



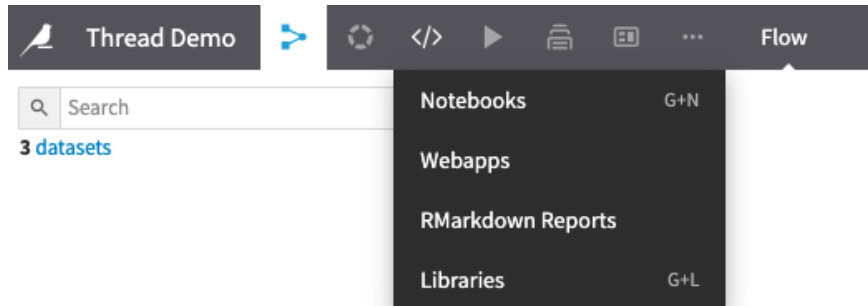
v1.1 Installation Instructions

Updated 6/23/22

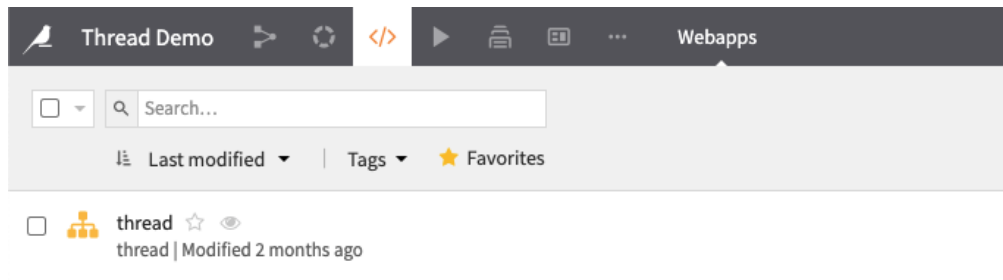
Getting Started	2
Home Screen	3
Search	3
Dataiku Instance Stats	3
View Search Results from Home Page	4
Project	4
Dataset	5
Column	7
Create & Edit Definitions	9
Apply a Definition to a Column	9
Catalog Screen	11
Adding New Data	12
Sending Links	13
FAQ	14

Getting Started

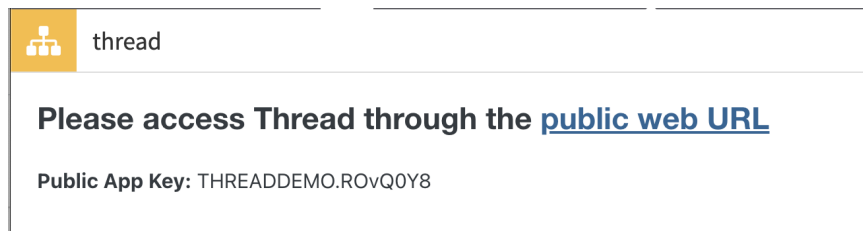
You must be logged into Dataiku to access Thread. To access Thread Data while in Dataiku go to your Thread project and then Webapps



Select Thread Data




To open this as a public webapp click the link



Home Screen

THREAD

HomeCatalogvmaus ▾

Search 

Dataiku Instance Stats

Projects

170

Datasets

1,904

Columns

32,023

Definitions

22

Projects Recently Created in DSS

Scan for new Projects

Name	Last Modified
DKU_CHURN	6/9/2022, 5:35:53 AM
VIDADEMO	6/2/2022, 7:47:11 AM
DKU_TUTORIAL_BASIC_101_2	5/16/2022, 1:51:44 PM

Search

This search bar allows you to search for columns, datasets, projects and definitions. Uncheck the boxes to filter the results.

Dataiku Instance Stats

This is a summary of all projects, datasets, columns, and definitions in **only this node** of Dataiku that have been indexed.


