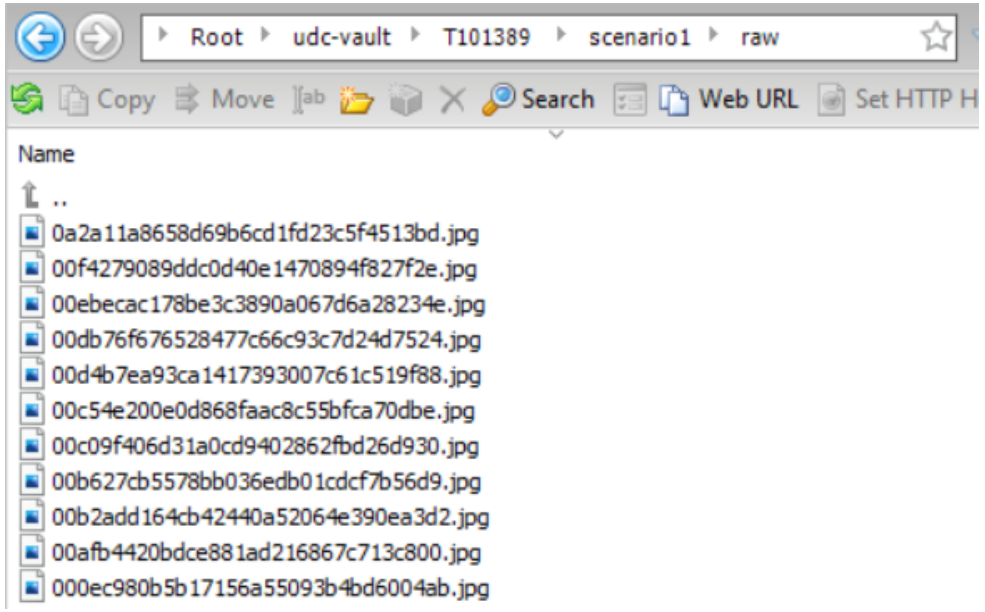
Demo: Unified Data Catalog for Military Aircraft

Scenario 1: Initialize Catalog with Metadata

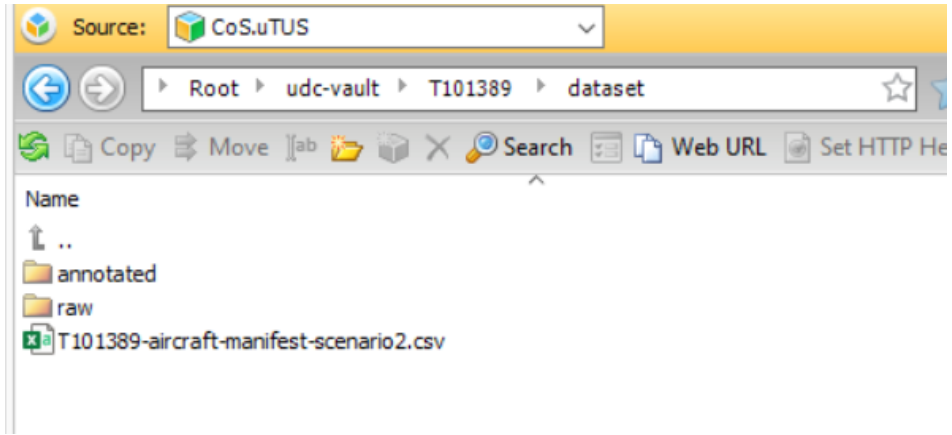
Frank Lee

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

Upload Dataset to Cloud Object Storage Bucket



Upload Manifest to Cloud Object Storage Bucket



Search for Dataset using Query Builder

Query builder

Visual query | **SQL query**

Select * from Spectrum Discover where...

```
filename LIKE ('%T101389%') AND filetype LIKE ('%jpg%')
```

Group by

Choose groupings

Limit results

10000

Setting a lower limit may increase speed of results.

Sort by

e.g. owner asc

Set a sort order to choose which results are returned first.

Reset query | View results

Query Results

Results view:

Condensed | Individual

Inspecting records
22

Grouped By

- Cluster
- Datasource
- Filegroup
- Filename
- Fileset

Show more...

