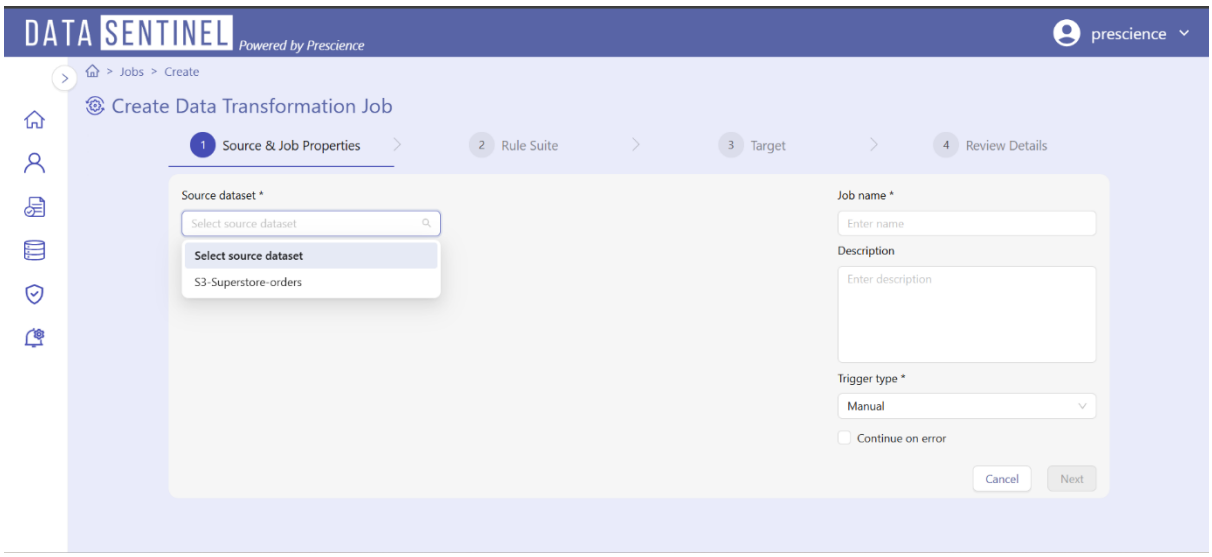

Image 46: DT source step

Continue on Error option ensures the job continues running, even if a rule fails. When selected, any failed rules will be skipped, allowing the remaining rules to execute successfully.

Trigger type dropdown provides two options: Manual and Schedule

Manual - User needs to manually run the job.

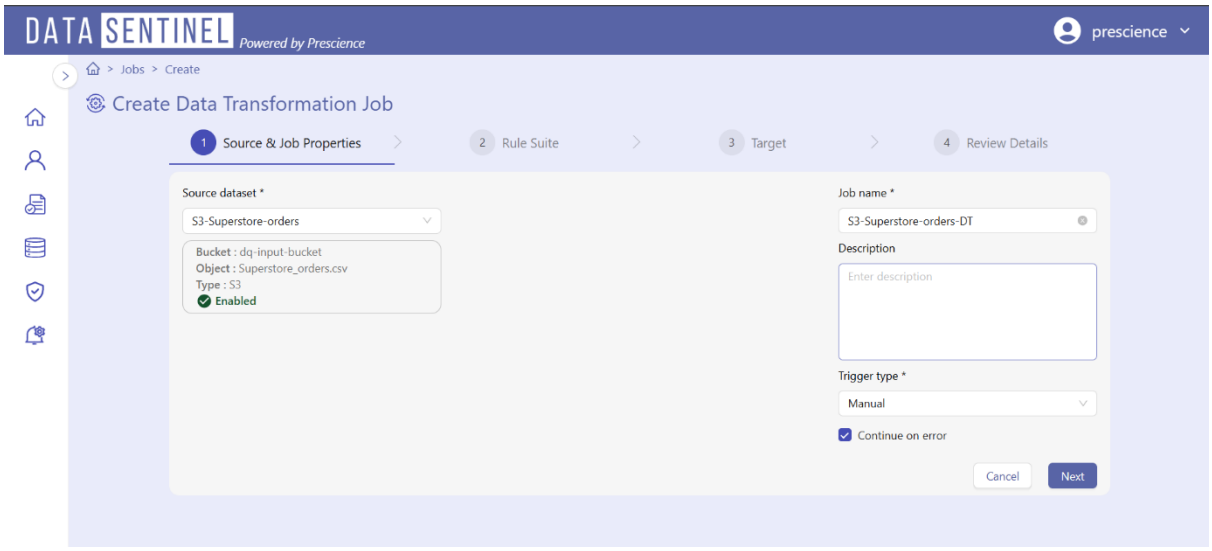

Image 47: Trigger type set to "Manual" for DT

Schedule - User needs to schedule the job run for a certain time interval. An example is shown below.

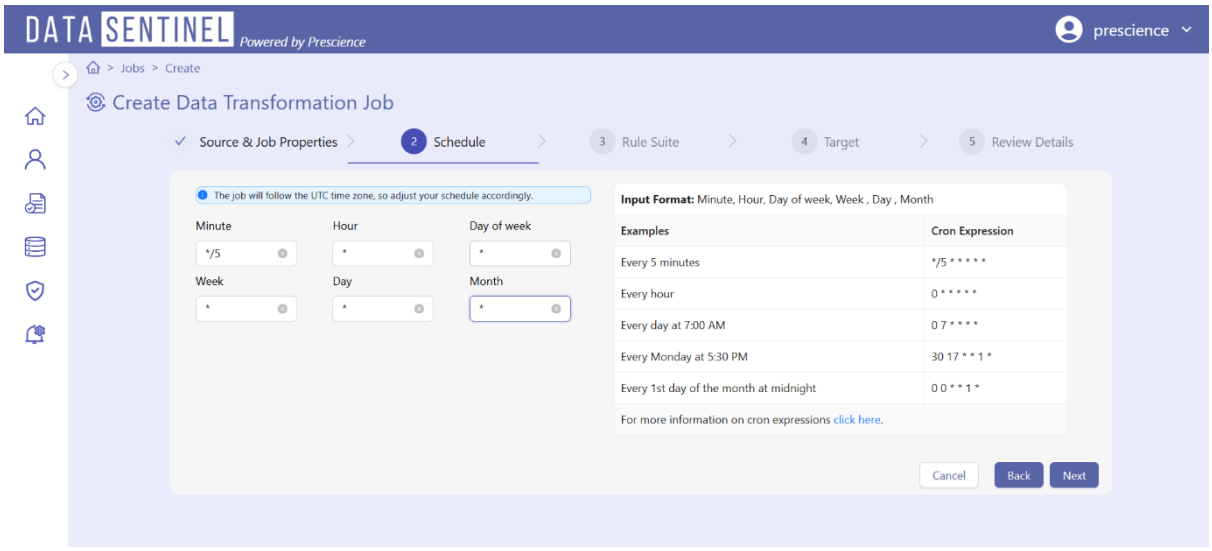

Image 48: Trigger type set to "Schedule" for DT

Click next to go to the next step.

