#### Onboarding to OS-Climate

- Obtain a GitHub Account
  - Make yourself an access token
- Request an invitation to the os-climate GitHub organization
  - Particular os-climate team or teams
  - odh-env-users team for Open Data Hub / OpenShift tools
- Obtain a Trino Account
  - File an issue against <a href="https://github.com/os-climate/os c data commons">https://github.com/os-climate/os c data commons</a>
  - Or send an email to eje@redhat.com
  - Trino account name will be same as your GitHub account name
  - We send you a random password
- JupyterHub
  - https://jupyterhub-odh-jupyterhub.apps.odh-cl1.apps.os-climate.org/
- SuperSet
  - https://superset-secure-odh-superset.apps.odh-cl1.apps.os-climate.org/
- Example Jupyter Notebooks
  - https://github.com/os-climate/data-platform-demo/tree/master/notebooks

quay.io GitHub Elyra jupyter aicoe-ci jupyter itr-ingest-raw-em.. jupyter esg\_data\_ingesti... o | itr-ingest-raw-se...

Elyra pipelines

[10]: # a way to examine the structure of a pandas data frame df.info(verbose=True)

> <class 'pandas.core.frame.DataFrame'> RangeIndex: 35 entries, 0 to 34 Data columns (total 16 columns):

Column Non-Null Count Dtype 31 non-null string company name company id 31 non-null string isic 0 non-null Int64 country 31 non-null string region 31 non-null strina 0 non-null industry level 1 Int64 industry level 2 0 non-null Int64 industry level 3 0 non-null Int64 industry level 4 0 non-null Int64 31 non-null string sector 10 company\_revenue 30 non-null Float64 30 non-null Float64 company market cap 12 company enterprise value 30 non-null Float64 30 non-null Float64 13 company total assets 14 company cash equivalents 30 non-null Float64 15 target probability 30 non-null Float64 dtypes: Float64(6), Int64(5), string(5) memory usage: 4.9 KB







[10]: # a way to examine the structure of a pandas data frame df.info(verbose=True)

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
0.00	000000		
0	company_name	31 non-null	string
1	company_id	31 non-null	string
2	isic	0 non-null	Int64
3	country	31 non-null	string
4	region	31 non-null	string
5	industry_level_1	0 non-null	Int64
6	industry_level_2	0 non-null	Int64
7	industry_level_3	0 non-null	Int64
8	industry_level_4	0 non-null	Int64
9	sector	31 non-null	string
10	company_revenue	30 non-null	Float64
11	company market cap	30 non-null	Float64
12	company_enterprise_value	30 non-null	Float64
13	company_total_assets	30 non-null	Float64
14	company_cash_equivalents	30 non-null	Float64
15	target_probability	30 non-null	Float64
	es: Float64(6), Int64(5), ry usage: 4.9 KB	string(5)	







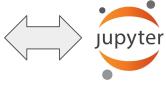


[10]: # a way to examine the structure of a pandas data frame df.info(verbose=True)

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 16 columns):

Column Non-Null Count Dtype 31 non-null string company name company id 31 non-null string isic 0 non-null Int64 country 31 non-null string region 31 non-null strina 0 non-null industry level 1 Int64 industry level 2 0 non-null Int64 industry level 3 0 non-null Int64 industry level 4 0 non-null Int64 31 non-null sector string 10 company\_revenue 30 non-null Float64 11 company market cap 30 non-null Float64 12 company enterprise value 30 non-null Float64 Float64 13 company total assets 30 non-null 14 company cash equivalents 30 non-null Float64 15 target probability 30 non-null Float64 dtypes: Float64(6), Int64(5), string(5) memory usage: 4.9 KB

Jupyter for Iteration & Development



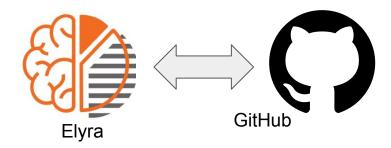






esg\_data\_ingesti... o

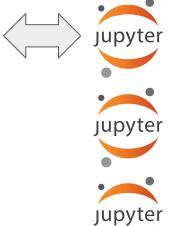
Elyra pipelines

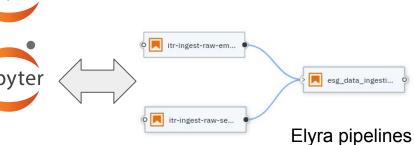


[10]: # a way to examine the structure of a pandas data frame
df.info(verbose=True)

