



Creating Datasets with ConverSight

Efficiently Build and Manage Datasets for Data-Driven
Insights.

Datasheet

Table of Contents

1. Introduction	03
2. Steps Involved in Creating Datasets	03
3. Input Dataset Information and Locale Settings	04
4. Select one or more Active Connectors for the Data Source	04
5. Select Tables for Dataset	05
6. Select Relevant Columns	06
7. Initiate Loading Process	07
7.1 Check Load Status	07
7.2 Configure SME	09
7.3 Save and Publish the Datasets	10
7.4 Dataset Status	10
8. Conclusion	11

1. Introduction

In the world of data analytics, creating high-quality datasets is critical to achieving accurate and reliable insights. The ConverSight platform provides a powerful toolset for creating and managing datasets that can be leveraged by organizations of all sizes and industries. By building datasets in ConverSight, analysts and data scientists ensure that their insights are based on the most comprehensive and relevant data available.

In this datasheet, we will explore the significance of creating datasets in ConverSight and its contribution to achieving data analytics objectives for organizations. It also highlights the essential features of the platform that benefit data professionals, along with best practices for constructing datasets that are optimized for analysis. Whether you're new to ConverSight or an experienced data expert, this datasheet offers useful insights and actionable steps for creating datasets.

2. Importance of Creating Datasets

Creating datasets is an important step in making the data available for efficient data analysis and visualization within the ConverSight platform. However, creating a dataset can be a complex process, requiring multiple steps and careful attention to detail. Here, we will walk you through the steps involved in creating datasets using ConverSight.

2.1 Steps Involved in Creating Datasets

Here are the overall steps involved for creating a dataset:

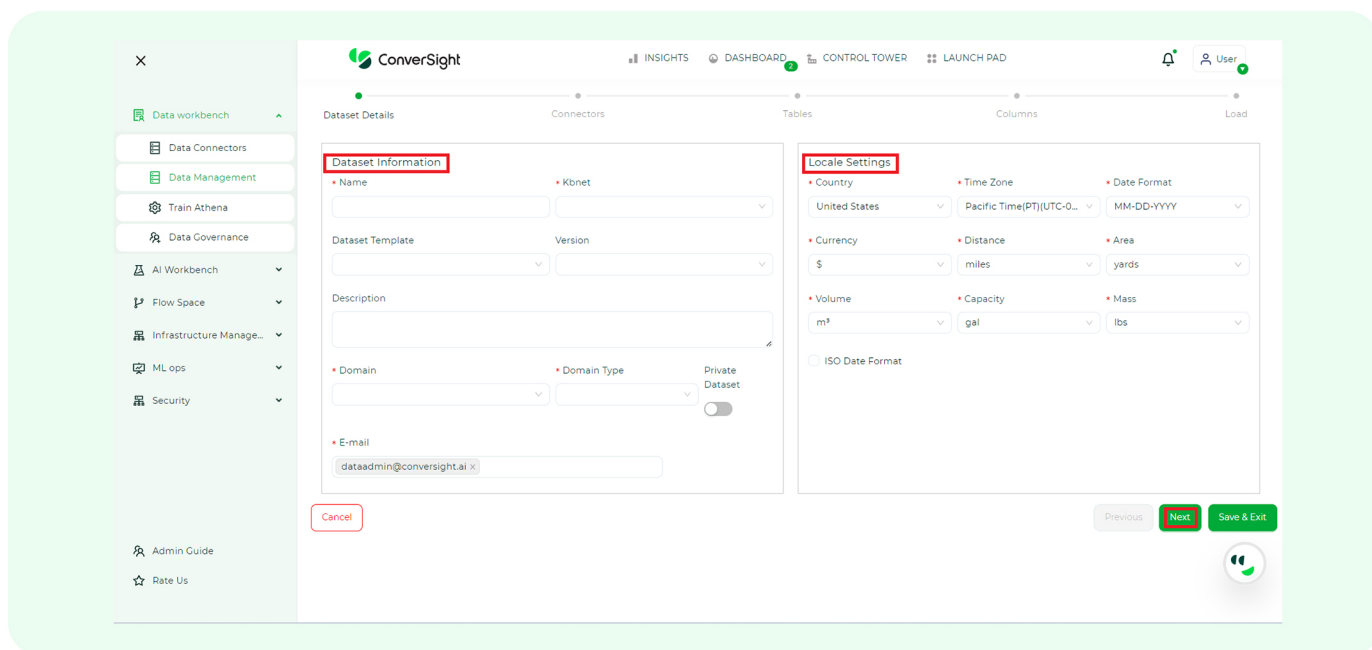
1. Navigate to the Configurations section, select 'Data Management' under the 'Data Workbench' menu, and click 'Create New'.
2. Input the necessary 'Dataset Information' and 'Locale Settings'.
3. Choose a connected data source and click 'Next'.
4. Choose tables to include in your dataset and click 'Next'.
5. Select the relevant columns and click 'Next'.
6. Click 'Load' to initiate the loading process.
7. Click 'Configure SME' to train and provide knowledge for Athena and AI/ML models.
8. Save the changes made and publish the dataset by clicking on the 'Save' and 'Publish' buttons respectively.
9. Once published, the dataset status will change to 'Active', indicating it's ready to be used for querying, analyzing, and visualizing data.