Rule suite - you can either select an existing rule suite or create a new one.

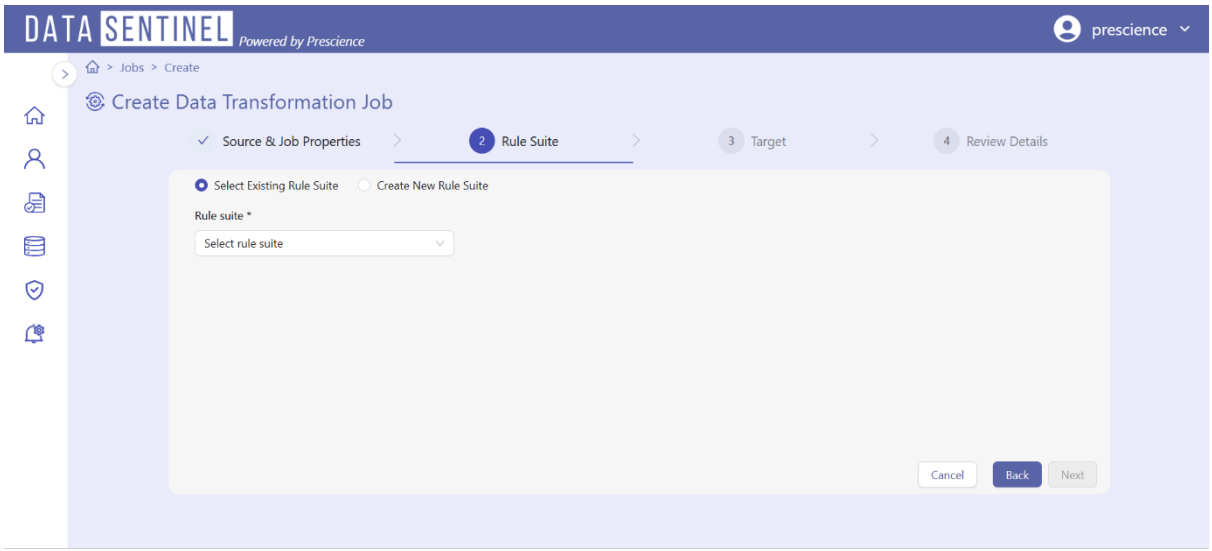

Image 49: Rule suite step in DT

Existing Rule Suite

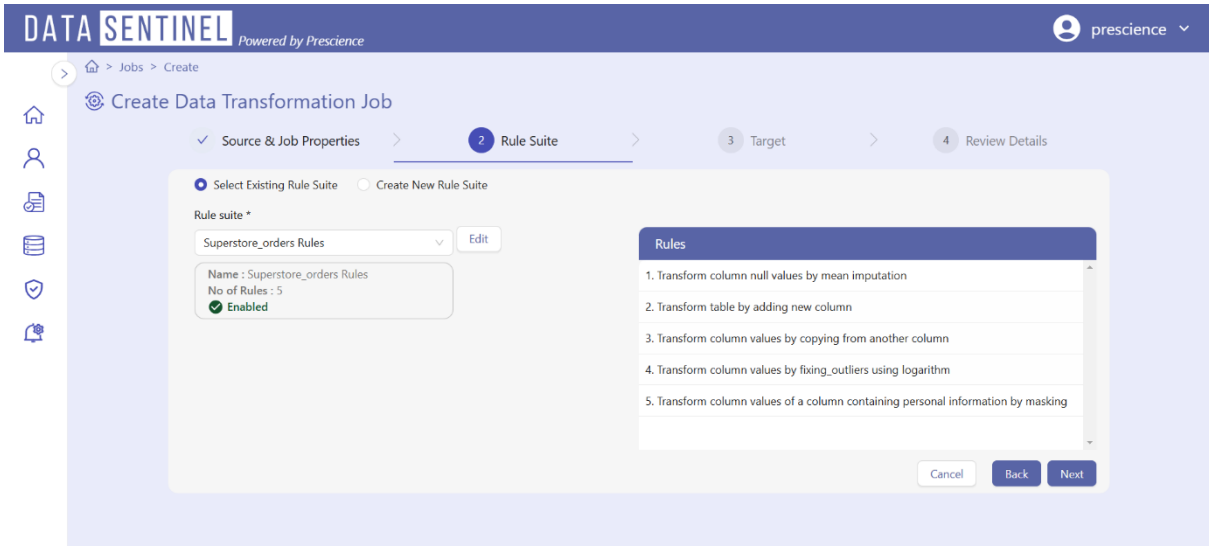

Image 50: Existing rule suite step in DT

Create new Rule suite - Add the rules by clicking **+** button.

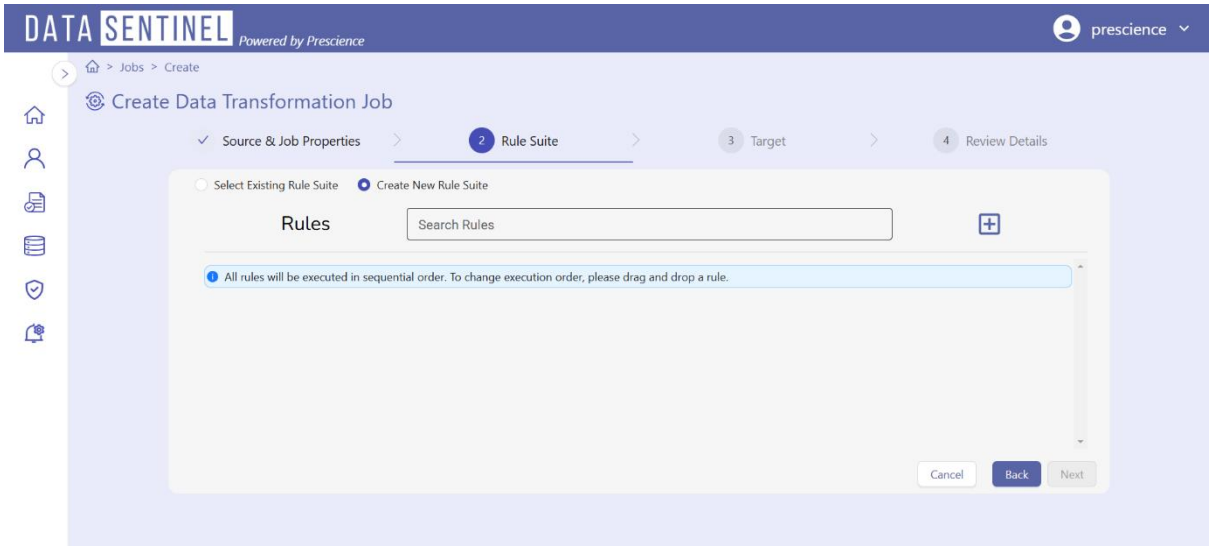

Image 51: Create new rule suite in DT

Select the data transformation rules to be added for your dataset from the dropdown.