Matched 22 records from metaocean_view table in 0.084276 seconds

path	filename	udc dem2	u2 type	u2 id	udc dem3	u2 width
udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	NULL	NULL	NULL	NULL	NULL
udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	NULL	NULL	NULL	NULL	NULL
udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	NULL	NULL	NULL	NULL	NULL
udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdcf7b56d9.jpg	NULL	NULL	NULL	NULL	NULL
udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	NULL	NULL	NULL	NULL	NULL
udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	NULL	NULL	NULL	NULL	NULL
udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	NULL	NULL	NULL	NULL	NULL
udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	NULL	NULL	NULL	NULL	NULL

Items per page: 10 | 1–10 of 22 items | 1 of 3 pages

Create AutoTag Policy

Create the autotag policy in json

```
{
"pol_id": "T101389_autotag_u2source_pol",
"action_id": "AUTOTAG",
"action_params": {
    "rule": {
    "name":"setFromExistingField",
    "action": "regex",
    "existingField": "filename",
    "regexPattern": "[^/]+",
    "matchNo": 3,
    "newField": "u2-source"
    }
},
"schedule": "NOW",
"pol_filter": "filename LIKE ('%T101389%') AND filetype LIKE ('%jpg%')",
"pol_state": "active",
"explicit": "true",
"collections": []
}
```

Execute the policy on server

```
[moadmin@discover1 T101389]$ pwd
/home/moadmin/T101389
[moadmin@discover1 T101389]$ ls
T101389_aircraft_import_scenario1.json  T101389_autotag_u2source.json
T101389_aircraft_import_scenario2.json
[moadmin@discover1 T101389]$
[moadmin@discover1 T101389]$ gettoken
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
[moadmin@discover1 T101389]$
[moadmin@discover1 T101389]$
[moadmin@discover1 T101389]$ curl -k -H "Authorization: Bearer ${TOKEN}" https://localhost/policyengine/v1/
policies/T101389_autotag_u2source_pol -X POST -d @./T101389_autotag_u2source.json -H "Content-Type: applica
tion/json"
Policy 'T101389_autotag_u2source_pol' added[moadmin@discover1 T101389]$
```

Create AutoTag Policy: Result

Tag management

Tags | **Policies** | Policy history

Add Policy +

Policy	Type	Schedule (UTC)	State	State	Collections	Last modified by	Last modified
ap_cs_body	CONTENTSEARCH	Done	Active	Complete		sdadmin	9/17/2021, 2:51:16 PM
ap_cs_meta	CONTENTSEARCH	Done	Active	Complete		sdadmin	9/16/2021, 1:18:06 PM
T101389_autotag_u2source_pol	AUTOTAG	Done	Active	Complete		sdadmin	9/21/2021, 2:32:38 PM
u3-iphone-gps	CONTENTSEARCH	Done	Active	Complete		sdadmin	9/17/2021, 12:35:42 AM

Items per page: 20 | 1–4 of 4 items | 1 of 1 pages

Create AutoTag Policy: Result

Tag management

Tags | Policies | **Policy history**

Policy name	Type	Started	Ended	Status	Total records	Successful	Skipped	Failed
T101389_autotag_u2source_pol	AUTOTAG	9/21/2021, 2:32:39 PM	9/21/2021, 2:32:41 PM	Complete	22	22	0	0
T101389_contentsearch_u2source_pol	AUTOTAG	9/21/2021, 2:27:44 PM	9/21/2021, 2:27:46 PM	Complete	22	22	0	0
T101389_aircraft_import_scenario2_pol	IMPORT_TAGS	9/21/2021, 1:03:52 AM	9/21/2021, 1:04:11 AM	Complete	22	22	0	0
T101389_aircraft_import_scenario1_pol	IMPORT_TAGS	9/20/2021, 11:21:32 PM	9/20/2021, 11:21:49 PM	Complete	11	11	0	0
T101389_aircraft_import_scenario1_pol	IMPORT_TAGS	9/20/2021, 11:17:06 PM	9/20/2021, 11:17:22 PM	Complete	11	11	0	0
T101389_aircraft_import_pol	IMPORT_TAGS	9/20/2021, 9:18:41 PM	9/20/2021, 9:18:59 PM	Complete	11	11	0	0
T101389_aircraft_import_pol	IMPORT_TAGS	9/20/2021, 9:16:06 PM	9/20/2021, 9:16:13 PM	Complete	0	0	0	0
T101389_aircraft_import_pol	IMPORT_TAGS	9/20/2021, 9:08:05 PM	9/20/2021, 9:08:40 PM	Complete	0	0	0	0
ap_cs_body	CONTENTSEARCH	9/17/2021,	9/17/2021,	Complete	16	16	0	0

