Demo: Unified Data Catalog for Aircraft Images

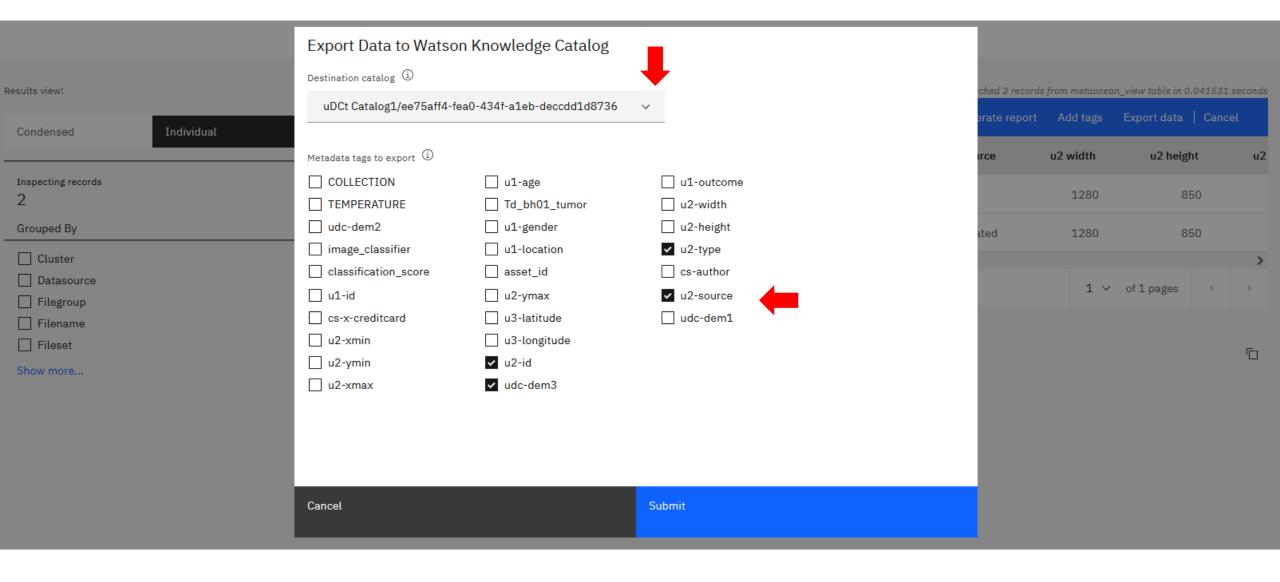## Scenario 3: Extend Catalog to Data & AI Workbench

Frank Lee, Joanna Wong, Reginald D'Souza

https://github.com/drfranknlee/demo-udc-aircraft

Working with Images

| 0 | NULL | T101389-10000 | T101389-10001 | T101389-10002 | T101389-10003 | T101389-10004 | T101389-10005 | T101389-10006 | T101389-10007 | T101389-10008 |

■ null ■ T101389-10000 ■ T101389-10001 ■ T101389-10002 ■ T101389-10003 ■ T101389-10004 ■ T101389-10005 ■ T101389-10006 ■ T101389-10007 ■ T101389-10008

Query builder

Visual query   SQL query

`udc-dem3 IN ('T101389-s2')`

Sizerange	0/4
State	0/2
Td_bh01_tumor	0/10
Temperature	0/1
Tier	0/2
Timesinceaccess	0/5
U2-id	0/12
U2-source	0/3
U2-type	0/10
Udc-dem2	0/4
Udc-dem3	1/2

1 item selected   Update query | Cancel

u2-id	count
☐ NULL	21,972
☐ T101389-10000	2
☐ T101389-10001	2
☑ T101389-10002	2
☐ T101389-10003	2
☐ T101389-10004	2
☐ T101389-10005	2
☐ T101389-10006	2
☐ T101389-10007	2
☐ T101389-10008	2

Results view:

Matched 2 records from metaocean_view table in 0.064482 seconds

Condensed	**Individual**

Inspecting records
2

Grouped By

☐ Cluster
☐ Datasource
☐ Filegroup
☐ Filename
☐ Fileset

Show more...