View Search Results from Home Page

Project

TH-READ

[Home](#)[Catalog](#)[vmaus](#)

House Prices(HOUSEPRICES)



House Prices

(HOUSEPRICES)

Project owned by **Ryan Moore** in **training** folder

imported

Options







Documentation Status

3%

9 / 331 columns

Datasets

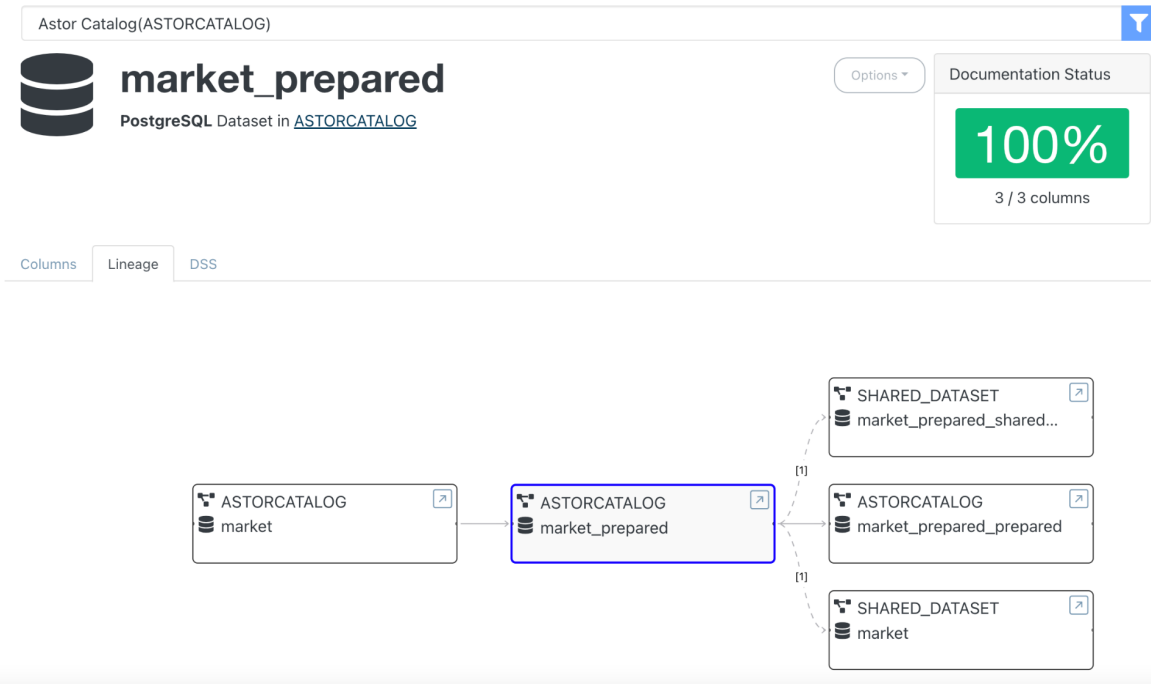
DSS

Name	Documentation Status
 data_key	0 / 2 (0%)
 sample_submission	0 / 2 (0%)
 test	3 / 80 (4%)
 test_train_joined	2 / 81 (2%)
 test_train_joined_prepared	2 / 85 (2%)
 train	2 / 81 (2%)

You can see on this page

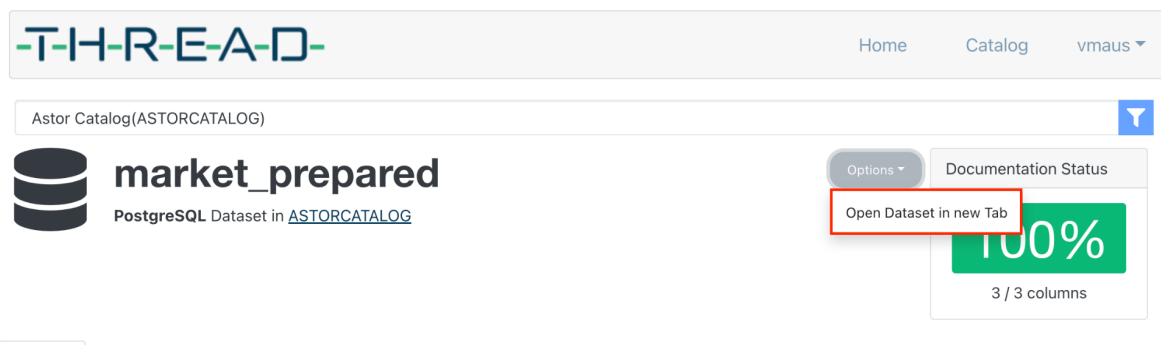
1. Project Name
2. Project ID
3. Project owner
4. Project folder
5. Project Tags
6. Datasets tab
 - a. Name: All datasets in this project. Clicking this will take you to this dataset's result page
 - b. Documentation Status: How many columns have been documented in each data set
7. DSS tab
 - a. Project flow in Dataiku
8. Documentation Status of all columns in the entire project
9. Options
 - a. Open project in new tab: new tab to view this project information page
 - b. Rescan project: for bringing in new datasets and columns from Dataiku