Create AutoTag Policy: Result

Results view:

Condensed | Individual

Inspecting records
22

Grouped By

- Cluster
- Datasource
- Filegroup
- Filename
- Fileset

Show more...

Matched 22 records from metaocean_view table in 0.082699 seconds

path	filename	u2 source
udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	annotated
udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	annotated
udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	annotated
udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdcf7b56d9.jpg	annotated
udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	annotated
udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	annotated
udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	annotated
udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	annotated
udc-vault/	T101389/dataset/annotated/00ebecac178be3c3890a067d6a28234e.jpg	annotated
udc-vault/	T101389/dataset/annotated/00f4279089ddc0d40e1470894f827f2e.jpg	annotated
udc-vault/	T101389/dataset/annotated/0a2a11a8658d69b6cd1fd23c5f4513bd.jpg	annotated
udc-vault/	T101389/dataset/raw/000ec980b5b17156a55093b4bd6004ab.jpg	raw
udc-vault/	T101389/dataset/raw/00afb4420bdce881ad216867c713c800.jpg	raw
udc-vault/	T101389/dataset/raw/00b2add164cb42440a52064e390ea3d2.jpg	raw
udc-vault/	T101389/dataset/raw/00b627cb5578bb036edb01cdcf7b56d9.jpg	raw
udc-vault/	T101389/dataset/raw/00c09f406d31a0cd9402862fbd26d930.jpg	raw

Items per page: 50 | 1–22 of 22 items | 1 of 1 pages

Query builder

```
filename LIKE ('%T101389%') AND filetype LIKE ('%jpg%')
```

Custom-Tag Dataset

Results view:

Condensed | Individual

Inspecting records
22

Grouped By

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

Show more...

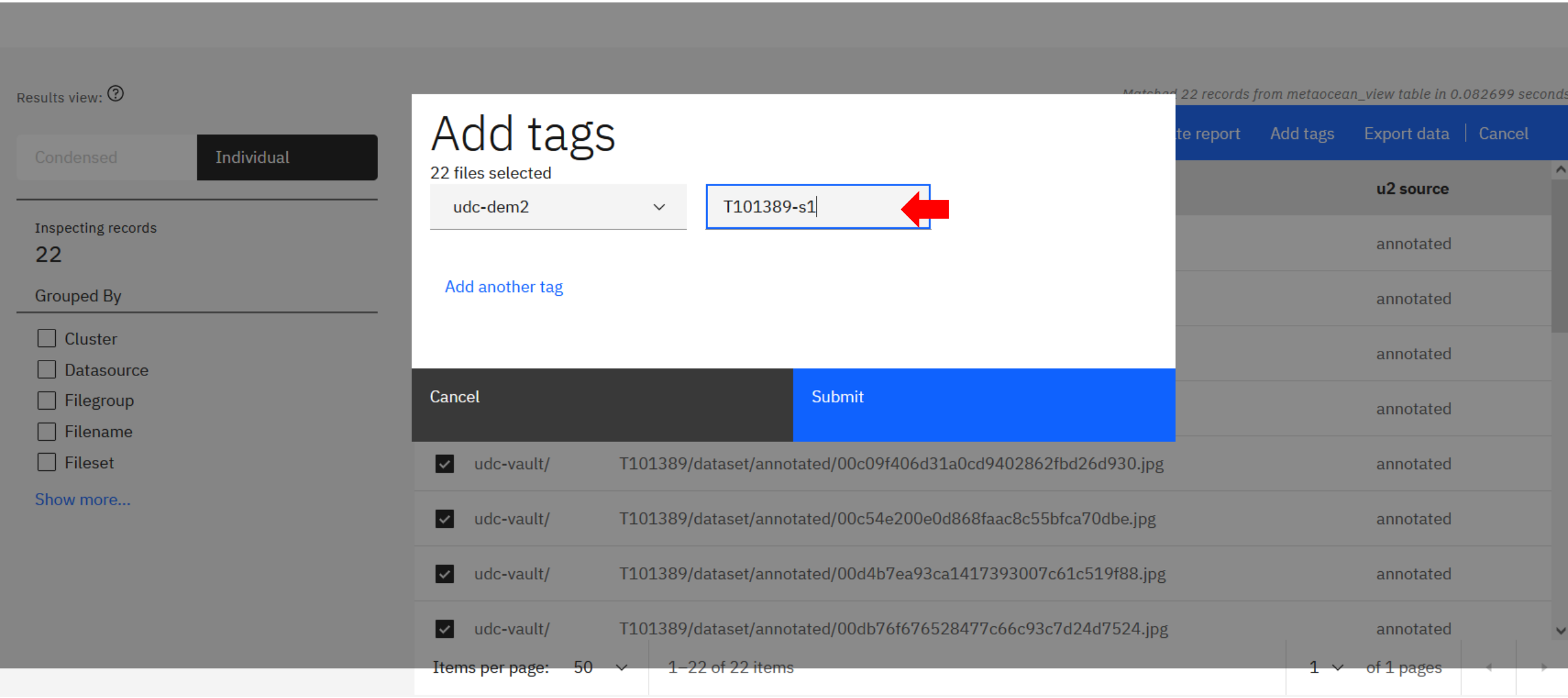
Matched 22 records from m_auocean_view table in 0.082699 seconds