The following section presents a complete guide to creating datasets with ConverSight platform, including detailed instructions and screenshots to facilitate the process.

3. Input Dataset Information and Locale Settings

To create a new dataset, go to the Configurations section and choose 'Data Management' under the 'Data Workbench' menu. Then, select '**Create New**'.

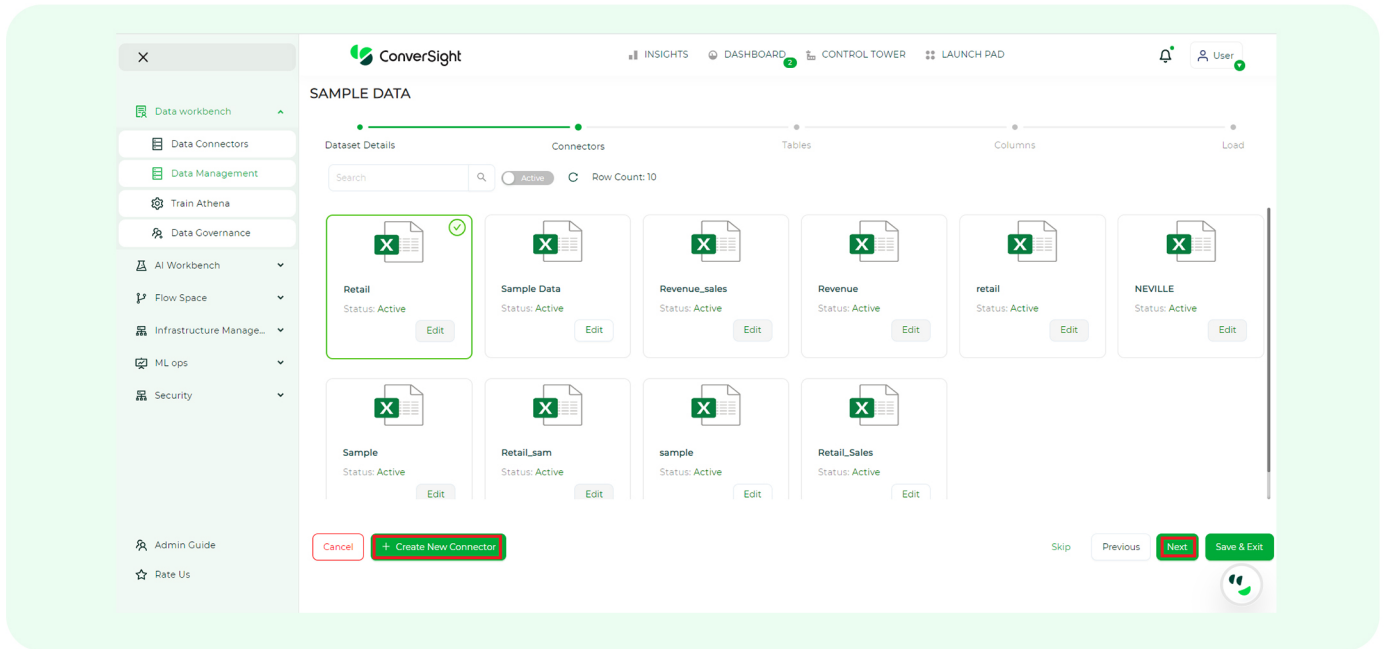
This will take you to the '**Dataset Details**' page where you can enter the necessary information, including the dataset name, description, template, and Locale Settings based on your geographic location.



4. Select one or more Active Connectors for the Data Source

When creating a dataset, it's important to note that only currently '**Active**' connectors can be selected. On the '**Connectors**' page, you will see a list of all the available connectors that were loaded into ConverSight. You can select one or more connectors of your choice from this list. These data sources are loaded previously using the Data Connector module in ConverSight.