	path	filename	u2 id	udc dem3	u2 source	u2 width	u2 height	u2
☐	udc-vault/	T101389/dataset/raw/00b2add164cb42440a52064e390ea3d2.jpg	T101389-10002	T101389-s2	raw	1280	850	
☐	udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	T101389-10002	T101389-s2	annotated	1280	850	

Items per page: 10 ⌄   1–2 of 2 items   1 ⌄ of 1 pages ◁ ▷

Query builder

udc-dem3 IN ('T101389-s2') AND u2-id IN ('T101389-10002')

Select files to be Exported to Watson Knowledge Catalog (in Cloud Pak for Data)

Results view:

Condensed	Individual

Inspecting records

2

Grouped By

- ☐ Cluster
- ☐ Datasource
- ☐ Filegroup
- ☐ Filename
- ☐ Fileset

Show more...

Matched 2 records from metaocean_view table in 0.064482 seconds

2 items selected   Generate report   Add tags   Export data   Cancel

☑	path	filename	u2 id	udc dem3	u2 source	u2 width	u2 height	u2
☑	udc-vault/	T101389/dataset/raw/00b2add164cb42440a52064e390ea3d2.jpg	T101389-10002	T101389-s2	raw	1280	850	
☑	udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	T101389-10002	T101389-s2	annotated	1280	850	

Items per page:	10 ⌄	1–2 of 2 items		1 ⌄	of 1 pages	◁	▷

Query builder

```
udc-dem3 IN ('T101389-s2') AND u2-id IN ('T101389-10002')
```

Select Metadata Tags to be Exported to Watson Knowledge Catalog "uDCt Catalog1"

Export Data to Watson Knowledge Catalog

Destination catalog ⓘ

uDCt Catalog1/ee75aff4-fea0-434f-a1eb-deccdd1d8736 ⌄

Metadata tags to export ⓘ

- ☐ COLLECTION
- ☐ TEMPERATURE
- ☐ udc-dem2
- ☐ image_classifier
- ☐ classification_score
- ☐ u1-id
- ☐ cs-x-creditcard
- ☐ u2-xmin
- ☐ u2-ymin
- ☐ u2-xmax

- ☐ u1-age
- ☐ Td_bh01_tumor
- ☐ u1-gender
- ☐ u1-location
- ☐ asset_id
- ☐ u2-ymax
- ☐ u3-latitude
- ☐ u3-longitude
- ☑ u2-id
- ☑ udc-dem3

- ☐ u1-outcome
- ☐ u2-width
- ☐ u2-height
- ☑ u2-type
- ☐ cs-author
- ☑ u2-source
- ☐ udc-dem1

Cancel

Submit

Results view:

Condensed | Individual

Inspecting records
2

Grouped By

- ☐ Cluster
- ☐ Datasource
- ☐ Filegroup
- ☐ Filename
- ☐ Fileset

Show more...

_ched 2 records from metaocean_view table in 0.041531 seconds_

erate report | Add tags | Export data | Cancel

rce	u2 width	u2 height	u2
	1280	850	
ated	1280	850	

1 ⌄ of 1 pages

All

Search

Welcome, admin!

Discover services

Extend the functionality of the platform by installing services.
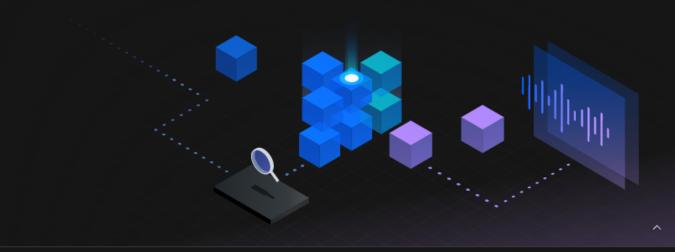
Browse the catalog

Add users

Connect to your LDAP server. Then, give users access to the platform.

Manage users

Stay informed

Monitor the services that are running and understand how you are using cluster resources.