22 items selected | Generate report | Add tags | Export data | Cancel

path	filename	u2 source
udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	annotated
udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	annotated
udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	annotated
udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdcf7b56d9.jpg	annotated
udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	annotated
udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	annotated
udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	annotated
udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	annotated

Items per page: 50 | 1–22 of 22 items | 1 of 1 pages

Custom-Tag Dataset



Custom-Tag Dataset with T101389-s1

Results view:

Condensed | Individual

Inspecting records
22

Grouped By

- Cluster
- Datasource
- Filegroup
- Filename
- Fileset

Show more...

Matched 22 records from metaocean_view table in 0.05813 seconds

path	filename	udc dem2	u2 source
udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdcf7b56d9.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	T101389-s1	annotated

Items per page: 10 | 1–10 of 22 items | 1 of 3 pages

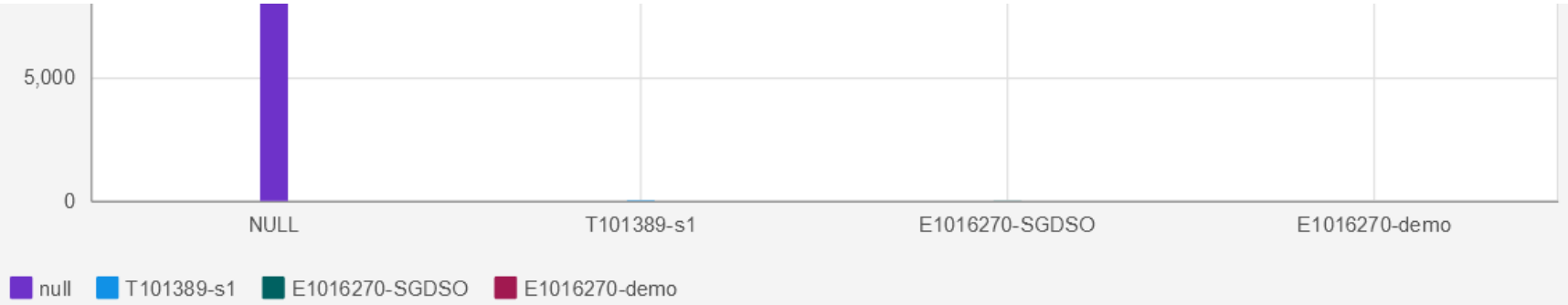
Find the Newly Tagged Dataset with Visual Query

Query builder

Visual query | SQL query

```
Query preview...
```

Temperature	0/1
Tier	0/2
Timesinceaccess	0/5
U2-id	0/1
U2-source	0/3
U2-type	0/1
Udc-dem2	0/4
Udc-dem3	0/1



udc-dem2	count
NULL	21,958
T101389-s1	22
E1016270-SGDSO	8
E1016270-demo	1

Items per page: 10 | 1–4 of 4 items | 1 of 1 pages

New Dataset (T101389-s1) Can be Found using a Single-query Search

Query Results

Results view:

Condensed | Individual

Displaying records
22

Condensed results
1

Grouped By

Udc-dem2

Show more...

