CLOUDERA DATA PLATFORM DATA CATALOG LAB

Step-by-step instructions:

If you can read this you found the PDF. If you are reading it in a Browser please download the PDF to your local computer. This will let you cut and paste from the PDF into the Cloudera Web User Interface.

Part 1 - Data Catalog [10 minutes]

Overview: What is Cloudera Data Catalog?

Data Catalog is a service that enables you to understand, manage, secure, and govern data assets across the enterprise. Data Catalog helps you understand data across multiple clusters and across multiple CDP environments. You can search to locate relevant data of interest based on various parameters. Using Data Catalog, you can understand how data is interpreted for use, how it is created and modified, and how data access is secured and protected.

Purpose: Search for a dataset (table) in Data Catalog, called “flights”.

- Find what database(s) the table “flights” is located.
- Find out at least one year that the “flights” table was generated from.
- Find out how many columns the table “flights” contains.

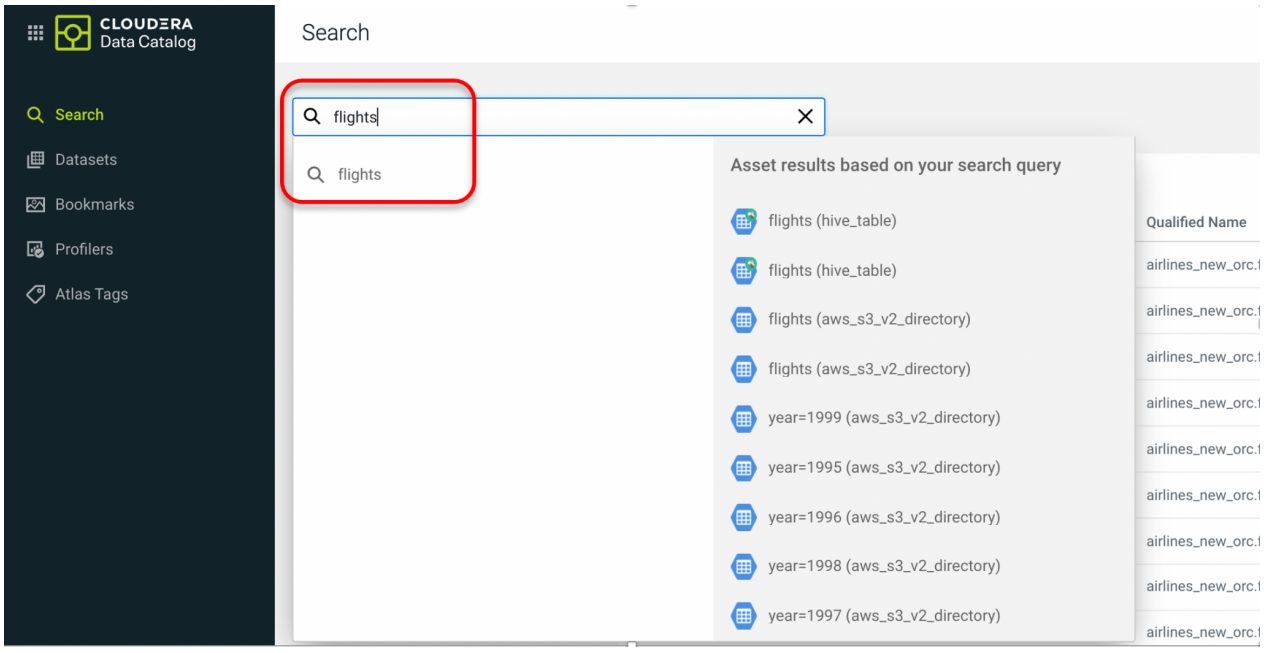
1) Open CDP, using the “admin” user within the Test Drive link.

Your link should look something like (remember click the link in your email not the link below)
http://login.trycdp.com/auth/realms/trycdp-trialxx/protocol/saml/clients/samlclient?tn=trialxx_admin@trycdp.com&p=X
*xx represents the trial user #