Manage the platform

Quick navigation

All projects

Instances

Databases

Catalogs

Support

📄 Documentation

👥 Community

⌨ Support

💡 Share an idea
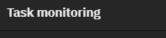
Overview

Recent projects

IBM-UDF-Testing	Yesterday at 11:37 PM

🌐 🌐 🍃 +1

Recent catalogs

Platform assets catalog	Today at 4:49 AM
Default Catalog	Today at 6:16 AM
uDCt Catalog1	Today at 6:27 AM

Requests

Data requests

0

Completed publish to catalog requests

0

Pending publish to catalog requests

1

Task inbox

Notifications
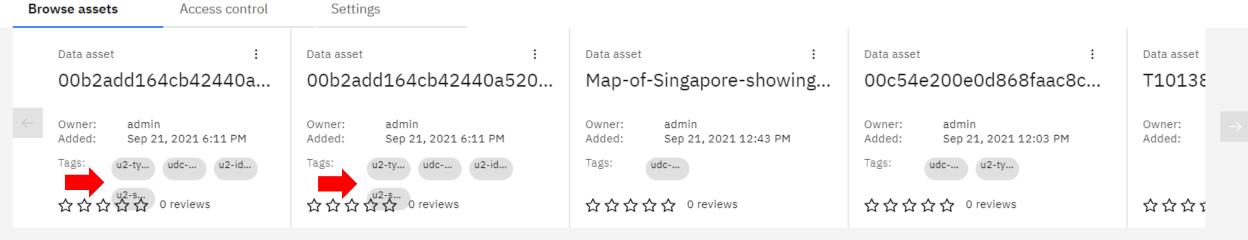
ⓘ No notifications

You will see your most recent notifications here.

Task monitoring

Total active tasks

0

Catalogs

Add to Catalog + ⌄

uDCt Catalog1

Browse assets   Access control   Settings

Data asset ⋮	Data asset ⋮	Data asset ⋮	Data asset ⋮	Data asset
00b2add164cb42440a...	00b2add164cb42440a520...	Map-of-Singapore-showing...	00c54e200e0d868faac8c...	T10138
Owner: admin	Owner: admin	Owner: admin	Owner: admin	Owner:
Added: Sep 21, 2021 6:11 PM	Added: Sep 21, 2021 6:11 PM	Added: Sep 21, 2021 12:43 PM	Added: Sep 21, 2021 12:03 PM	Added:
Tags: u2-ty... udc-... u2-id... u2-s...	Tags: u2-ty... udc-... u2-id... u2-s...	Tags: udc-...	Tags: udc-... u2-ty...	Tags:
☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews	☆☆☆☆

Showing 9 of 9 items

☐	Name ⌃⌄	Owner	Tags	Business terms	Type	Date added	
☐	00b2add164cb42440a52064e390ea3...	Ⓐ admin	u2-typ... udc-de... + 2 more		Data asset	Sep 21, 2021	⋮
☐	00b2add164cb42440a52064e390ea3...	Ⓐ admin	u2-typ... udc-de... + 2 more		Data asset	Sep 21, 2021	⋮

Catalogs / uDCt Catalog1

Add to Catalog + ⌄

Data asset

00b2add164cb42440a52064e390ea3d2.jpg ✎

Remove   **Add to project +**

| **Overview** | Asset | Access | Review | Activities |

Added: Sep 21, 2021 6:11 PM
Format: Image ✎

Tags ✎

u2-id:T101389-10002   u2-source:raw   u2-type:B1

udc-dem3:T101389-s2

Reviews

☆☆☆☆☆   0 reviews

Connection
Source: udc_tucson_icos
Source type: Cloud Object Storage (infrastr...

⌃ General

Description ✎

No description added yet.

Path

udc-vault / T101389 / dataset / raw / 00b2add164cb42440a52064e390ea3d2.jpg
/

⌃ Governance artifacts

Business terms ⊕

No business terms added yet.

Classifications ⊕

No classifications added yet.

⌃ Related assets

Show: Any relationship ⌄   Add asset +

Catalogs / uDCt Catalog1

Add to Catalog + ⌄

Data asset

00b2add164cb42440a52064e390ea3d2.jpg ✎

Remove   **Add to project +**

Overview | Asset | Access | Review | Activities

Added: Sep 21, 2021 6:11 PM
Format: Image ✎

Tags ✎
u2-id:T101389-10002   u2-source:annotated
u2-type:B1   udc-dem3:T101389-s2

Reviews
☆ ☆ ☆ ☆ ☆   0 reviews

Connection
Source: udc_tucson_icos
Source type: Cloud Object Storage (infrastr...

⌃ General

Description ✎

No description added yet.