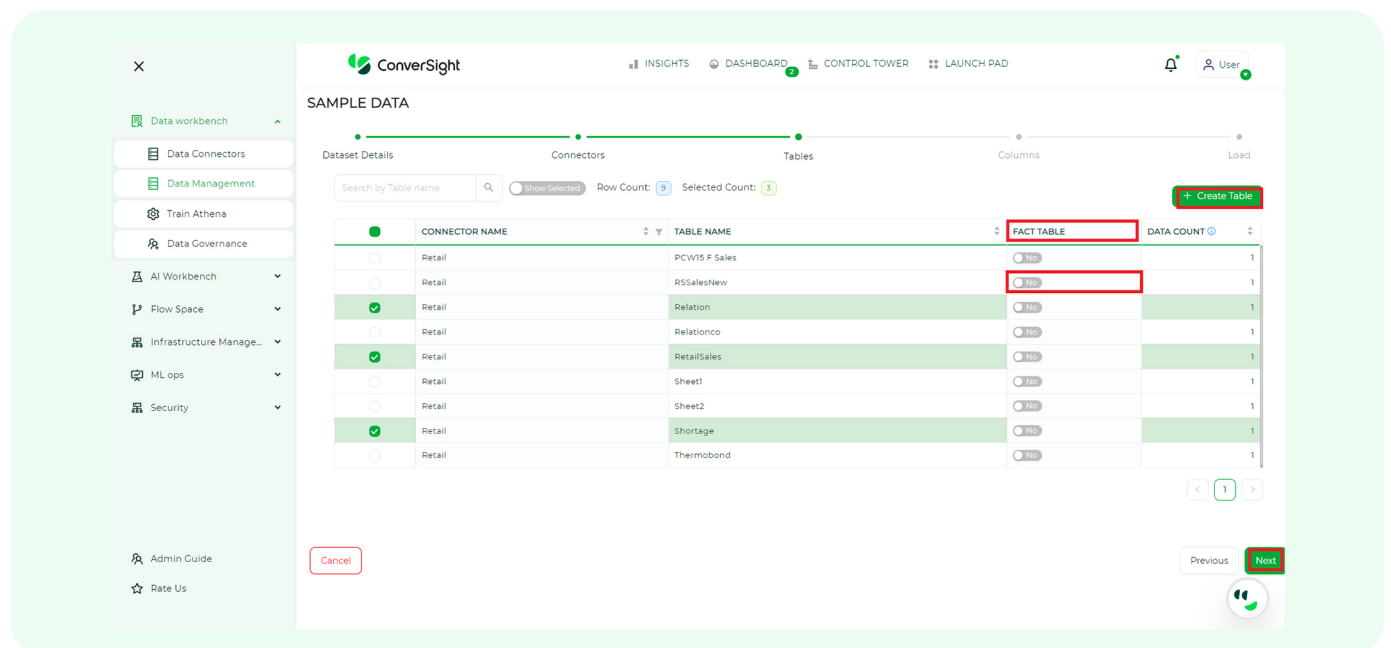
To connect a new data source to ConverSight, click on the '**Create New Connector**' button, and follow the prompts to create and load the data. This is helpful if you need to connect a data source that hasn't been loaded previously, and you can then select it from the list of available connectors during dataset creation.



5. Select Tables for Dataset

Upon choosing the connector, the data will be loaded. The 'Tables' section will be populated with all the tables that are available in the connected data source, allowing you to select the relevant tables you want to access.

This section enables you to view and manage the tables associated with each connected data source, including fact tables. A Fact Table is a table in a data warehouse that holds quantitative information for analysis and reporting. It's often denormalized to improve performance and can be easily enabled using the Fact Table toggle button during dataset creation. To analyze the data in a Fact Table, you'll need Dimension Tables that store information on how the data can be analyzed. Joining the Fact Table's foreign key column with a Dimension Table enables in-depth analysis. Fact Tables are crucial for data governance and accurate reporting in ConverSight.



In addition to fact tables, ConverSight also allows you to create custom tables using SQL queries. Custom tables are tables that you create in ConverSight to meet your specific data needs. To create a custom table, you can select the **'Create Table'** option in the Tables section and write SQL queries to extract data from the connected data source. With custom tables, users can perform all operations just like they would with a normal table that is imported. However, columns will not be listed for custom tables, but you can perform data analysis and transformations on the data as you would with a regular table.

6. Select Relevant Columns

In the **'Columns'** section, you will be able to view a list of all the columns available in the selected tables, and you can select specific columns that you want to include in your dataset. You can also modify the data type of each column by selecting from the available options in the **'Data Type'** drop-down menu next to each column name. This is helpful if you want to customize the data types to match the requirements of your data analysis tasks.

After selecting the columns and modifying their data types (if necessary), you can click on the **'Next'** button to proceed with the dataset creation process.

The screenshot displays the 'Columns' selection step in the ConverSight interface. The main area shows a table with the following data:

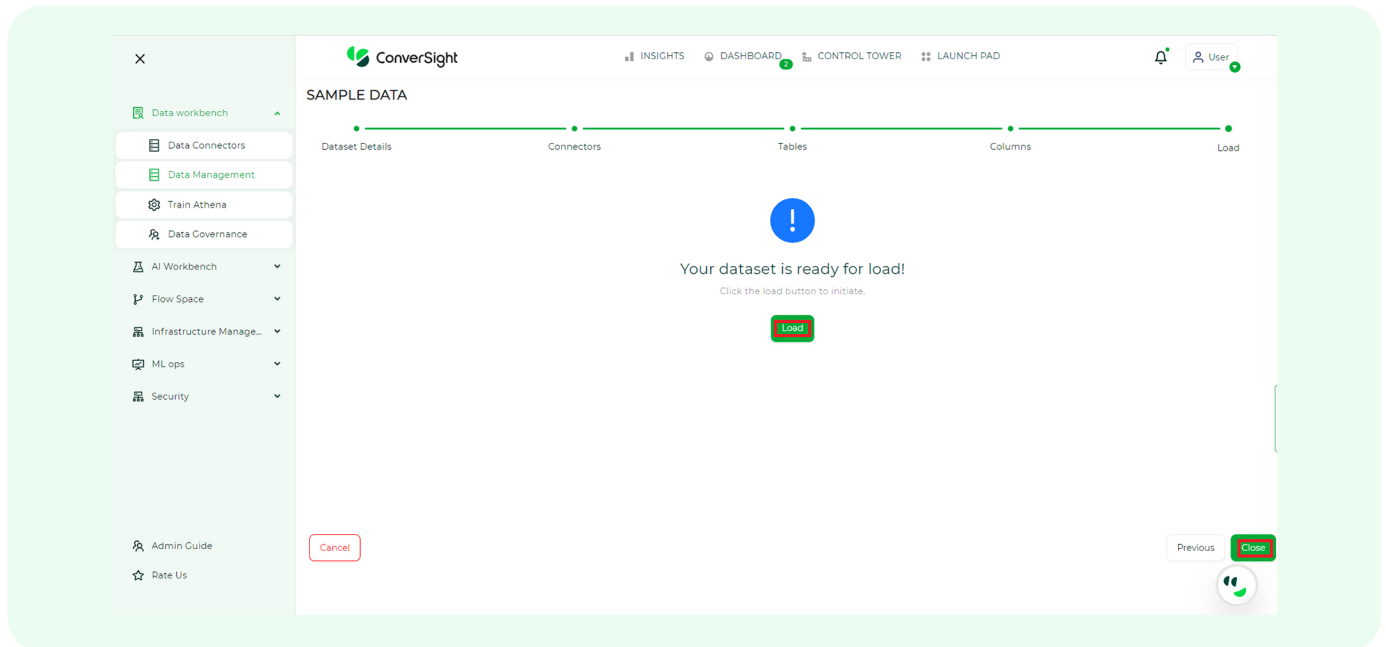
CONNECTOR NAME	TABLE NAME	COLUMN NAME	DATA TYPE	FORMAT
<input checked="" type="checkbox"/> Retail	RetailSales	Zip	String	
<input type="checkbox"/> Retail	RetailSales	email	String	
<input checked="" type="checkbox"/> Retail	RetailSales	product	String	
<input checked="" type="checkbox"/> Retail	RetailSales	rscost	Double	
<input checked="" type="checkbox"/> Retail	Shortage	Count	Double	
<input checked="" type="checkbox"/> Retail	Shortage	Item name	String	
<input checked="" type="checkbox"/> Retail	Shortage	Stock date	Timestamp	Enter Time Format
<input type="checkbox"/> Retail	Shortage	Stock_Order Date	Timestamp	Enter Time Format

At the bottom right, the 'Next' button is highlighted in red, indicating the next step in the dataset creation process.

7. Initiate Loading Process

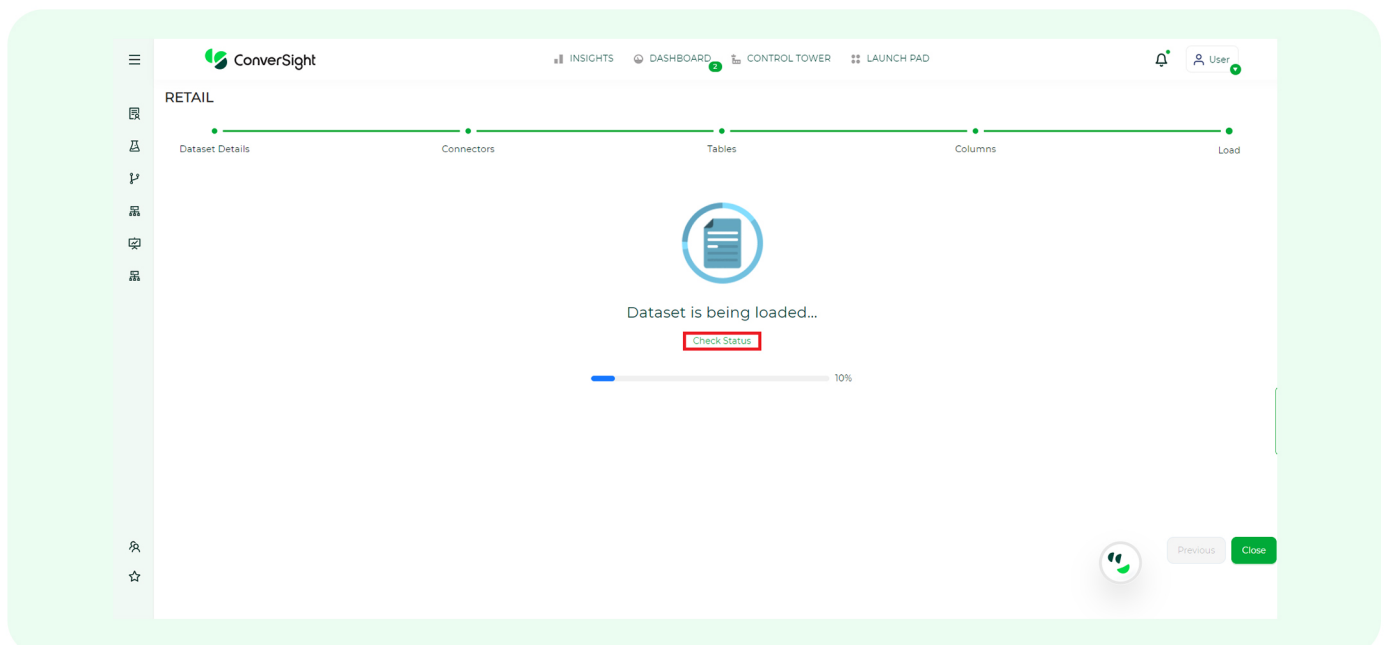
After you have selected the tables and columns you want to include in your dataset, you can initiate the loading process for your dataset.

