INTRODUCTION

CONVENTIONS

Data Catalyst training is done on a shared environment, just as you would use it in real life. When executing training exercises, that if trainees are given explicit names for objects, we will run into conflicts (you can't save two objects named "foo").

In order to avoid this, anything you create as a user should be prefixed with your username.

> Create a source called <username>_big_data.

If you see an instruction like this and your name is Ruth Bader Ginsburg, you would name this rginsburg_big_data.

ENVIRONMENT

The following environments are the default for these exercises. If your organization is mapped to different environments, please check your training welcome email for details.

QDC
http://training.qdc-demo.com:8465/qdc
User and password – see email

Ambari [for access to hdfs and hive]
http://training.qdc-demo.com:8888/
User: trainee
Password: trainee

AWS S3
Verify that the source connection called s3_demo is working

- Click on the Source module

- Click on Source Connections

- Select the row that has the "S3_demo" source connector

- Click on the pencil and then select **Test Connection**

GROUPS
You have been assigned to a group, to make the environment manageable for first time users.
[your_group] [you'll be assigned to a unique group to which you have admin access]
Demo   [you have access to this data, but can not modify]
You will assign all objects to the [your_group] group

QlikSense
For integration from QDC to Qliksense.
https://sense1.qdc-demo.com
username: IP-AC1F0E10\presales
password: Qlik!23$

Catalog Basics (QDC 4.0.6)

In this exercise you will explore the Catalog and its content, as well as publish data to Qlik and create new categories (datasets).

Working with the Catalog

The Catalog is where all meta data and entities are stored and cataloged. You will find that there is a lot of information in the Catalog, and we will explain some of it here.

- Click on the Catalog module
- You will see tiles for the entities in the **Catalog**. They have three composite metrics that describe aggregate characteristics of the data in them, amongst other things

-
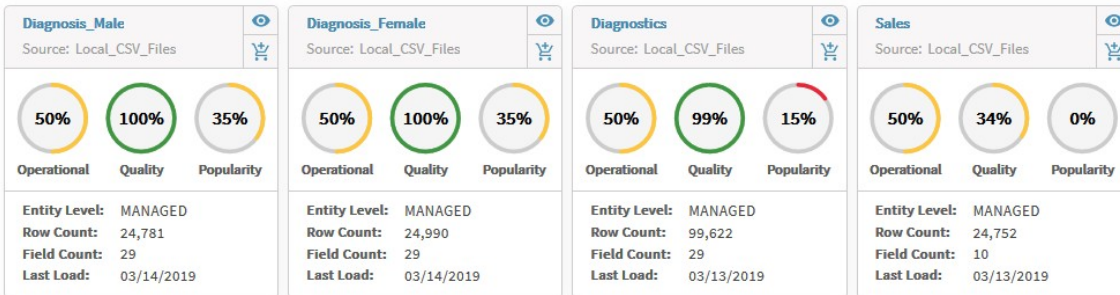
- The **Operational**, **Quality** and **Popularity** metrics describe in detail the following things:

  a. **Operational**
     A composite of the relative number of finished loads and how long ago the data was last refreshed/loaded. Newer and more regularly updated data will yield a higher %.

  b. **Quality**
     A composite of the relative amount of Bad and Ugly records. Indicates if the original data set is of good quality or not. Lower amounts of Bad and Ugly records will yield a higher %.

  c. **Popularity**
     A composite of the number of tags, comments, workflows in which the entity is involved and the number of publish jobs it is in. A higher presence in jobs and workflows, as well as strong activity and meta data will yield a higher %.

     (Note: These metrics are client configurable – these are the out of the box definitions)
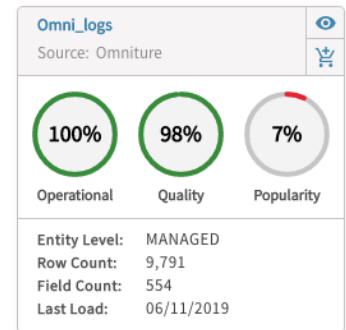
- Click on the eyeball ◎ icon on the **Omni_logs** entity. You will be taken to the **Details** of the entity.

- Note the Business metadata – you would have been able to find this entity if you had searched for "Adobe" or "events" or "entry."

- Note also the tags – tags can be added by users or automatically based on rules. Also a filtering capability in the product. Bronze in this case means validated Raw data with no modifications.

- Look in the **Field Statistics** and **Metadata table**. These values, like Min/Max value, Percentage Nulls, Cardinality and more, is generated automatically by Qlik Data Catalyst® and cataloged with the data.