Grouped 22 results from mrcapacity table in 0.268984 seconds

udc dem2	Total Files	Total Size
T101389-s1	22	5.33 MiB

Items per page: 10 | 1–1 of 1 items | 1 of 1 pages

Query builder

```
udc-dem2 IN ('T101389-s1')
```

Export the New Dataset into a Manifest

Query Results

Results view:

Condensed | Individual

Inspecting records
22

Grouped By

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

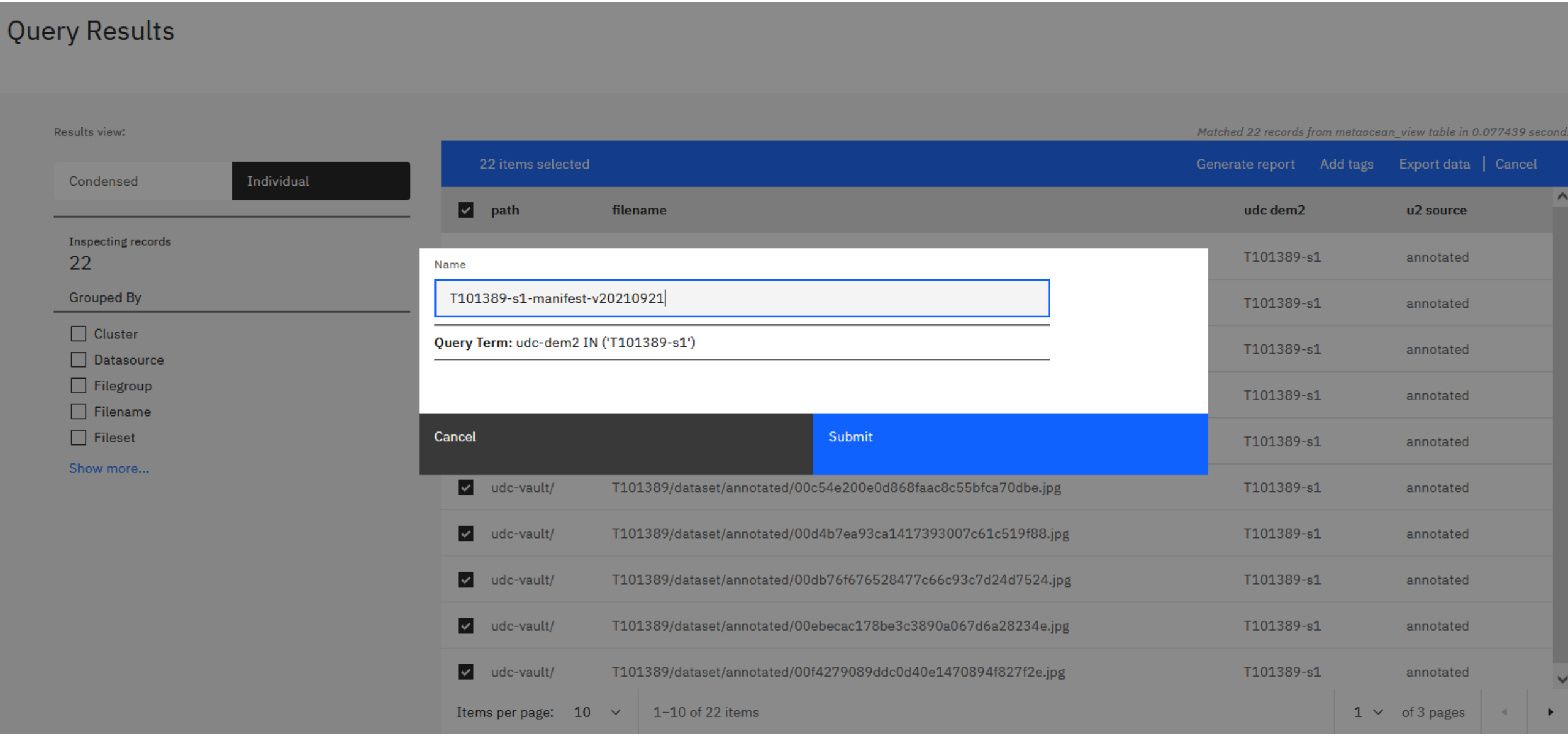
Show more...