Path

udc-vault / T101389 / dataset / annotated /
00b2add164cb42440a52064e390ea3d2.jpg /

⌃ Governance artifacts

Business terms ⊕

No business terms added yet.

Classifications ⊕

No classifications added yet.

⌃ Related assets

Show: Any relationship ⌄   Add asset +

Relationship	Asset name	Asset type

Catalogs / uDCt Catalog1

Add to Catalog

Data asset

00b2add164cb42440a52064e390ea3d2.jpg

Remove   Add to project
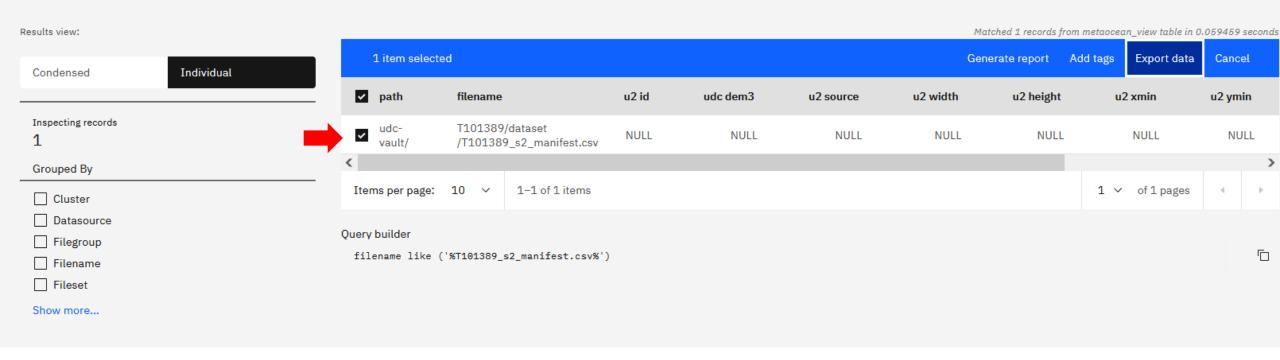
Overview   Asset   Access   Review   Activities

Preview: Image

Query builder

Visual query   **SQL query**

Select * from Spectrum Discover where...

filename like ('%T101389_s2_manifest.csv%')

Group by

| 1 × | Groups selected | ⌄ |

Limit results

10000 ↕

Setting a lower limit may increase speed of results.

Sort b

e.g

Set a

Reset query   **View results**

Working with Manifest

Query Results

Results view:

Condensed	Individual

Inspecting records
1

Grouped By

☐ Cluster
☐ Datasource
☐ Filegroup
☐ Filename
☐ Fileset

Show more...

Matched 1 records from metaocean_view table in 0.059459 seconds

			1 item selected			Generate report	Add tags	**Export data**	Cancel

☑	path	filename	u2 id	udc dem3	u2 source	u2 width	u2 height	u2 xmin	u2 ymin
☑	udc-vault/	T101389/dataset/T101389_s2_manifest.csv	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Items per page: 10 ⌄   1–1 of 1 items   1 ⌄   of 1 pages   ◀ ▶

Query builder

```
filename like ('%T101389_s2_manifest.csv%')
```

Query Results

Results view:

Condensed	Individual

Inspecting records
1

Grouped By

- [ ] Cluster
- [ ] Datasource
- [ ] Filegroup
- [ ] Filename
- [ ] Fileset

Show more...

Export Data to Watson Knowledge Catalog

Destination catalog ⓘ

➡️ uDCt Catalog1/ee75aff4-fea0-434f-a1eb-deccdd1d8736 ⌄

Metadata tags to export ⓘ

- [ ] COLLECTION
- [ ] TEMPERATURE
- [ ] udc-dem2
- [ ] u1-gender
- [ ] u1-outcome
- [ ] Td_bh01_tumor
- [ ] u1-location
- [ ] asset_id
- [ ] u2-width
- [ ] u2-type

- [ ] u2-height
- [ ] cs-author
- [ ] cs-x-creditcard
- [ ] u2-ymin
- [ ] u2-ymax
- [ ] u1-id
- [ ] u2-xmin
- [ ] u2-xmax
- [ ] u3-latitude
- [ ] u3-longitude

