

Introduction to Elyra: AI-centric extensions to JupyterLab

—

Yiwen Li

Developer Advocate

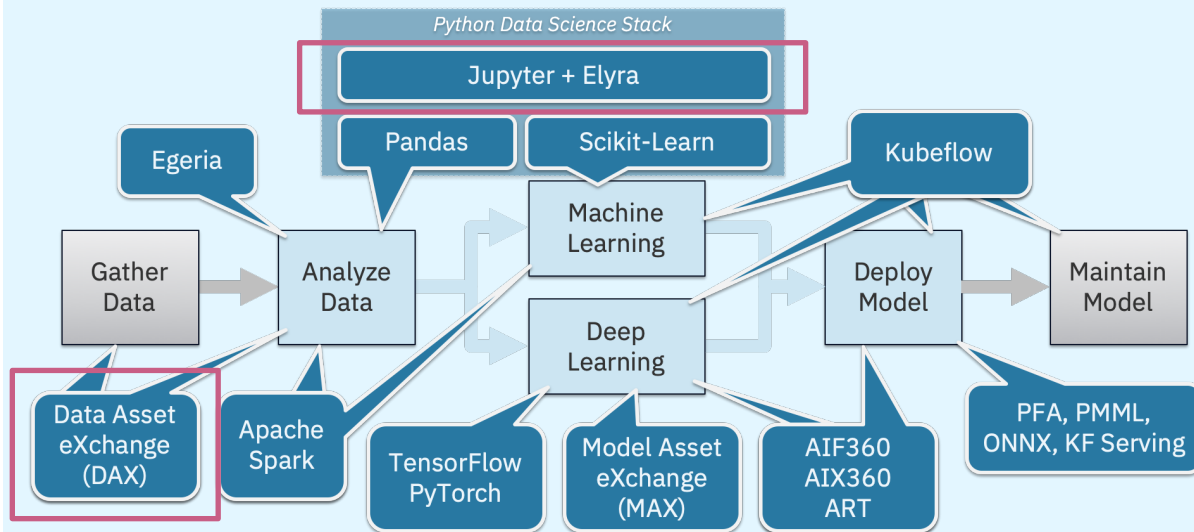
IBM CODAIT

Center for Open Source Data and AI Technologies

yiwen.Li@ibm.com | LinkedIn: <https://www.linkedin.com/in/yiwen-li-47a019119>

- CODAIT aims to make AI solutions dramatically easier to create, deploy, and manage in the enterprise.
- We contribute to and advocate for the open-source technologies that are foundational to IBM's AI offerings.
- 30+ open-source developers!

Improving the Enterprise AI Lifecycle in Open Source



What is Elyra?

And, why do we need Elyra?

Data Science Pipeline



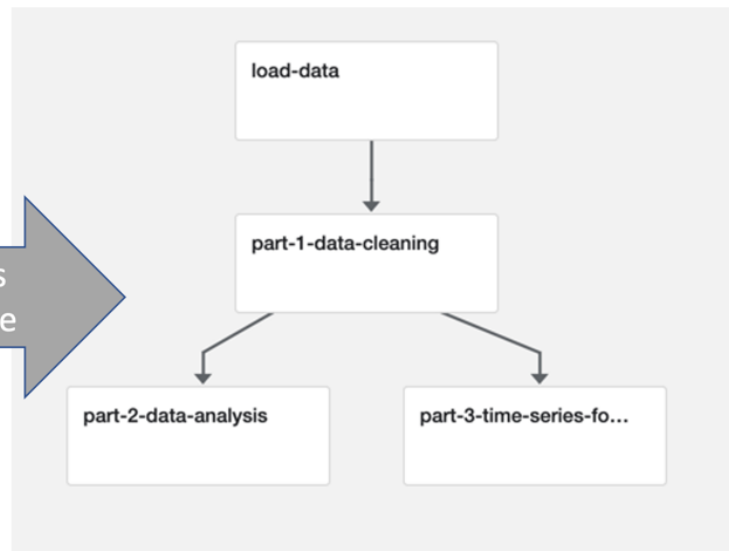
Creating notebook pipelines using Elyra and KubeFlow pipelines



First, we select only the subset of data columns of interest and inspect the column types.

```
[ ]: # Choose what columns to import from raw data
column_subset = [
    'DATE',
    'HOURLYVISIBILITY',
    'HOURLYDRYBULBTEMPF',
    'HOURLYWETBULBTEMPF',
    'HOURLYDewPointTempF',
    'HOURLYRelativeHumidity',
    'HOURLYWindSpeed',
    'HOURLYWindDirection',
    'HOURLYStationPressure',
```

Run as
pipeline



Getting Started

What are the pre-requisites to run?