To load your dataset, click on the 'Load' button, which will start the loading process. This process may take some time depending on the size and complexity of your dataset.



7.1 Check Load Status

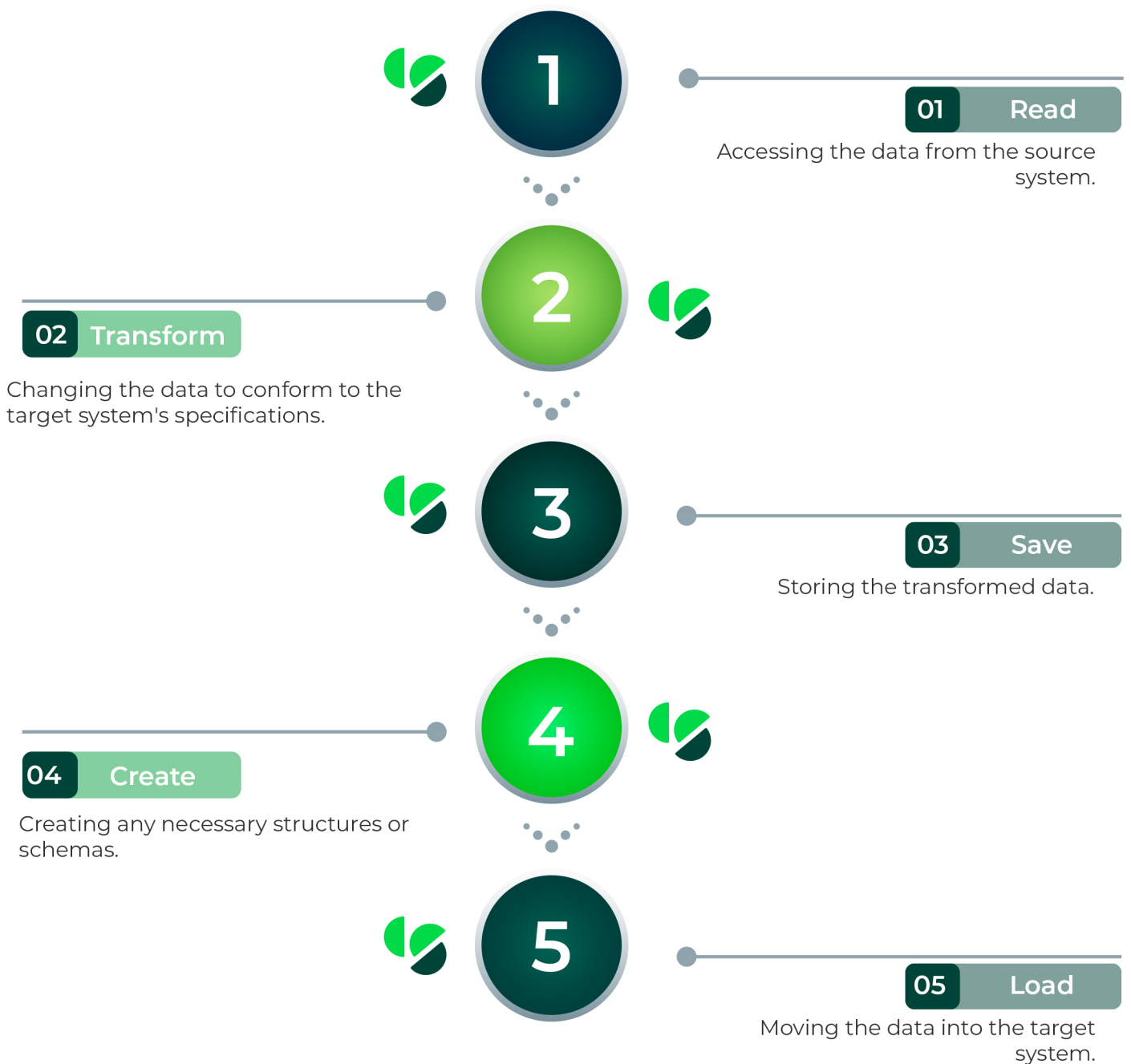
During the loading process, the selected dataset goes through various stages, and you can monitor the progress by clicking 'Check Status'. The load status includes details such as the table name, status, level, runtime, and any errors encountered during the process.