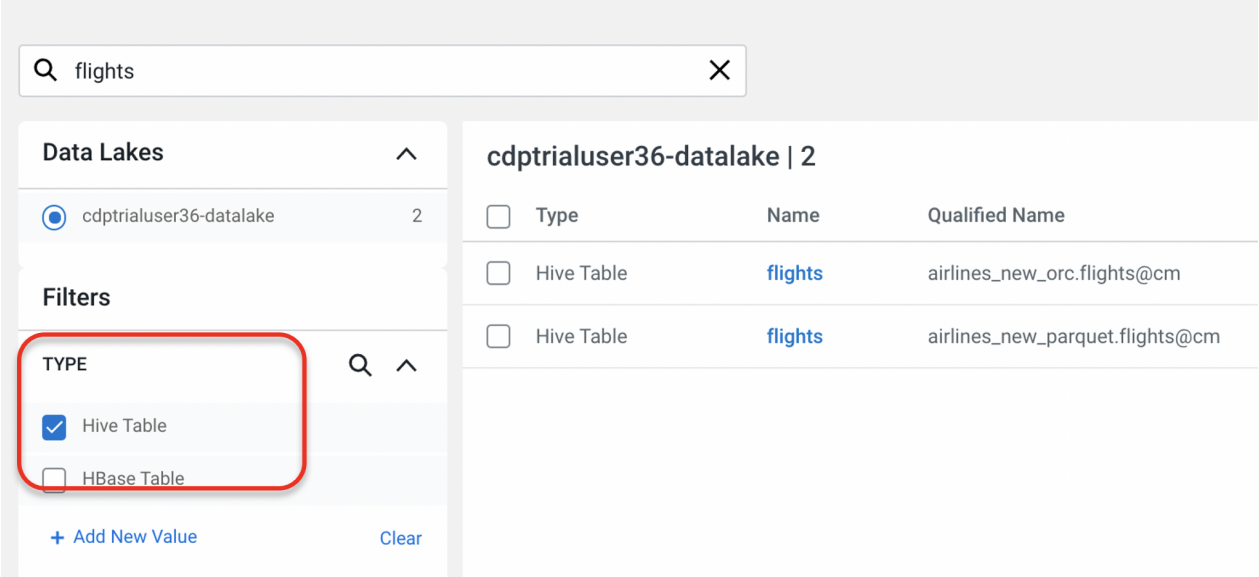
*X represents the password

2) Click the “Data Catalog” within the CDP Home Screen

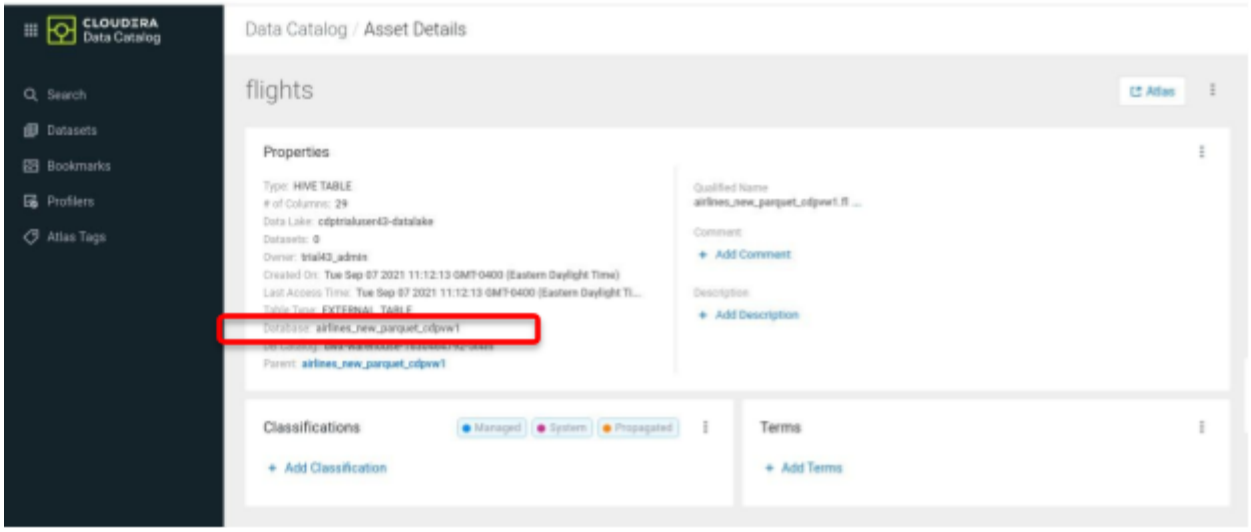
3) Type “flights” in the search box and press enter on your keyboard (or you might click “flights” under suggestions)



