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 https://github.com/os-climate/os_c_data_commons
  - 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

Jupyter and Elyra

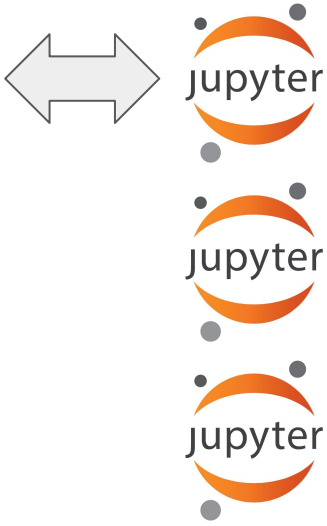
```
[10]: # a way to examine the structure of a pandas data frame
      df.info(verbose=True)

<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
dtypes: Float64(6), Int64(5), string(5)
memory usage: 4.9 KB
```

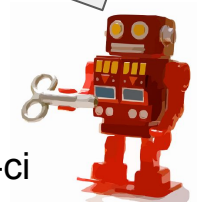
Jupyter for Iteration & Development

Elyra

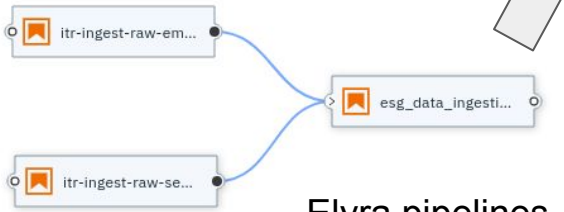
GitHub

quay.io





aicoe-ci



Elyra pipelines

Jupyter and Elyra

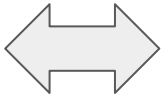
Elyra

Jupyter and Elyra

Elyra

```
[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
dtypes: Float64(6), Int64(5), string(5)
memory usage: 4.9 KB
```



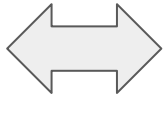
Jupyter for
Iteration &
Development

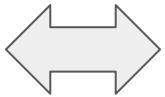
Jupyter and Elyra

Elyra

```
[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
dtypes: Float64(6), Int64(5), string(5)
memory usage: 4.9 KB
```



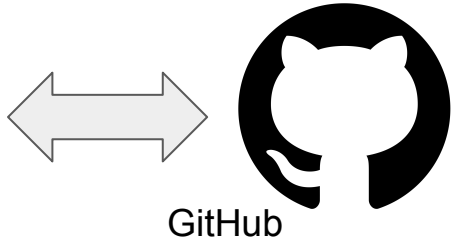


Jupyter for Iteration & Development

Elyra pipelines

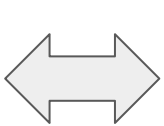
Jupyter and Elyra

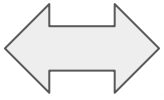
Elyra



GitHub

```
[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
dtypes: Float64(6), Int64(5), string(5)
memory usage: 4.9 KB
```



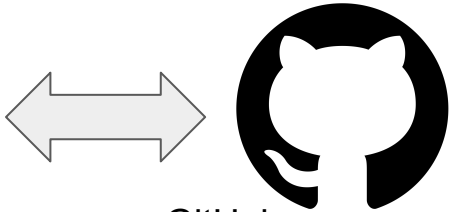


Jupyter for Iteration & Development

Elyra pipelines

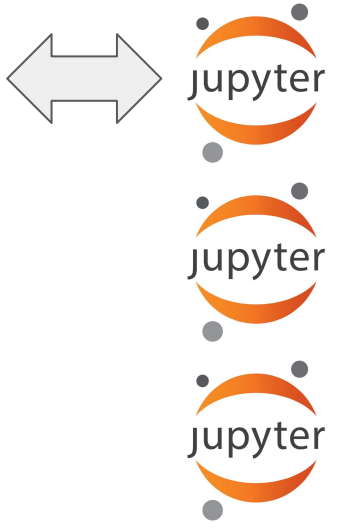
Jupyter and Elyra

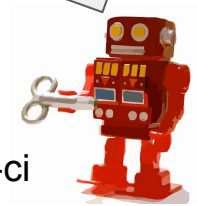
Elyra



GitHub

```
[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
dtypes: Float64(6), Int64(5), string(5)
memory usage: 4.9 KB
```





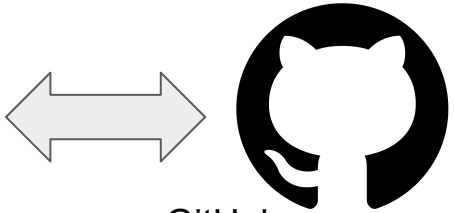
aicoe-ci

Jupyter for Iteration & Development

Elyra pipelines

Jupyter and Elyra

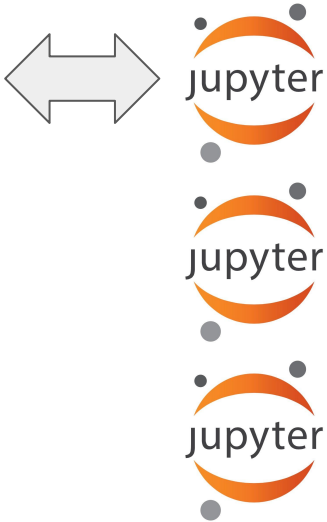
Elyra

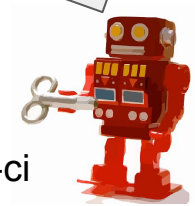


GitHub

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
---  ------                     --------------  -----
 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
dtypes: Float64(6), Int64(5), string(5)
memory usage: 4.9 KB
```





aicoe-ci

Jupyter for Iteration & Development

Elyra pipelines

Jupyter and Elyra

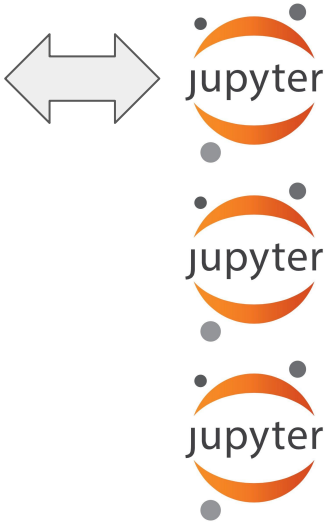
```
[10]: # a way to examine the structure of a pandas data frame
      df.info(verbose=True)

<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
dtypes: Float64(6), Int64(5), string(5)
memory usage: 4.9 KB
```

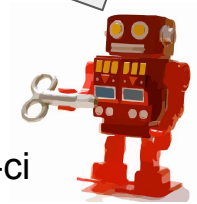
Jupyter for Iteration & Development

Elyra

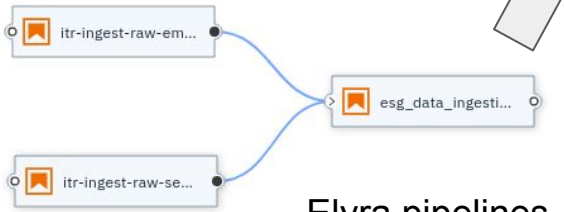
GitHub

quay.io





aicoe-ci



Elyra pipelines

Data Ingest Model