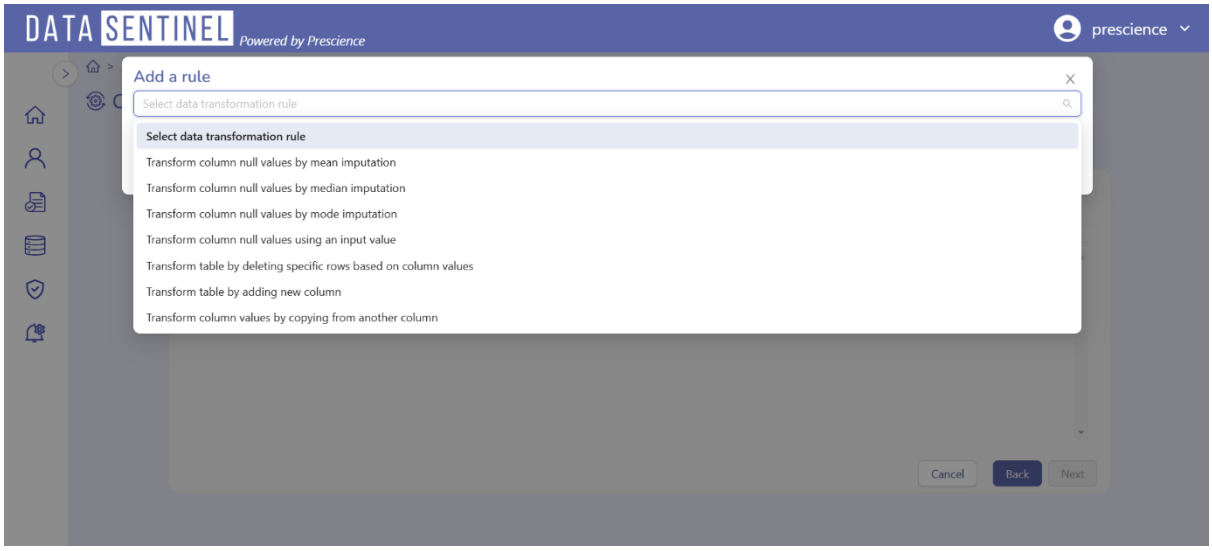

Image 52: Add a rule in DT

Please make sure you replace **#COLUMN_NAME** with column for which the rule needs to be applied.

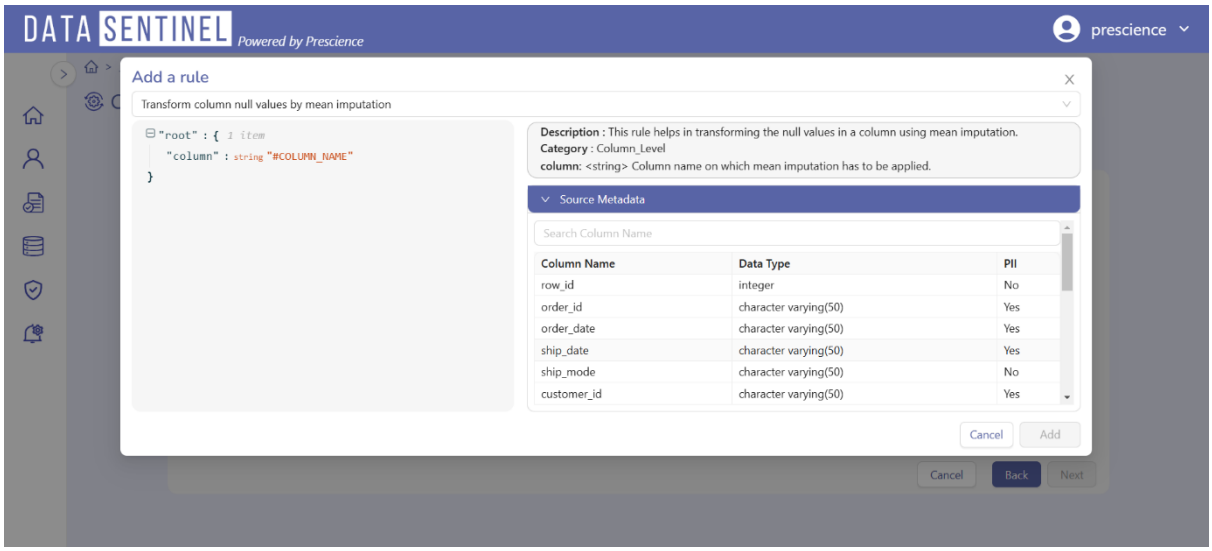

Image 53: Add a rule detail in DT

The rule suite can be created based on the user requirements for the dataset.

After adding all rules. Select Create button to save the rules.