- [ ] u2-id
- [x] udc-dem3
- [ ] image_classifier
- [ ] udc-dem1
- [ ] u1-age
- [ ] u2-source
- [ ] classification_score

Cancel	Submit

Matched 1 records from metaocean_view table in 0.059459 seconds

Generate report	Add tags	Export data	Cancel

u2 height	u2 xmin	u2 ymin
NULL	NULL	NULL

1 ⌄ of 1 pages ◄ ►

Catalogs

Add to Catalog + ⌄

uDCt Catalog1

Browse assets   Access control   Settings

| Recently added | Watson recommends | Highly rated | | Collapse |

Data asset ⋮	Data asset ⋮	Data asset ⋮	Data asset ⋮	Data as
T101389_s2_manifest....	00b2add164cb42440a520...	00b2add164cb42440a520...	Map-of-Singapore-showing...	00c5
Owner: admin	Owner: admin	Owner: admin	Owner: admin	Owner:
Added: Sep 21, 2021 6:27 PM	Added: Sep 21, 2021 6:11 PM	Added: Sep 21, 2021 6:11 PM	Added: Sep 21, 2021 12:43 PM	Added:
	Tags: u2-ty... udc-... u2-id... u2-s...	Tags: u2-ty... udc-... u2-id... u2-s...	Tags: udc-...	Tags:
☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews	☆☆☆☆☆ 0 reviews	☆☆

Showing 10 of 10 items

☐	Name	Owner	Tags	Business terms	Type	Date added ⇕	
☐	T101389_s2_manifest.csv	Ⓐ admin			Data asset	Sep 21, 2021	⋮
☐	00b2add164cb42440a52064e390ea...	Ⓐ admin	u2-typ... udc-d... + 2 more		Data asset	Sep 21, 2021	⋮
☐	00b2add164cb42440a52064e390ea...	Ⓐ admin	u2-typ... udc-d... + 2 more		Data asset	Sep 21, 2021	⋮

Catalogs / uDCt Catalog1   Add to Catalog + ∨

Data asset

T101389_s2_manifest.csv ✎   Remove   **Add to project   +**

| Overview | Asset | Access | Review | Profile | Activities |

Added: Sep 21, 2021 6:27 PM

Format: CSV ✎

Tags ✎

There are no tags available for this asset.

Reviews

☆ ☆ ☆ ☆ ☆   0 reviews

Connection

Source: udc_tucson_icos

Source type: **Cloud Object Storage (infr...**

⌃ **General**

Description ✎

No description added yet.

Path

udc-vault / T101389 / dataset / T101389_s2_manifest.csv /

⌃ **Governance artifacts**

Business terms ⊕

No business terms added yet.

Classifications ⊕

No classifications added yet.