- c. Final dataset(s)
- d. Numbers between the boxes indicate count of datasets between these datasets
- e. The blue arrow in the top right corner of each box allows you to jump to that datasets results page



DSS: View the dataset explore page in Dataiku

- 6. Documentation Status: Total number and percentage of columns in this dataset that have been defined
- 7. Options
 - a. Open Dataset in a new Tab

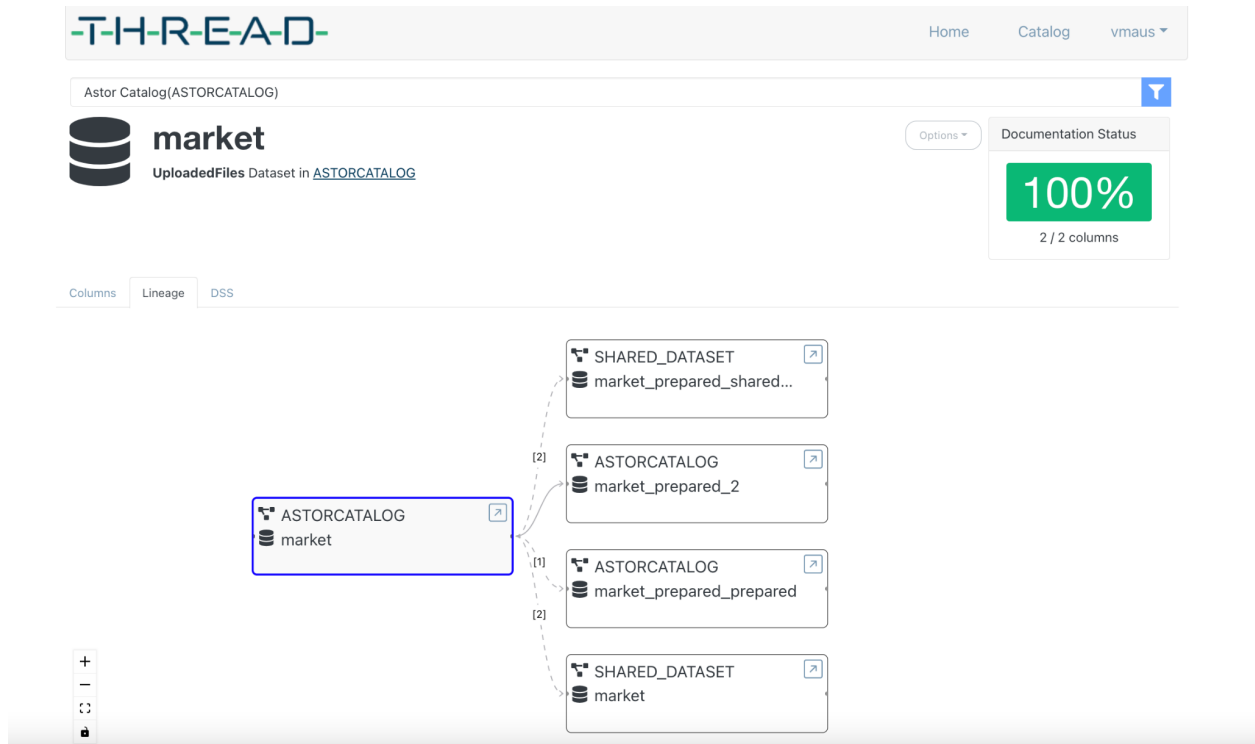


Column

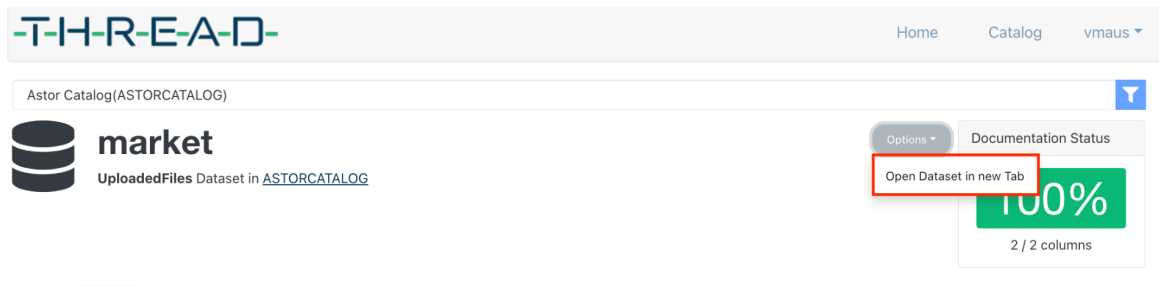
The screenshot shows the 'Thread' application interface. At the top, there's a navigation bar with 'Home', 'Catalog', and 'vmaus'. Below this is a search bar containing 'RoofMatl'. The main content area displays the 'RoofMatl' column definition. It includes a 'Definition' tab, a 'Lineage' tab, and a 'DSS' tab. The 'Definition' tab is active, showing the column name 'RoofMatl', its data type 'string', and its location 'column in HOUSEPRICES | test'. There are also tags 'housing express' and 'pii'. A 'Description' field states: 'Roof Materials: CompShg (Standard (Composite) Shingle) and Style is Primarily Gable, then Hip.' To the right, there's a 'Documentation Status' section with a 'Documented' button. A blue arrow icon is visible in the top right corner of the main content area.

You can see on this page

1. Column Name
2. Data type
3. Project Name. You can click this to view this project results page
4. Dataset name. You can click this to view the dataset results page
5. Definition
 - a. Definition Name
 - b. Definition Tags
 - c. Description of definition
6. Lineage
 - a. Source dataset of this column (even if in a different project)
 - b. Dataset the column you are viewing is located in highlighted in blue
 - c. Final dataset(s) this column goes to
 - d. Numbers between the boxes indicate count of datasets between these datasets
 - e. The blue arrow in the top right corner of each box allows you to jump to that dataset results page



7. DSS: View the dataset with this column in Dataiku explore page
8. Documentation Status: Whether or not this column has a definition
9. Options
 - a. Open Dataset in a new Tab