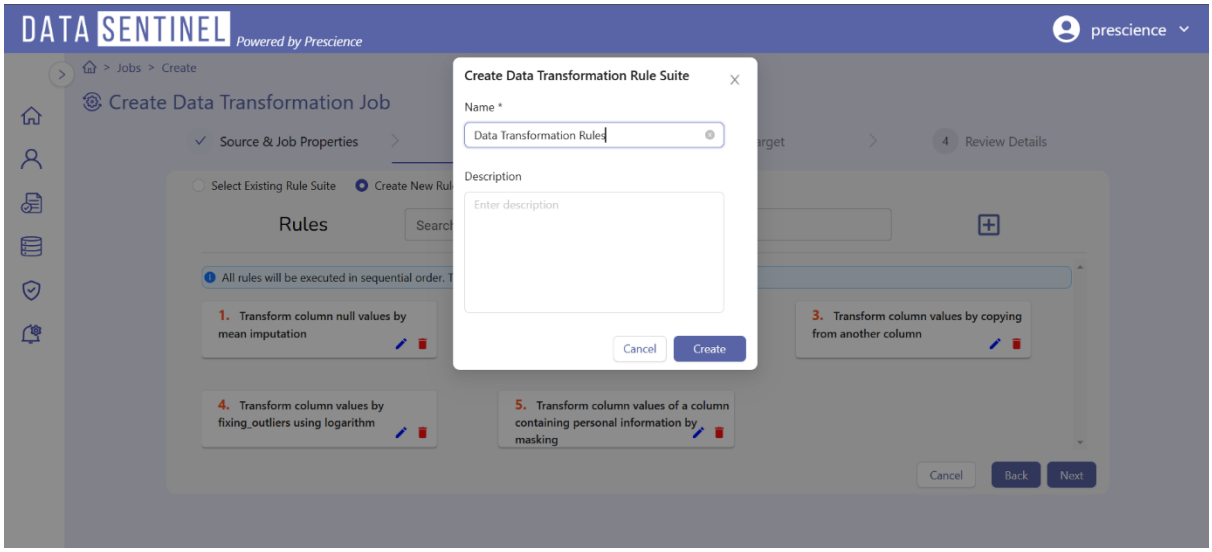

Image 54: Create DT rule suite

Select the target connection and target dataset details as shown in the figure.

Choose Compare source and target datasets option to show comparison of source and target datasets metadata in the results.

To enable custom mapping, select the appropriate option.

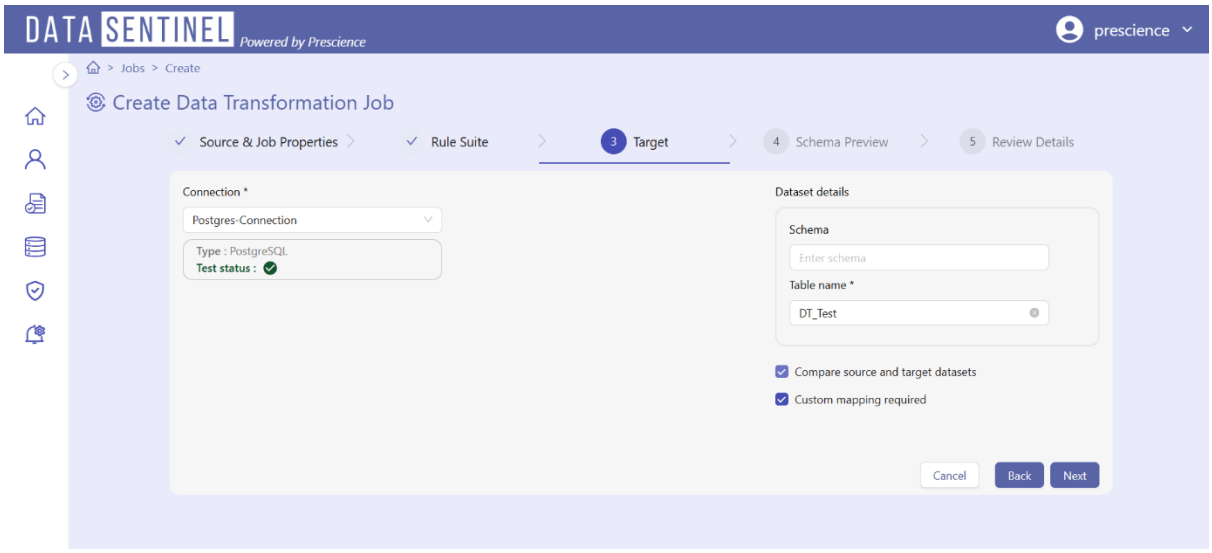

Image 55: Target step in DT

Custom mapping allows you to change the target dataset based on the rule suite. You can modify the data types as needed and proceed to the next step.

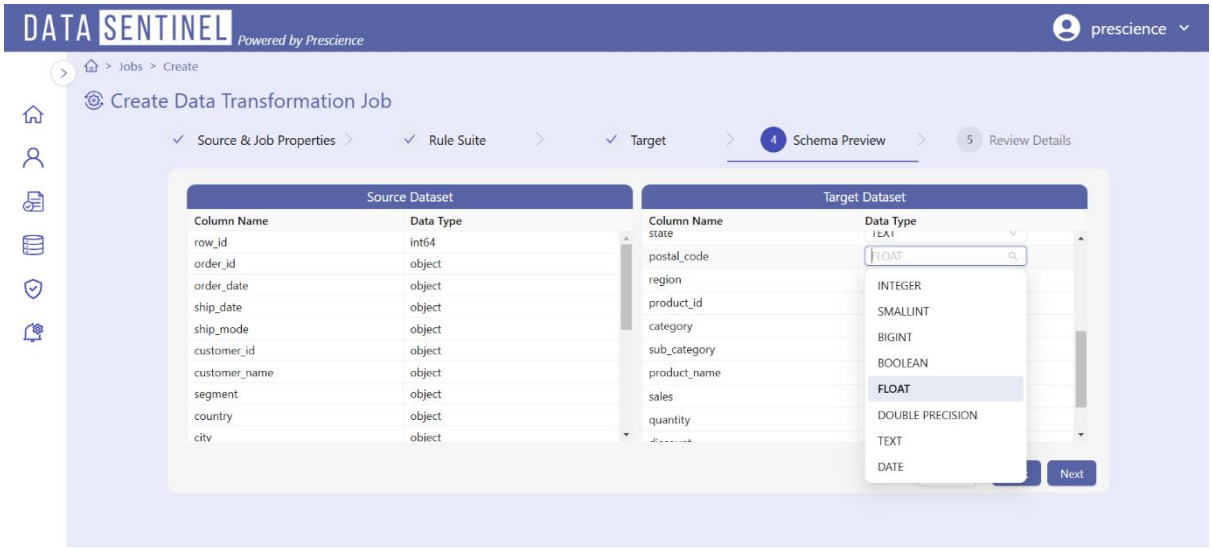

Image 56: Schema Preview step in DT

Now you can either select Save or **Save & Run** by selecting the Submit button.