Matched 22 records from metaocean_view table in 0.077439 seconds

22 items selected | Generate report | Add tags | Export data | Cancel

path	filename	udc dem2	u2 source
udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdcf7b56d9.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	T101389-s1	annotated
udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	T101389-s1	annotated

Export the New Dataset into a Manifest



Export the New Dataset into a Manifest

Reports

Report has been Created
The report T101389-s1-manifest-v20210921 was created

Report	Last run	Duration	Status	Output size
T101389-s1-manifest-v20210921	Tue Sep 21 2021 14:49:17 GMT-0500 (Central Daylight Time)	0 seconds	Complete	10.32 KiB
demo-csv	Thu Sep 09 2021 18:39:49 GMT-0500 (Central Daylight Time)	0 seconds	Complete	2.34 KiB
E101475_d1-list	Thu Apr 29 2021 08:16:33 GMT-0500 (Central Daylight Time)	0 seconds	Complete	8.17 KiB
E1014753-data-manifest	Thu Apr 29 2021 00:36:52 GMT-0500 (Central Daylight Time)	0 seconds	Complete	8.08 KiB
BH01 Dataset Report for FS21EMEA	Tue Feb 09 2021 00:14:30 GMT-0600 (Central Standard Time)	0 seconds	Complete	3.75 MiB
TD-BH01 Report	Sat Jan 30 2021 18:15:17 GMT-0600 (Central Standard Time)	0 seconds	Complete	3.73 MiB
BreakHist Summary	Sat Jan 30 2021 17:56:08 GMT-0600 (Central Standard Time)	0 seconds	Complete	3.33 MiB
Age Report Detail. 90-180 Days	Tue Dec 22 2020 02:24:47 GMT-0600 (Central Standard Time)	0 seconds	Complete	308 Bytes
Age Report Summary. 90-180 Days	Tue Dec 22 2020 02:24:47 GMT-0600 (Central Standard Time)	0 seconds	Complete	41 Bytes
Age Report Detail. 720+ Days	Tue Dec 22 2020 02:24:46 GMT-0600 (Central Standard Time)	0 seconds	Complete	308 Bytes

Items per page: 20 | 1–20 of 21 items | 1 of 2 pages

Export the New Dataset into a Manifest

A1 path

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	path	filename	filetype	datasource	owner	group	revision	site	platform	cluster	inode	permissior	fileset	uid	gid	recordvers	state	migloc	mtime	atime	ctime	tier	size
2	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:03.0	08:03.0			100863
3	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:02.9	08:02.9			185155
4	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:03.3	08:03.3			109370
5	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:03.5	08:03.5			100985
6	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:04.1	08:04.1			276382
7	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:04.5	08:04.5			372057
8	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:04.8	08:04.8			76399
9	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:05.0	08:05.0			69311
10	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:05.4	08:05.4			231879
11	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:05.7	08:05.7			174566
12	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:06.1	08:06.1			150729
13	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								07:57.6	07:57.6			57205
14	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								07:58.4	07:58.4			595383
15	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								07:58.1	07:58.1			66842
16	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								07:58.7	07:58.7			124600
17	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:01.1	08:01.1			1933439
18	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								07:59.9	07:59.9			284812
19	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:00.8	08:00.8			46241
20	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:01.3	08:01.3			25959
21	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:01.7	08:01.7			215413
22	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:02.3	08:02.3			329510
23	udc-vault/	T101389/c	jpg	udc-vault				Tucson	IBM COS	df3b7ae7-	0								08:02.4	08:02.4			62207
24																							
25																							
26																							