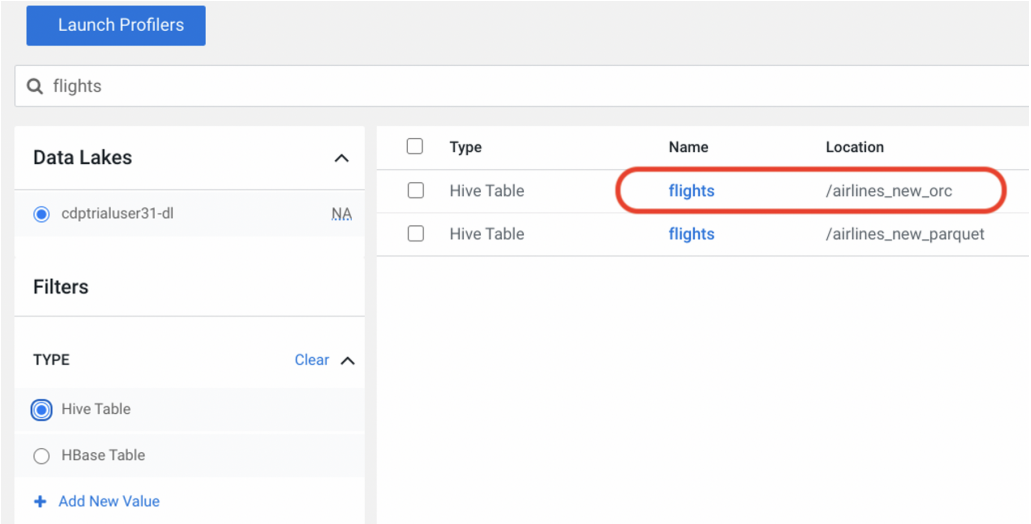
4) Click “Hive Table” under Filters on the left



*Find what database(s) the table “flights” is located.

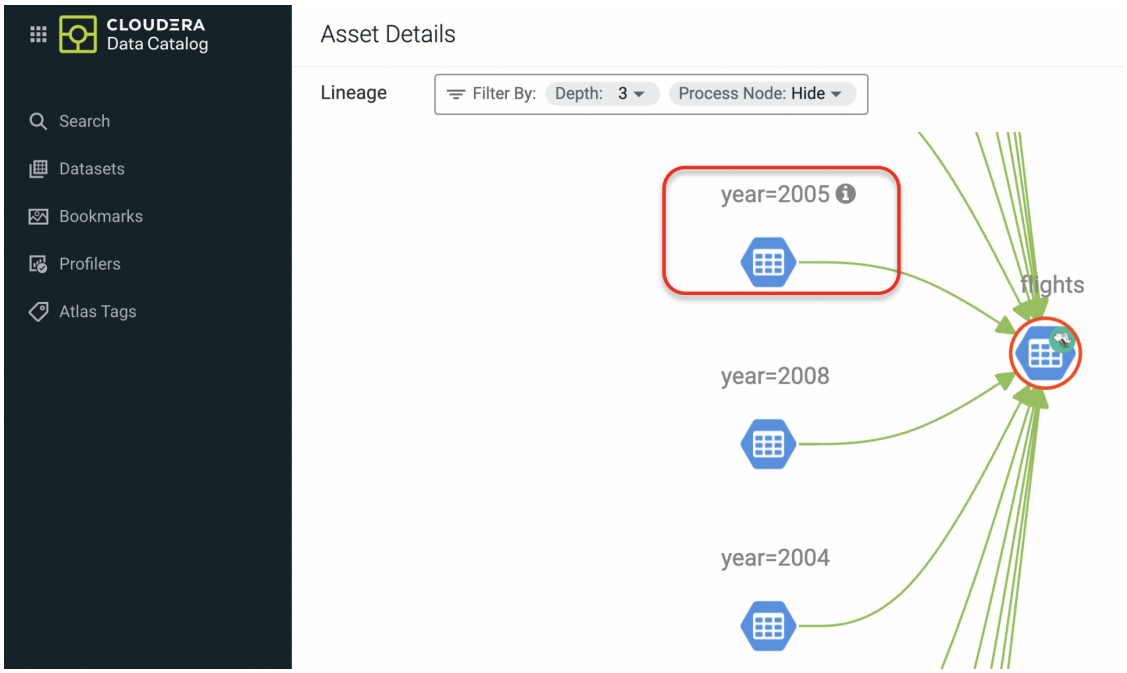


5) Click “flights” where the Location = /airlines_new_orc



*Find what database(s) the table “flights” is located.

6) Zoom into the Lineage and scroll over one of the year=2005 input dataset, clicking the “i” for more information



*Find out at least one year that the “flights” table was generated from.

*Find out how many columns the table “flights” contains.

END