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
0	company name	31 non-null	string
1	company id	31 non-null	string
2	isic	0 non-null	Int64
3	country	31 non-null	string
4	region	31 non-null	string
5	industry_level_1	0 non-null	Int64
6	industry_level_2	0 non-null	Int64
7	industry_level_3	0 non-null	Int64
8	industry_level_4	0 non-null	Int64
9	sector	31 non-null	string
10	company_revenue	30 non-null	Float64
11	company_market_cap	30 non-null	Float64
12	company_enterprise_value	30 non-null	Float64
13	company_total_assets	30 non-null	Float64
14	company_cash_equivalents	30 non-null	Float64
15	target_probability	30 non-null	Float64
	es: Float64(6), Int64(5), ry usage: 4.9 KB	string(5)	





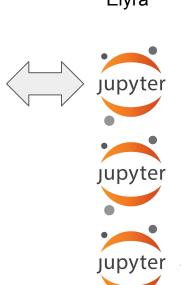


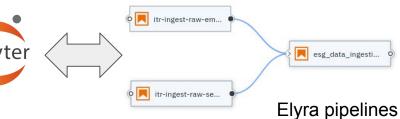
[10]: # a way to examine the structure of a pandas data frame df.info(verbose=True)

> <class 'pandas.core.frame.DataFrame'> RangeIndex: 35 entries, 0 to 34 Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
0	company_name	31 non-null	string
1	company_id	31 non-null	string
2	isic	0 non-null	Int64
3	country	31 non-null	string
4	region	31 non-null	string
5	industry level 1	0 non-null	Int64
6	industry level 2	0 non-null	Int64
7	industry level 3	0 non-null	Int64
8	industry level 4	0 non-null	Int64
9	sector	31 non-null	string
10	company_revenue	30 non-null	Float64
11	company_market_cap	30 non-null	Float64
12	company_enterprise_value	30 non-null	Float64
13	company_total_assets	30 non-null	Float64
14	company_cash_equivalents	30 non-null	Float64
15	target_probability	30 non-null	Float64
	es: Float64(6), Int64(5), ry usage: 4.9 KB	string(5)	

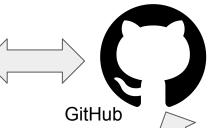
Jupyter for Iteration & Development





aicoe-ci





aicoe-ci



quay.io

[10]: # a way to examine the structure of a pandas data frame df.info(verbose=True)

> <class 'pandas.core.frame.DataFrame'> RangeIndex: 35 entries, 0 to 34 Data columns (total 16 columns):

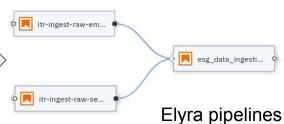
#	Column	Non-Null Count	Dtype
000			
Θ	company_name	31 non-null	string
1	company_id	31 non-null	string
2	isic	0 non-null	Int64
3	country	31 non-null	string
4	region	31 non-null	string
5	industry_level_1	0 non-null	Int64
6	industry_level_2	0 non-null	Int64
7	industry_level_3	0 non-null	Int64
8	industry_level_4	0 non-null	Int64
9	sector	31 non-null	string
10	company_revenue	30 non-null	Float64
11	company_market_cap	30 non-null	Float64
12	company_enterprise_value	30 non-null	Float64
13	company_total_assets	30 non-null	Float64
14	company_cash_equivalents	30 non-null	Float64
15	target_probability	30 non-null	Float64
	es: Float64(6), Int64(5), ry usage: 4.9 KB	string(5)	

jupyter









quay.io GitHub Elyra jupyter aicoe-ci jupyter itr-ingest-raw-em.. jupyter esg\_data\_ingesti... o | itr-ingest-raw-se...

Elyra pipelines

[10]: # a way to examine the structure of a pandas data frame df.info(verbose=True)

> <class 'pandas.core.frame.DataFrame'> RangeIndex: 35 entries, 0 to 34 Data columns (total 16 columns):

Column Non-Null Count Dtype 31 non-null string company name company id 31 non-null string isic 0 non-null Int64 country 31 non-null string region 31 non-null strina 0 non-null industry level 1 Int64 industry level 2 0 non-null Int64 industry level 3 0 non-null Int64 industry level 4 0 non-null Int64 31 non-null string sector 10 company\_revenue 30 non-null Float64 30 non-null Float64 company market cap 12 company enterprise value 30 non-null Float64 30 non-null Float64 13 company total assets 14 company cash equivalents 30 non-null Float64 15 target probability 30 non-null Float64 dtypes: Float64(6), Int64(5), string(5) memory usage: 4.9 KB

# **Data Ingest Model** Trino Analyst **DMZ Buckets** Elyra Org A **Pipelines** Org B Superset\* jupyter Org C Trino Mounted **Bucket**