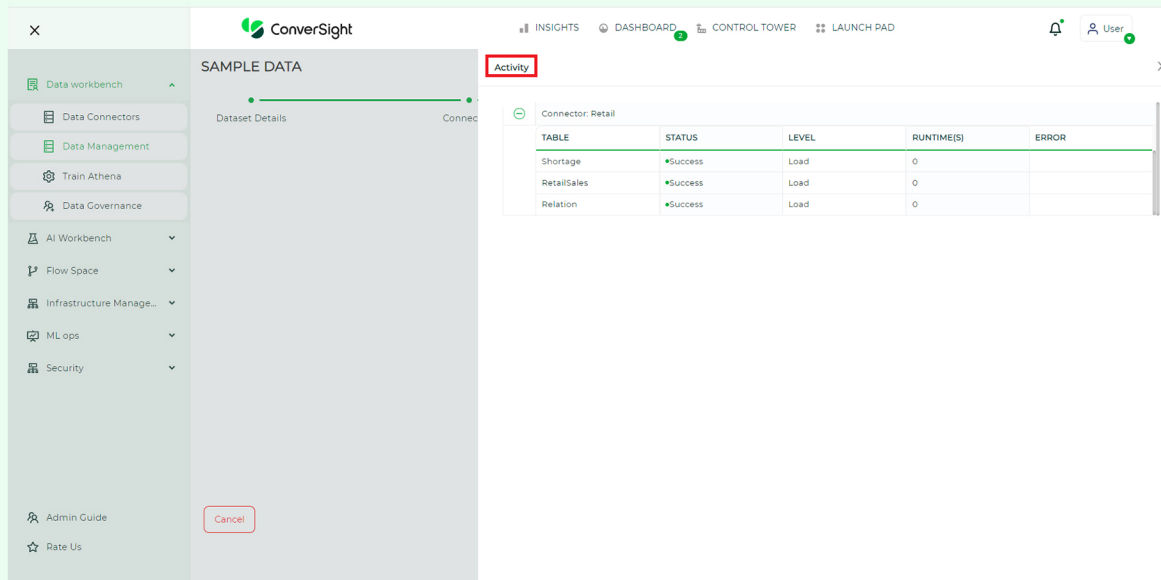


After initiating the loading process of a selected dataset, the status of the load can be viewed on the ConverSight platform. There are two possible statuses:

1. **'Processing'** indicates that the load is still ongoing.
2. **'Success'** indicates that the load has been completed successfully.

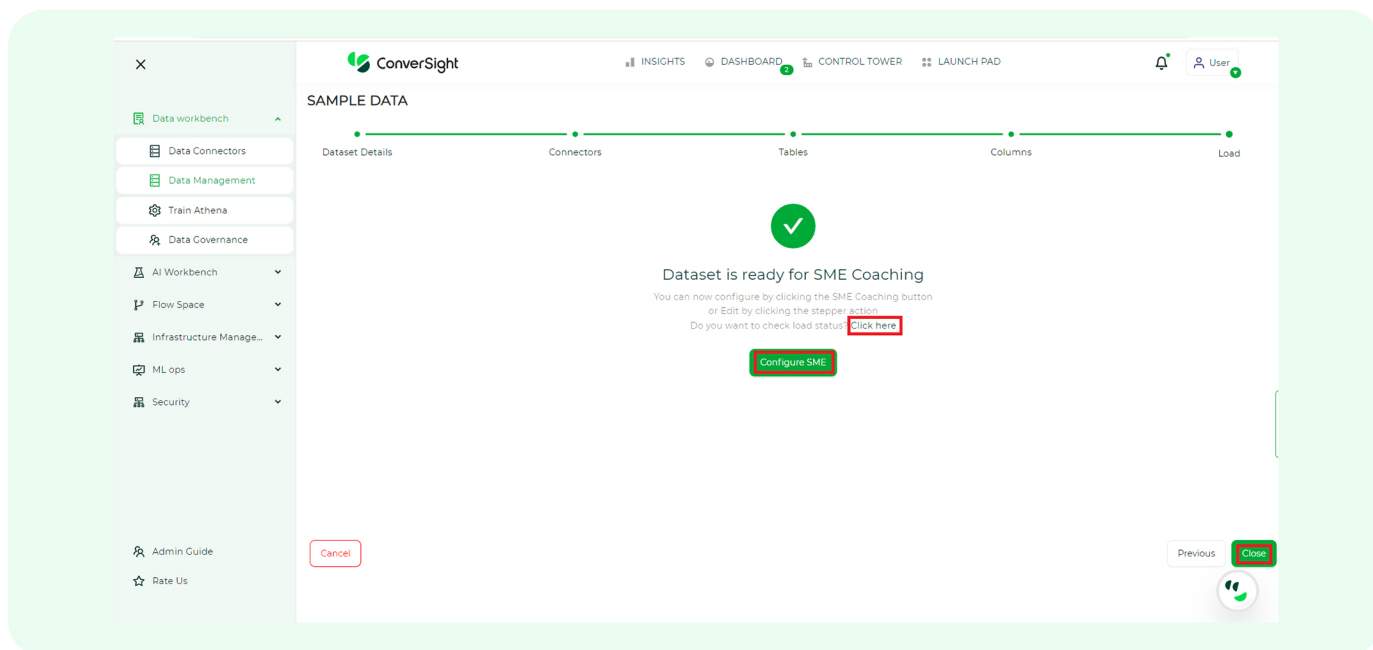
Additionally, the status of the load process is broken down into several levels that indicate the various stages of the loading process, including:





7.2 Configure SME

After the successful loading of a dataset into the ConverSight platform, the next step is to provide knowledge and training for Athena and AI models by proceeding to SME Coaching. To do this, click on the **'Configure SME'** button, which will redirect you to the Configure SME page.

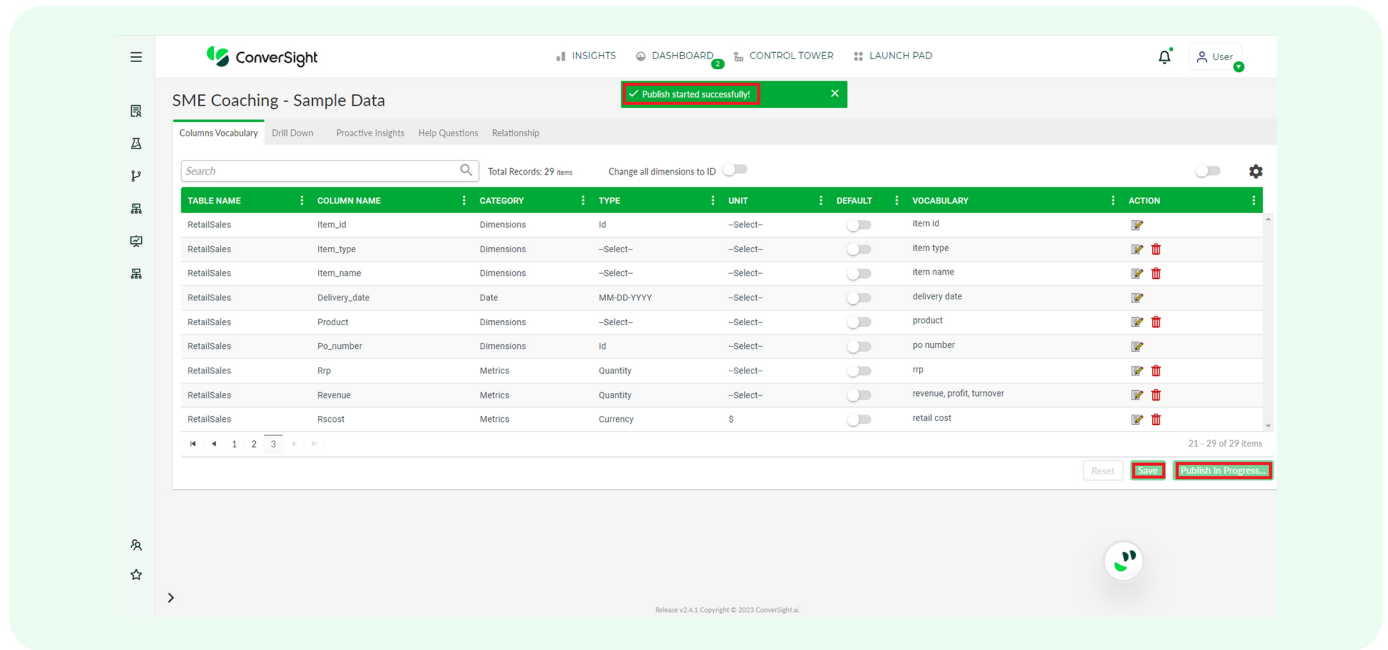


Here, you can update the vocabulary of the columns to make it more meaningful and improve query efficiency.

Optimizing the vocabulary of a dataset can facilitate the process of accessing and extracting insights by Athena and AI/ML models. This can be achieved by training Athena to better understand the dataset and its specific vocabulary.