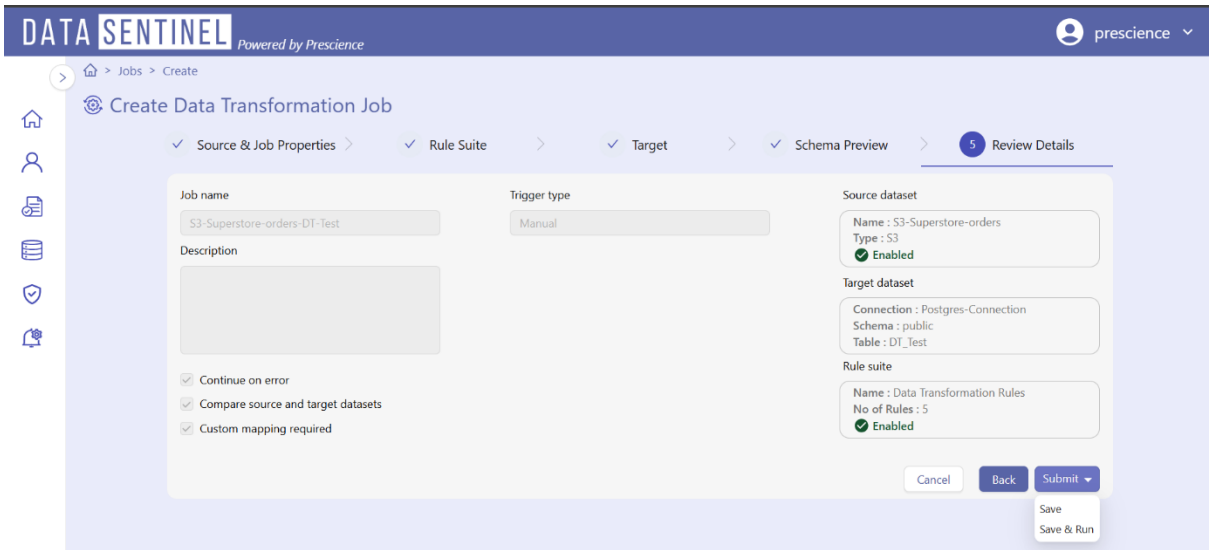

Image 57: Review details step in DT

When you click the Save button, job will be created.

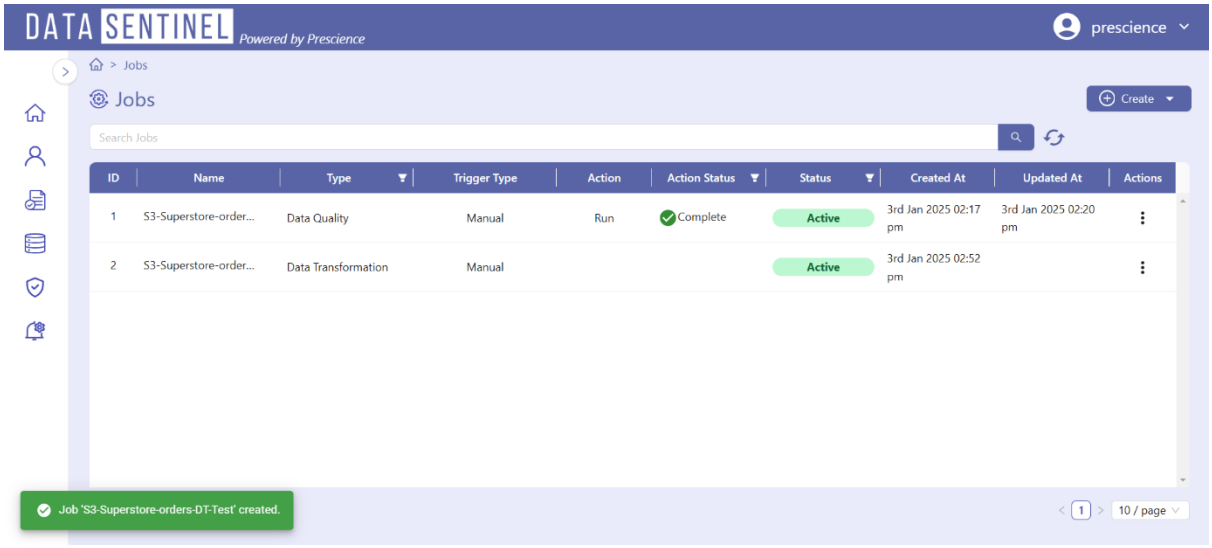

Image 58: Jobs page

To manually run the job, click the action icon on the selected job and choose **Run** from the dropdown menu.

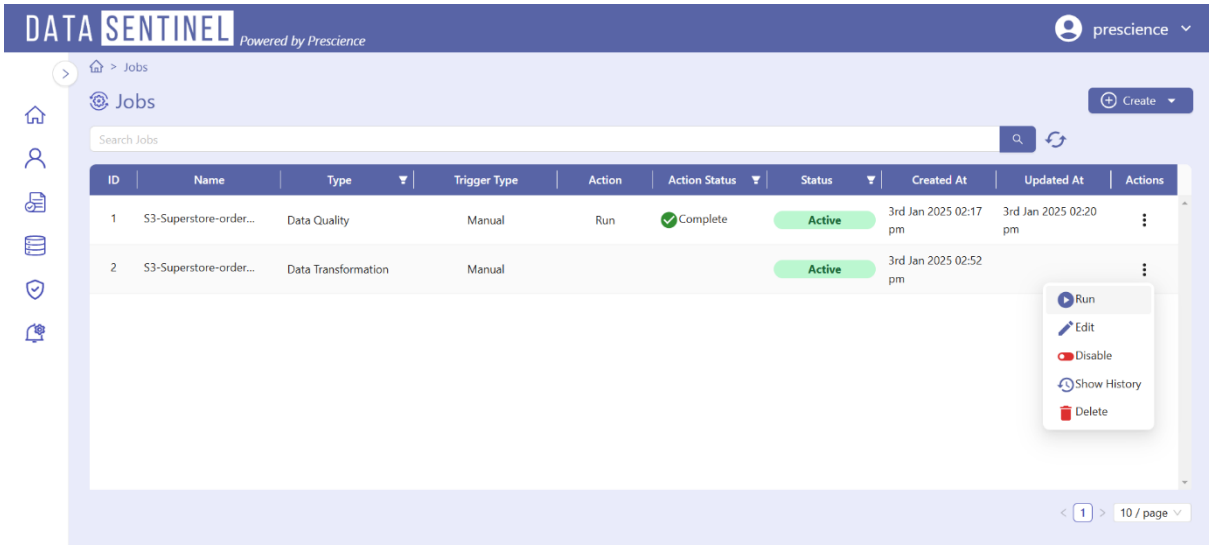

Image 59: Jobs page

Click the Show History icon from the dropdown menu to go to the history page. Then, select an action and click the **Show Report** option.