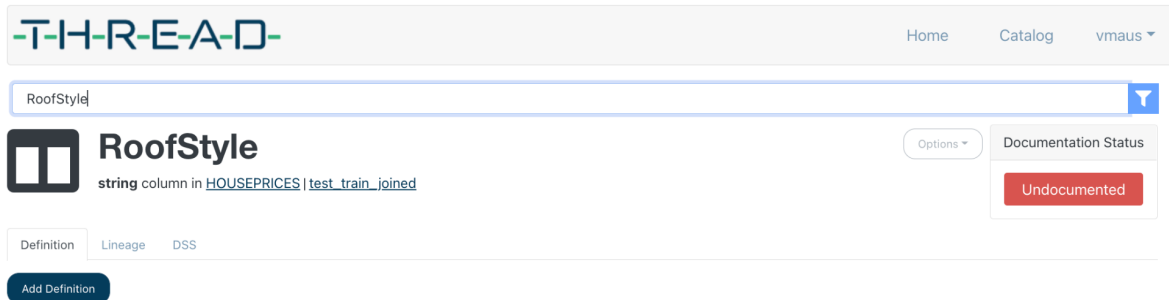


Create & Edit Definitions

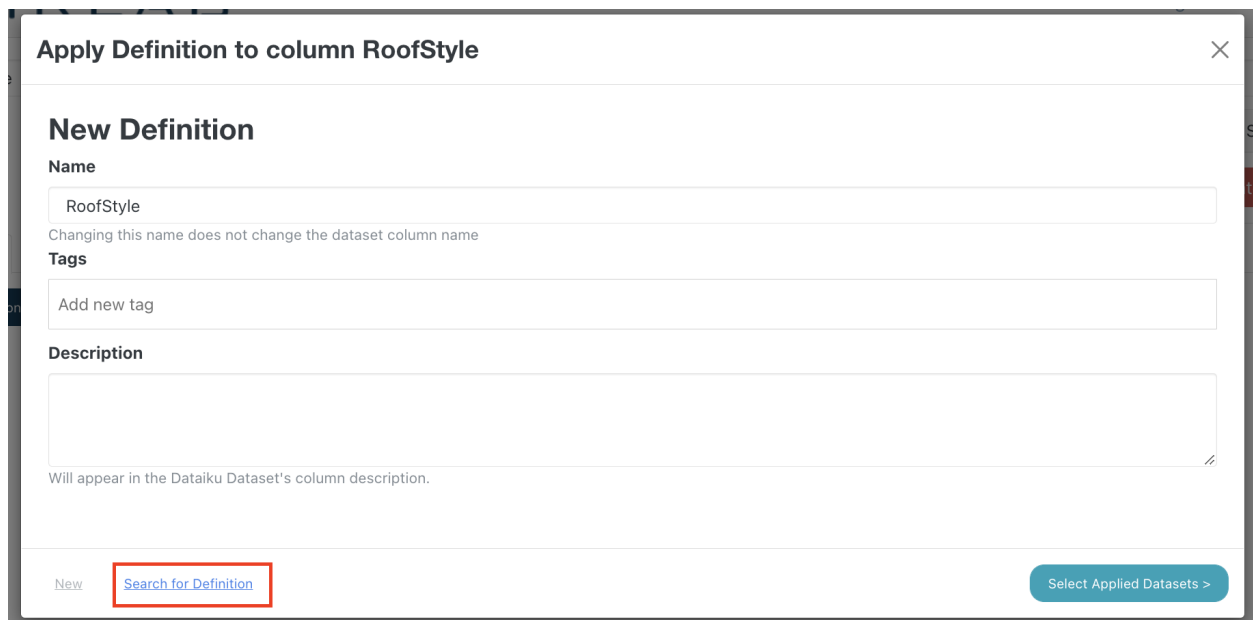
To create, edit or remove a definition from a column find the column using the home screen search.

Apply a Definition to a Column

Click Add Definition



If a definition for this column already exists, “Search for Definition” in the bottom left corner. Find the correct definition and apply to this column.



If this is a brand new definition enter:

1. Name of definition: This is a business-friendly name for the definition that may be applied to many columns that do not have the same name.
2. Tags: Apply new tags or existing tags by typing the tag name
3. Description: definition of the word

Select Applied Data Sets

1. Select which datasets you would like this to apply to in the lineage. Any places this has been applied previously will automatically select and can be unselected.

Apply Definition To Lineage



Select the up and downstream datasets to apply this definition

This Column

☒ RoofStyle

[All](#) | [None](#)

Upstream

☐ HOUSEPRICES|test|RoofStyle

☐ HOUSEPRICES|train|RoofStyle

Downstream

☐ HOUSEPRICES|test_train_joined_prepared|RoofStyle

< Edit Definition

Apply

Catalog Screen

View all definitions on this node of Dataiku. One definition may be applied to many columns, even if they don't have the same name.