Catalogs / uDCt Catalog1

Add to Catalog + ⌄

Data asset

T101389_s2_manifest.csv ✎

Remove   **Add to project   +**

| Overview | **Asset** | Access | Review | Profile | Activities |

Schema: 9 Columns
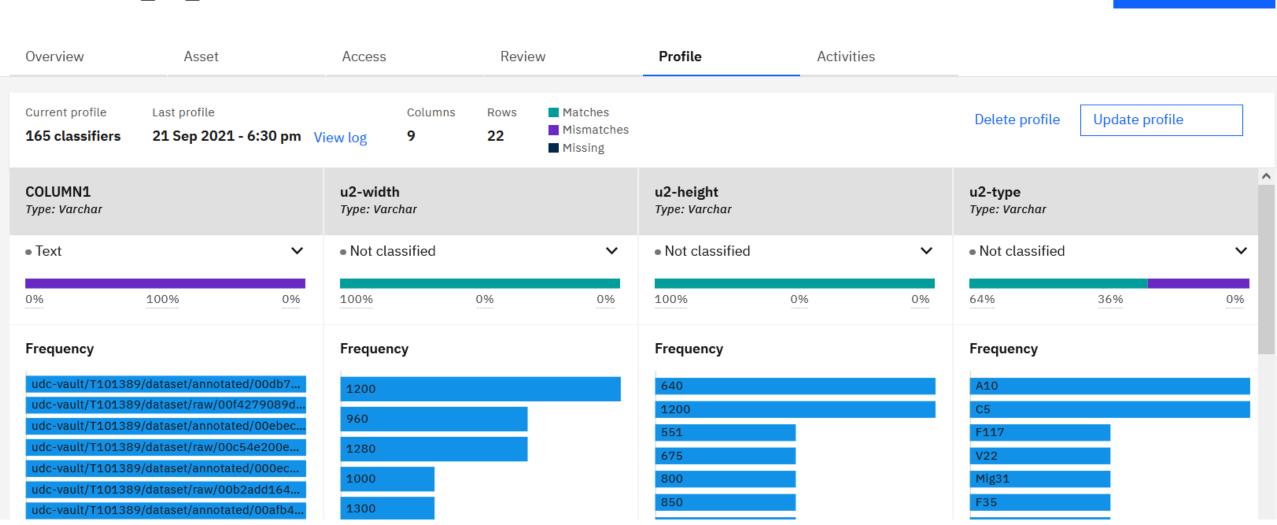Preview: 22 rows

Last refresh: just now ⟳ ⤓

COLUMN1 ⊙ String	u2-width ⊙ String	u2-height ⊙ String	u2-type ⊙ String	u2-xmin ⊙ String	u2-ymin ⊙ String	u2-xmax ⊙ String	u2-ymax ⊙ String	u2-id ⊙ String
udc-vault/T101389/	1000	551	Be200	308	126	651	201	T101389-10000
udc-vault/T101389/	1200	800	Mig31	299	197	1102	569	T101389-10001
udc-vault/T101389/	1280	850	B1	322	112	893	618	T101389-10002
udc-vault/T101389/	1300	867	F35	607	204	1202	529	T101389-10003
udc-vault/T101389/	1800	1200	A10	32	60	1800	766	T101389-10004
udc-vault/T101389/	1200	1200	V22	647	521	973	768	T101389-10005
udc-vault/T101389/	1200	675	A10	560	308	647	329	T101389-10006
udc-vault/T101389/	1280	857	F117	176	263	1280	553	T101389-10007
udc-vault/T101389/	960	640	C5	58	122	880	397	T101389-10008

Catalogs / uDCt Catalog1

Add to Catalog + ⌄

Data asset

T101389_s2_manifest.csv ✎

Remove   **Add to project +**

Overview	Asset	Access	Review	**Profile**	Activities

Current profile	Last profile		Columns	Rows	■ Matches	Delete profile	**Update profile**
165 classifiers	**21 Sep 2021 - 6:30 pm** View log		**9**	**22**	■ Mismatches ■ Missing		

COLUMN1 Type: Varchar	**u2-width** Type: Varchar	**u2-height** Type: Varchar	**u2-type** Type: Varchar
● Text ⌄	● Not classified ⌄	● Not classified ⌄	● Not classified ⌄
0% 100% 0%	100% 0% 0%	100% 0% 0%	64% 36% 0%
Frequency	**Frequency**	**Frequency**	**Frequency**
udc-vault/T101389/dataset/annotated/00db7...	1200	640	A10
udc-vault/T101389/dataset/raw/00f4279089d...	960	1200	C5
udc-vault/T101389/dataset/annotated/00ebec...	1280	551	F117
udc-vault/T101389/dataset/raw/00c54e200e...	1000	675	V22
udc-vault/T101389/dataset/annotated/000ec...	1300	800	Mig31
udc-vault/T101389/dataset/raw/00b2add164...		850	F35
udc-vault/T101389/dataset/annotated/00afb4...			

Catalogs / uDCt Catalog1

Successfully added your review. ✕

Data asset

T101389_s2_manifest.csv ✎

Remove   Add to project   +

Overview   Asset   Access   **Review**   Profile   Activities

Overall rating

5.0

★ ★ ★ ★ ★   1 review

Review summary

5 ▓▓▓▓▓▓ (1)

4 (0)

3 (0)

2 (0)

1 (0)

Your review

admin   Sep 21, 2021

★ ★ ★ ★ ★

Created the manifest for a demo

Catalogs / uDCt Catalog1   Add to Catalog   +   ⌄

Data asset
T101389-aircraft-manifest-scenario2.csv ✎   Remove   **Add to project** +