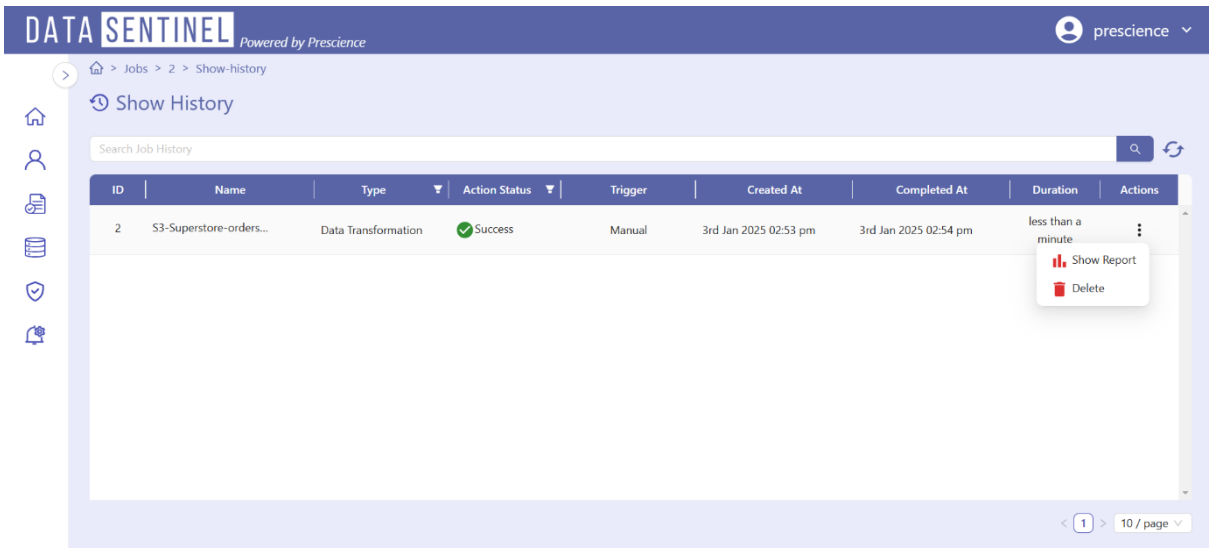

Image 60: Show History page

The Data Transformation results include source details, rule details, target details, and a comparison of the source and target datasets.

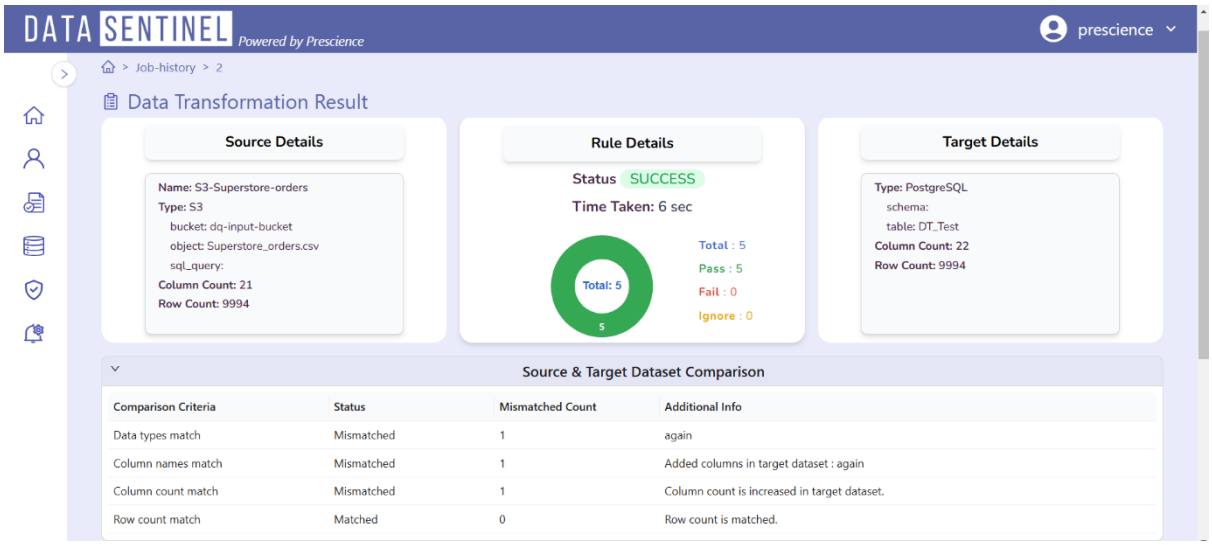

Image 61: Data transformation result

Rule Details contain the Source Column Name, Transformed Column name, Scanned Rows, Fixed Rows.

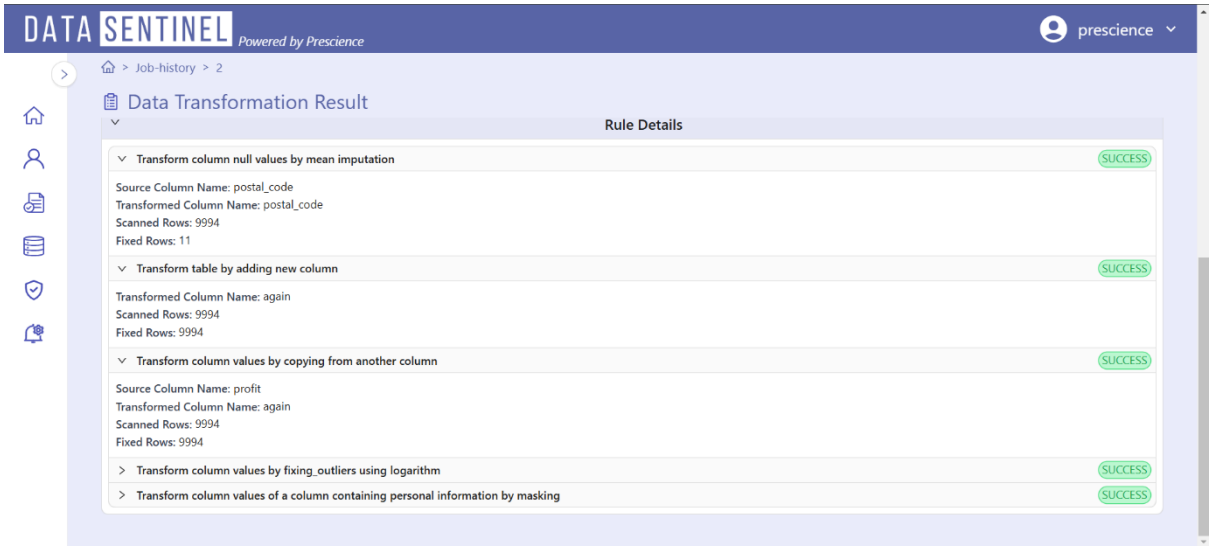

Image 62: Data transformation result

3. Setup Notification Policy

In the Super Admin settings page, there are two main configuration sections that need to be set up to enable notification for DQ score: SMTP Configuration (JSON), RabbitMQ Configuration (JSON).