You are able to view

1. Definition name: Sort this alphabetically ascending or descending by pressing the arrow next to the title. Click this link to edit or delete (using options) the definition
2. Description: Sort this alphabetically ascending or descending by pressing the arrow next to the title
3. Tag: Customizable metadata added when assigned to a column

Navigation

1. Search by name, definition or tag using the top right search bar
2. Filter by tag in the filter bar
3. Filter by tag by clicking a tag

THREAID

Home Catalog vmaus

≡ Definitions

Filter By Tag

Search Definitions

Name ▲	Description ▲ ▼	# Applied To	Tags
AreaCode	The Area Code	3	tag another
Customer Service Calls	Customer Service Calls	6	Taxi pii
day_number	Number of the day	7	
Distance to LaGuardia	Driving distance to LaGuardia in miles	2	airport Taxi
dropoff_latitude	dropoff_latitude - float for latitude coordinate of where the taxi ride ended.	2	
dropoff_longitude	float for longitude coordinate of where the taxi ride ended.	1	
Fence	GdPrv Good Privacy MnPrv Minimum Privacy GdWo Good Wood	1	fence-data pii

Click on the name to edit or delete

From this page you can view or edit

1. Column Name
2. Name of the definition (this does not need to match the column name)
3. Tags
4. Applied to: if you click this will take you to that dataset results
5. Description: definition in business term

THREAD

[Home](#)[Catalog](#)[vmaus](#)

Phone Number

Options

Definition

Details

Name

Phone Number

Tags

pii

Add new tag

Applied To

DKU_CHURN|Churn_Operations|Phone

Description

The Phone Number for the call in

Will appear in the Dataiku Dataset's column description.

Save

To delete this definition options > delete

THREAD

[Home](#)[Catalog](#)[vmaus](#)

Phone Number

Options

Definition

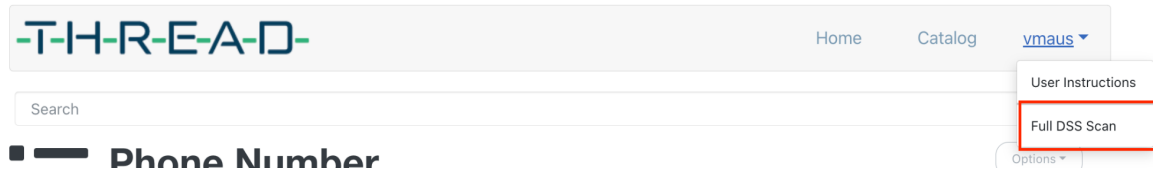
Details

Name

Phone Number

Adding New Data

In order to add new projects, datasets and columns from Dataiku into thread you need to rescan DSS using the arrow next to your username.



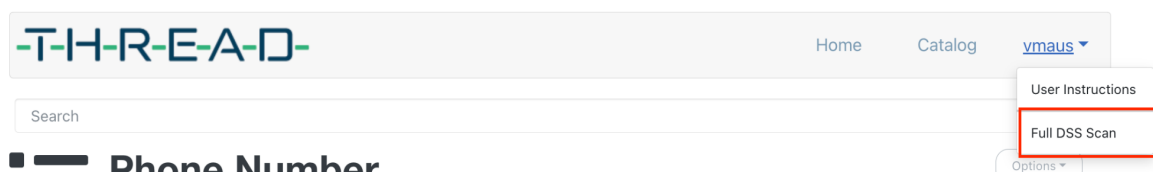
Sending Links

You can share a link to a specific page in Thread with anyone who has security access in Dataiku and permission to “read project content” in the Thread project.

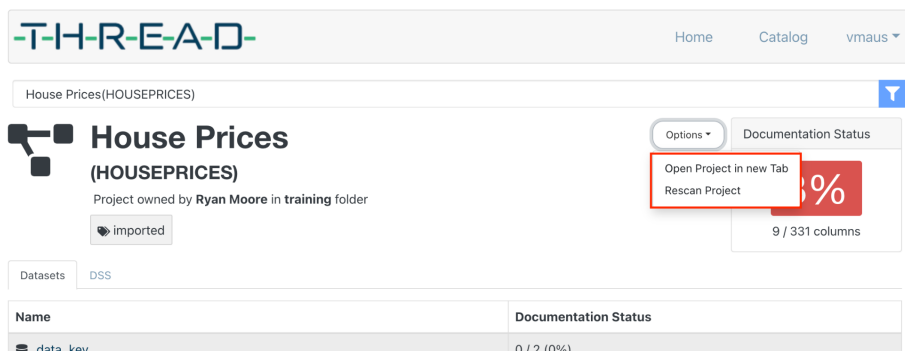
FAQ

Why can't I see my project, dataset, or column in the search results, or lineage screen?