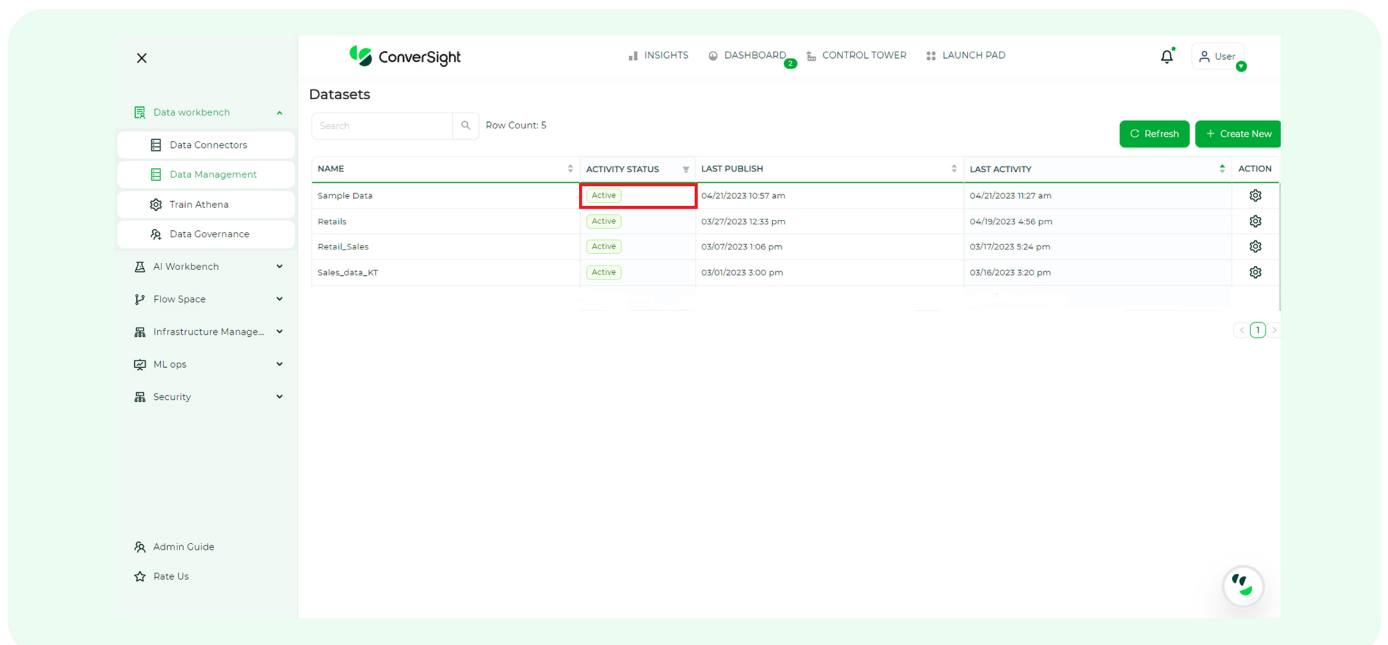
7.3 Save and Publish the Datasets

After making changes to a dataset, you can click the 'Save' button to save those changes. This ensures that any modifications made to the dataset are not lost and can be accessed later. Once you have made all the necessary changes to a dataset and are ready to make it available to other users, you can click on the 'Publish' button. Published datasets are made available for use in reports and dashboards, allowing users to analyze the data and gain insights.



7.4 Save and Publish the Datasets

After successfully publishing the dataset on the ConverSight platform, the status of the dataset will be marked as 'Active', indicating that the dataset is now available for use. This status ensures that all necessary processes have been completed successfully, and users can start querying, analyzing, and visualizing the data. The 'Active' status provides users with accurate, up-to-date, and reliable data, allowing them to create reports, generate insights, and perform various actions on the dataset.



8. Conclusion

By allowing users to easily import, clean, and transform data from a wide variety of sources, ConverSight enables organizations to make better, data-driven decisions. The platform's intuitive interface and advanced features make it a great choice for data analysts, business intelligence professionals, and anyone else who needs to work with data on a regular basis. ConverSight enables you to streamline your data workflows, minimize inaccuracies, and derive valuable insights that facilitate your organization in accomplishing its objectives. So, whether you're looking to optimize your operations, or gain a deeper understanding of your customers, ConverSight empowers you to reach your objectives swiftly and proficiently.

Join our customers who have accelerated growth with ConverSight



[Request a demo](#)

About ConverSight

ConverSight's Adaptive Analytics platform uses conversational AI, Natural Language Processing and machine learning to converge the distance between humans and data through data stories, presenting the meaning of data in the most effective, personalized and efficient form possible. ConverSight's patented AI business assistant, Athena, connects distributed databases to answer questions and Augment the consumers through 4 key functions: Information on demand, Automated Story Telling, Proactive Insights, and Recommended Actions.

For more information, visit www.conversight.ai