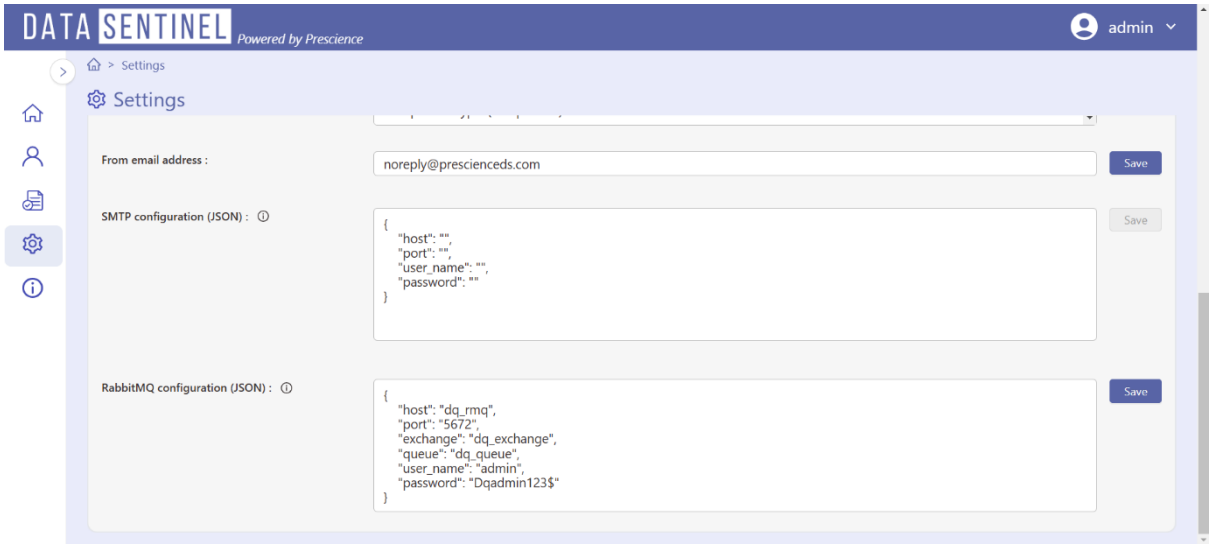

Image 63: SMTP and RabbitMQ configuration in settings page

SMTP and RabbitMQ Configuration Guide:

SMTP Configuration

To configure SMTP, follow these steps:

1. Navigate to the **Super Admin** page and go to **Settings**.
2. Add the following JSON configuration and **Save**:

```
{
  "host": "<SMTP server IP>",
  "port": <SMTP server port>,
  "user_name": "<SMTP user name>",
  "password": "<SMTP user password>"
}
```

RabbitMQ (RBMQ) Configuration

To configure RabbitMQ, follow these steps:

1. Navigate to the **Super Admin** page of your Data Sentinel instance and go to **Settings**.
2. **Save** the prepopulated RabbitMQ configuration JSON.

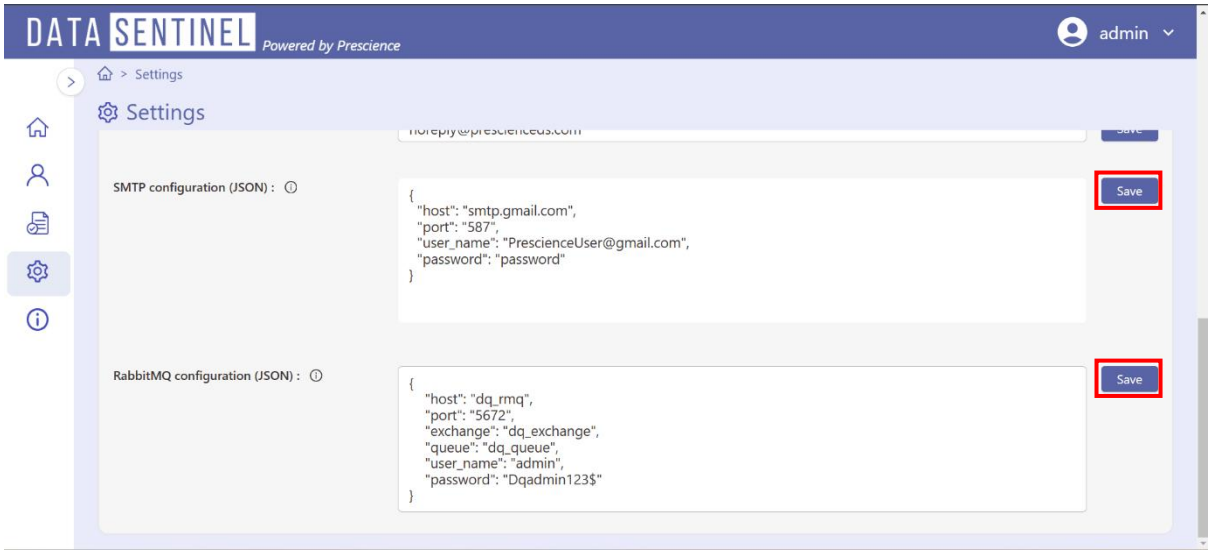

Image 64: SMTP & RabbitMQ configuration in settings page

On the Tenant Admin page, go to **Notification Policy** and fill in the details, including Name, Description, Condition Name, Comparison Type, Threshold Value, Email, and Webhook as shown in the figure below.