Thread needs to index any new projects, datasets, or columns. You can do this for all projects datasets and columns in your instance by going to the arrow in the top right corner next to your username.

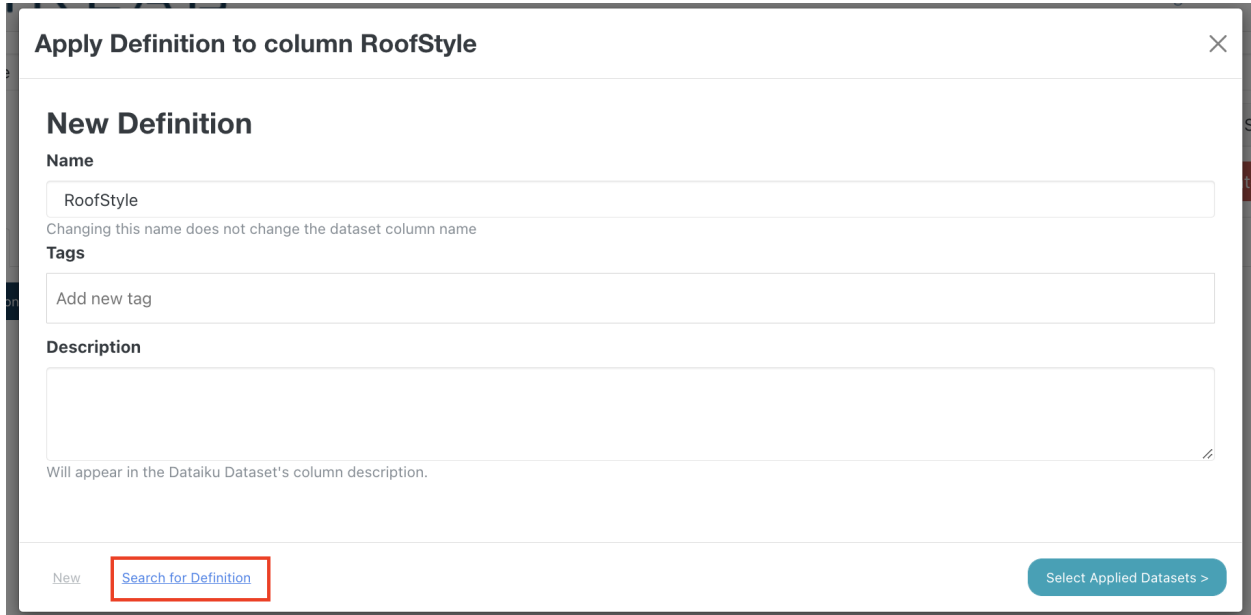


Alternatively, you can do this per project by searching for that project on the home screen and using options > Rescan Project. This method is suggested for those instances with a lot of projects.



How do I know if a definition already exists for this column?

From the home screen search for the column you would like to define. When you add definition use the “Search for Definition” in the bottom left corner. If you find the definition you can apply it here.



Alternatively, you can view all existing definitions in the catalog before going to edit the column.

Can everyone see all the data in my instance?

User must have Dataiku access to use Thread and permission of “read project content” to the Thread project. All users can see the column names and definitions for all projects on the instance. If the user doesn’t have access to a particular project, they cannot view the DSS tab in the project, dataset or column results screens.

Who can view, edit or delete definitions?

Any Dataiku user with “read project content” permission to the Thread project can view, edit and delete definitions in Thread and the column descriptions in Dataiku.

What if I change the name of a column?

Thread is able to recognize the same column with a different name as the same lineage if the column was renamed in a prep recipe.

How do I know how much is documented?

You can view by project, dataset, or column how much has been documented.

Can I have two Thread webapps running at the same time?

Yes, you can have two Thread webapps but they need to be in separate projects.