	Field Name	↑ Data type	Min Value	Max value	Percentage Nulls	Number of Nulls	Cardinality	Total Count
	accept_language	STRING	da-DK	vi-VN,vi;q=0.8,en-US	0.9%	93	58	9791
	browser	INTEGER	0	4239865298	0.0%	0	107	9791
	browser_height	INTEGER	0	1950	0.0%	0	537	9791

- On the top right of the page, click on the **Sample Data** icon ⬚ **Sample Data**, which will show a sample of the whole dataset. Note that this action takes you to the **Discover** module.

accept_language	↑ browser	browser_height	browser_width	campaign	c_color	channel
da-DK	638	933	1680		24	
de-DE	638	805	1300		24	
en	3118857255	137	128		24	
en-us	651211754	811	1440		24	sports
en-us	651211754	1128	1070		32	

> Now, let's use some of the filtering capabilities ot find data and get ready to publish this data to Qlik Sense.

- Click on the **Catalog** module ▦ **Catalog**.

- We're looking for worldwide sales by employee.

- Click on **Filters**

- Choose the dataset "TOPIC_Sales" Datasets: [TOPIC_Sales ▼] and click **Apply**

- Do a search for "world" in [Search by name... 🔍] to narrow it down to the desired data

- Click on the eyeball 👁 icon on the **Global_Transaction** entity.

- Scroll down to the bottom and look at the related entities. We wanted to look at sales by employee

- Ctrl-Click (Windows. ⌘-click on Mac) Global_Transactions and Employee_Lookup

- Click 🛒 Add To Cart

- At the top of the page, mouse over My Cart (which has two entities in it) and click on Publish to Qlik

▼	Related Entities
☐ Select All	🛒 Add To Cart

	Entity Name
▶ ☑	Global_Transactions — Source: Consolidated
☐ 👁	city_dim — Source: iot_demo
☐ 👁	country_dim — Source: iot_demo
☑ 👁	Employee_Lookup — Source: Consolidated

🛒 My Cart ② ▾ | Global Se

Take Action ▾
Create Dataset
Add to Dataset
Publish
Explore
Prepare
Publish to Qlik

- A new tab should open in Qlik Sense – you can now use that environment to analyze the data

a. Click on **Generate insights** to let Qlik Sense generate some suggested visualizations

b. In the upper left corner, click on Data Model Viewer to see the relationships

c. Also in upper left, click on Data load editor to see the load script QDC generated. Note the data connection that QDC used
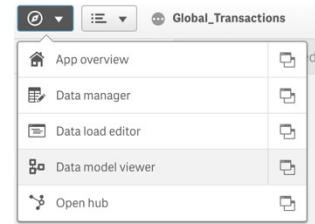
demo.qdc-demo.com (ip-ac1f0e10_presales)

d. If Qlik Sense does not open, check if

    i. Your browser is preventing pop-ups, so that it won't open the new tab and

    ii. That you're logged into the Sense server (logins on first page)

- Go back to the QDC tab and scroll to the top.

- Depending on your version, you will see either the Impact (older) or Lineage icon. Click on it.

Entity Lineage Report

Parent Source↑	Parent Entity	Job Name	Child Entity Type	Child Source	Child Entity	Relation Type
Consolidated	Global_Transactions		PREPARE_TARGET	Consolidated	EMEA_Sales_by_Employee_Detail	INGEST
Consolidated	Global_Transactions	Global_Sales_by_Employee	PREPARE_TARGET	Consolidated	EMEA_Sales_by_Employee_Detail	PREPARE
Consolidated	Global_Transactions	Global_Sales_by_Employee	PREPARE_TARGET	Consolidated	Sales_by_region_and_vertical	PREPARE
Consolidated	Global_Transactions		PREPARE_TARGET	Consolidated	Sales_by_region_and_vertical	INGEST

- Impact:

- Lineage (4.1 and later):

Entity Lineage: Global_Transactions
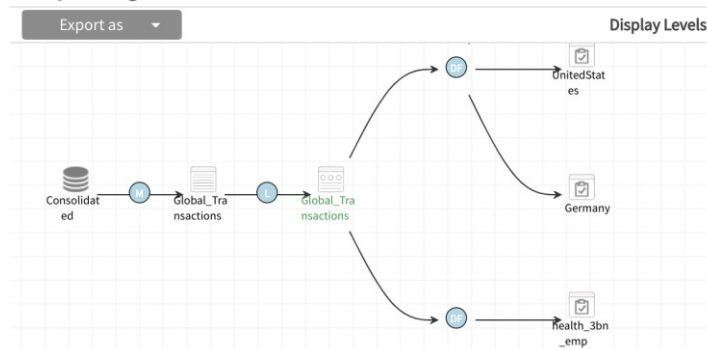
You are working on a project and want to create a dataset that will help others use the same data. Your user has permissions to create public datasets

- Click on the Catalog module

- From the catalog page with the tiles (or if you prefer a list view, click on the icon to the right of the filter button), add any five entities by clicking on the add to cart button .

- Click on the **My Cart** icon at the top choose Create Dataset
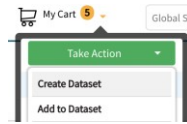
- Give your new dataset your login name (first initial+lastname) underscore dataset. (e.g. rginsburg_dataset), a description like "this was created for training," make it public and hit Save.

- Done! Now refresh your web page (it needs to reload the metadata to see it) and click on filter. Now filter using your new dataset.

**** DONE WITH CATALOG EXERCISES ****