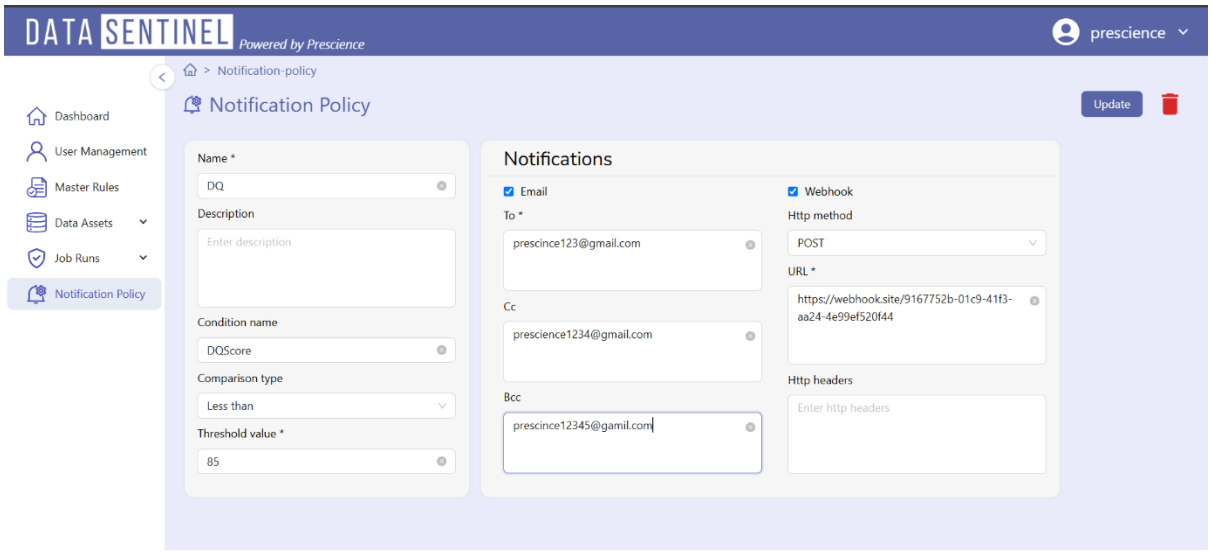

Image 65: Notification policy page

Notifications will be sent based on job run conditions via email. We will receive this email once the job is successfully completed.

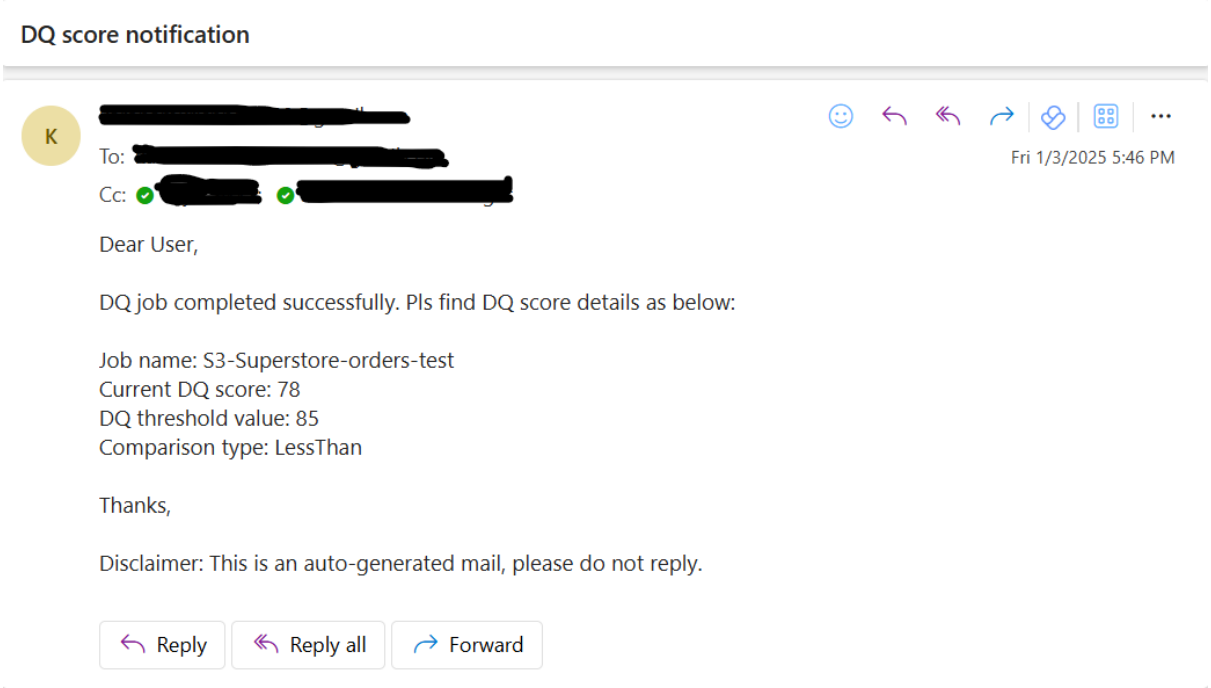

Image 66: Notification email message

4. License installation

After trial period expiry, data quality and data transformation jobs will stop running and a valid license will be needed to activate the product.

How to purchase a license?

To purchase a license, reach out to Prescience Support at info@prescienceds.com. Our team will get back to you shortly.

Steps to activate product:

Login to application as super admin and navigate to the **About** page and enter the required fields as mentioned below to activate the product.

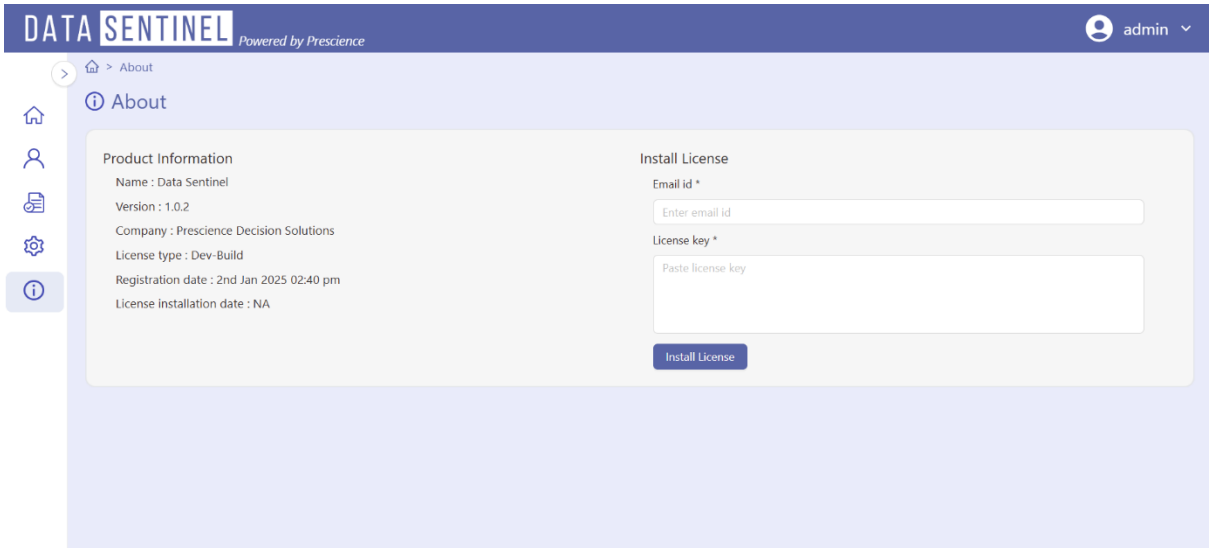

Image 67: License activation

Contact us:

In case of any sales and support queries, drop an email to info@prescienceds.com and our team will get back to you shortly.