Overview	Asset	Access	Review	Profile	Activities
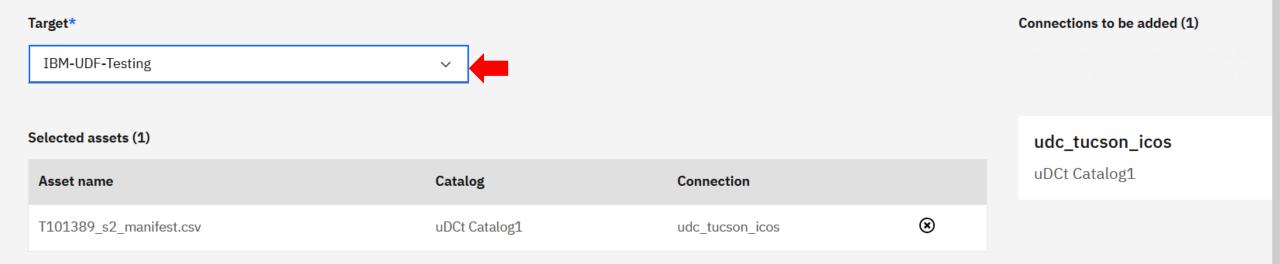
Schema: 9 Columns
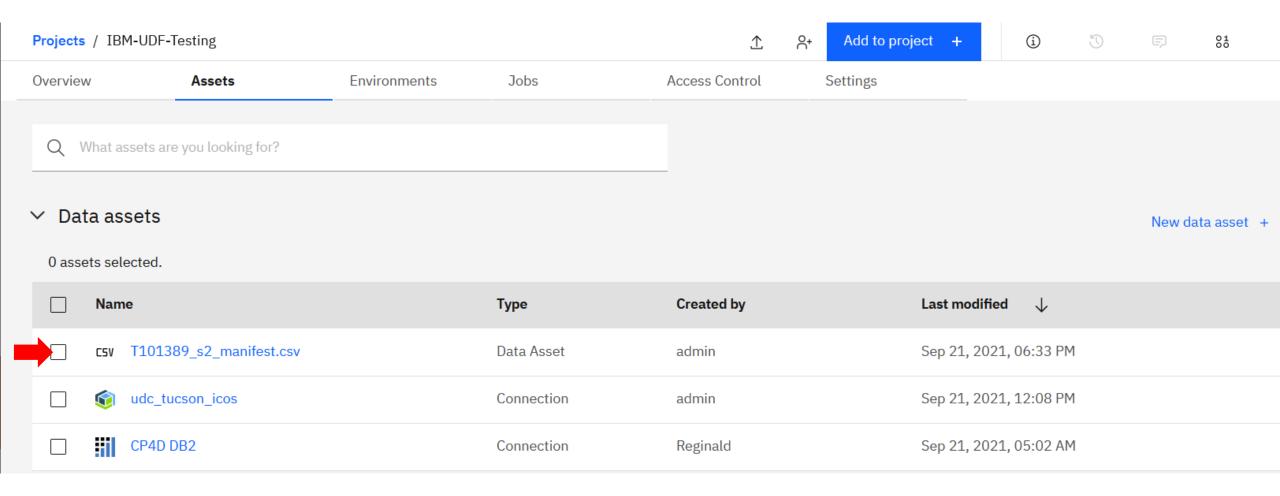Preview: 22 rows   Last refresh: 37 seconds ago   ↻   ↓

COLUMN1 👁 String	u2-width 👁 String	u2-height 👁 String	u2-type 👁 String	u2-xmin 👁 String	u2-ymin 👁 String	u2-xmax 👁 String	u2-ymax 👁 String	u2-id 👁 String
udc-vault/T101389/scen	1000	551	Be200	308	126	651	201	T101389-10000
udc-vault/T101389/scen	1200	800	Mig31	299	197	1102	569	T101389-10001
udc-vault/T101389/scen	1280	850	B1	322	112	893	618	T101389-10002
udc-vault/T101389/scen	1300	867	F35	607	204	1202	529	T101389-10003
udc-vault/T101389/scen	1800	1200	A10	32	60	1800	766	T101389-10004
udc-vault/T101389/scen	1200	1200	V22	647	521	973	768	T101389-10005
udc-vault/T101389/scen	1200	675	A10	560	308	647	329	T101389-10006
udc-vault/T101389/scen	1280	857	F117	176	263	1280	553	T101389-10007
udc-vault/T101389/scen	960	640	C5	58	133	889	397	T101389-10008
udc-vault/T101389/scen	2048	1365	C5	875	398	1110	629	T101389-10009
udc-vault/T101389/scen	960	640	Tu160	1	155	903	506	T101389-10010

Catalogs

Add to project

Target*

IBM-UDF-Testing   ⌄   ⬅

Connections to be added (1)

Selected assets (1)

Asset name	Catalog	Connection	
T101389_s2_manifest.csv	uDCt Catalog1	udc_tucson_icos	⊗

udc_tucson_icos

uDCt Catalog1

Cancel   Add

Projects / IBM-UDF-Testing

Add to project +

| Overview | **Assets** | Environments | Jobs | Access Control | Settings |

Q What assets are you looking for?

⌄ Data assets

New data asset +

0 assets selected.

	Name	Type	Created by	Last modified ↓
☐	CSV T101389_s2_manifest.csv	Data Asset	admin	Sep 21, 2021, 06:33 PM
☐	📦 udc_tucson_icos	Connection	admin	Sep 21, 2021, 12:08 PM
☐	⊞ CP4D DB2	Connection	Reginald	Sep 21, 2021, 05:02 AM

Projects / **IBM-UDF-Testing** / T101389_s2_manifest.csv

Preview   Profile   Activities

Schema: 9 Columns | 22 rows

Preview: 22 rows   Last refresh: just now   **Refine**

COLUMN1 String	u2-width String	u2-height String	u2-type String	u2-xmin String	u2-ymin String	u2-xmax String	u2-ymax String	u2-id String
Text ⌄	Not classif... ⌄	Not classif... ⌄	Not classif... ⌄	Not classif... ⌄	Not classif... ⌄	Not classif... ⌄	Not classif... ⌄	Not classif... ⌄
udc-vault/T101389/c	1000	551	Be200	308	126	651	201	T101389-10000
udc-vault/T101389/c	1200	800	Mig31	299	197	1102	569	T101389-10001
udc-vault/T101389/c	1280	850	B1	322	112	893	618	T101389-10002
udc-vault/T101389/c	1300	867	F35	607	204	1202	529	T101389-10003
udc-vault/T101389/c	1800	1200	A10	32	60	1800	766	T101389-10004
udc-vault/T101389/c	1200	1200	V22	647	521	973	768	T101389-10005
udc-vault/T101389/c	1200	675	A10	560	308	647	329	T101389-10006
udc-vault/T101389/c	1280	857	F117	176	263	1280	553	T101389-10007
udc-vault/T101389/c	960	640	C5	58	133	889	397	T101389-10008
udc-vault/T101389/c	2048	1365	C5	875	398	1110	629	T101389-10009
udc-vault/T101389/c	960	640	Tu160	1	155	903	506	T101389-10010

Projects / IBM-UDF-Testing / T101389-aircraft-manifest-scen... / Refine data

Operation +

Code an operation to cleanse and shape your data

Data   Profile   **Visualizations**

RELATIONSHIP CHART

CHART TYPE   Scatter plot   Line   Multi-series   Histogram   Population ...   Q-Q plot   Pie

ACTIONS ⌄

Columns*

👥 u2-type ⌄

👥 u2-id ⌄

Add another column +

Line style ⓘ
⦿ Curved
○ Straight

Label threshold ⓘ

20

Legend orient ⓘ
○ Horizontal
○ Vertical
⦿ Vertical bottom

Primary title ⓘ

u2-type

A10
4

■ u2-type
■ u2-id

Projects / IBM-UDF-Testing / T101389-aircraft-manifest-scen... / Refine data

Operation +

Code an operation to cleanse and shape your data

Data   Profile   **Visualizations**

HISTOGRAM CHART

CHART TYPE   Scatter plot   Line   Multi-series   Histogram   Population ...   Q-Q plot   Pie

ACTIONS

X-axis*

📏 u2-height

Split by ⓘ

♣ u2-type

Stacked ⓘ

◯ Stacked

◉ Panelled

Bin Method ⓘ

Auto bin

Show distribution curve ⓘ

🟢 On

Bin width ⓘ

1 ———●——— 20   5

u2-type
🟪 Histogram
🔵 A10 Distri...
🟢 C5 Distrib...