1. NodeJS 12+
2. Python 3.X
3. Anaconda (optional)
4. KubeFlow installation (optional)

Install Elyra

To install Elyra:

```
$ pip install elyra==1.1.0 && jupyter lab build
```

Or:

```
$ pip install --upgrade elyra && jupyter lab build
```

To verify installation:

```
$ jupyter serverextension list
```

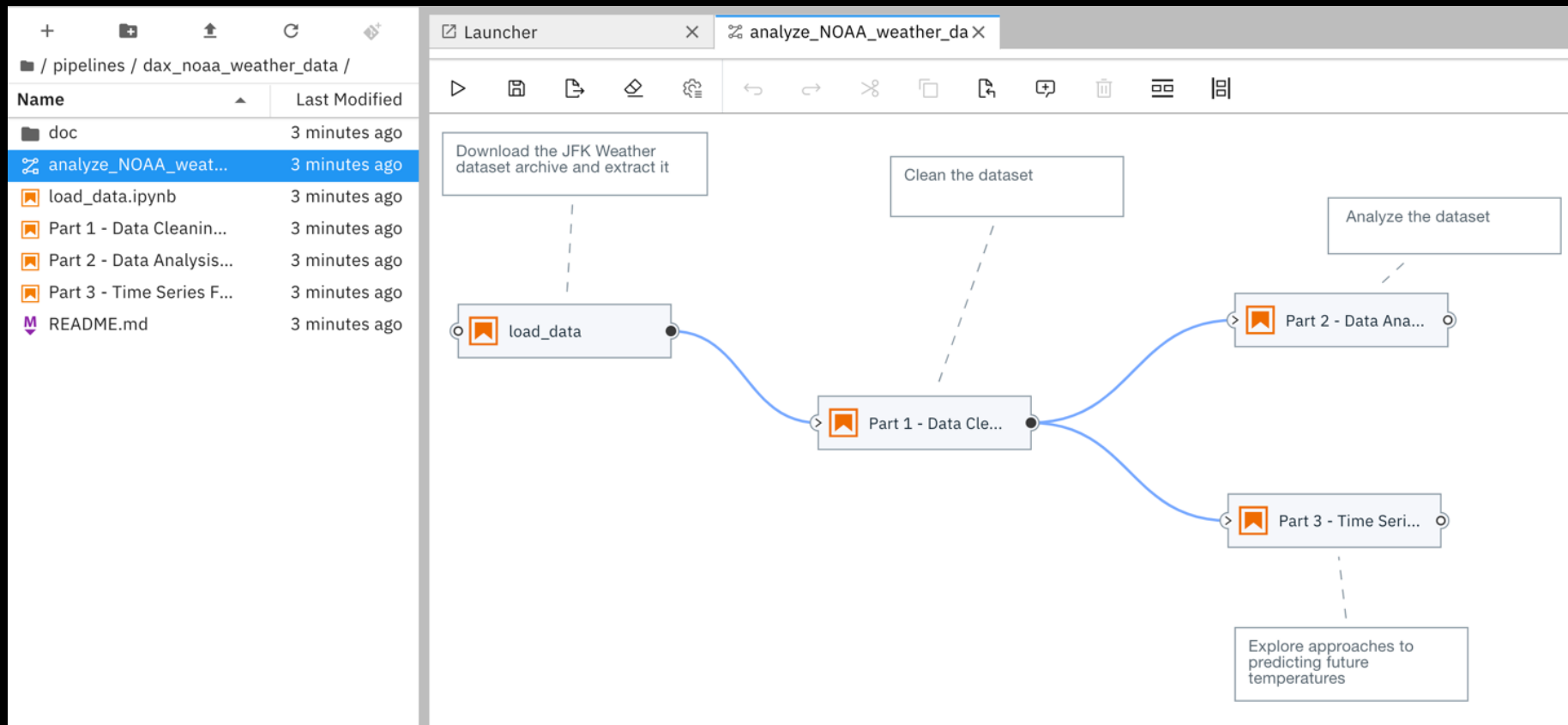
And

```
$ jupyter labextension list
```

Starting Elyra:

```
$ jupyter lab
```

Running notebooks as a pipeline



Let's run a pipeline locally and check out the notebooks!

Define Run Variables

Filename
load_data.ipynb [Browse...](#)

Runtime Image (docker image used as execution environment)
Pandas

File Dependencies [Add Dependencies...](#)
Local file dependencies that need to be copied to remote execution environment.
One filename or extension wildcard (eg. *.py) per line.

☐ Include Subdirectories in Dependencies (may increase submission time)

Environment Variables
DATASET_URL=https://dax-cdn.cdn.appdomain.cloud/dax-noaa-weather-data-jfk-airport/1.1.4/noaa-weather-data-jfk-airport.tar.gz

Output Files
data/noaa-weather-data-jfk-airport/jfk_weather.csv

[Cancel](#) [Save](#)

Several variables need to define:

- Docker runtime Image, such as Pandas image, TensorFlow image (w/ GPU support), Anaconda, or PyTorch (with CUDA-devel or with CUDA-runtime)
- File Dependencies
- Environment Variables
- Output files

Quickly add code snippet

The screenshot shows a code editor interface with a sidebar on the left containing icons for file explorer, search, and other tools. The main area displays a dialog titled 'Add new Code Snippet'. The dialog has a title bar with tabs: 'analyze_NOAA_weather_da X', 'New Code Snippet X', 'Part 1 - Data Cleaning.ipynb X', 'Part 2 - Data Analysis.ipynb X', and 'Part 3 - Time Series Foreca: X'. The dialog contains two input fields: 'Name (required)' and 'Description (optional)'. Below these is a 'Language (required)' dropdown menu currently set to '(No selection)'. A dropdown menu is open below the language selector, showing a search bar 'Filter...' and a list of languages: Python, Java, R, and Scala. At the bottom of the dialog is a 'Save & Close' button.

</> Code Snippets +

analyze_NOAA_weather_da X New Code Snippet X Part 1 - Data Cleaning.ipynb X Part 2 - Data Analysis.ipynb X Part 3 - Time Series Foreca: X

Add new Code Snippet

Name (required)

Description (optional)

Language (required)

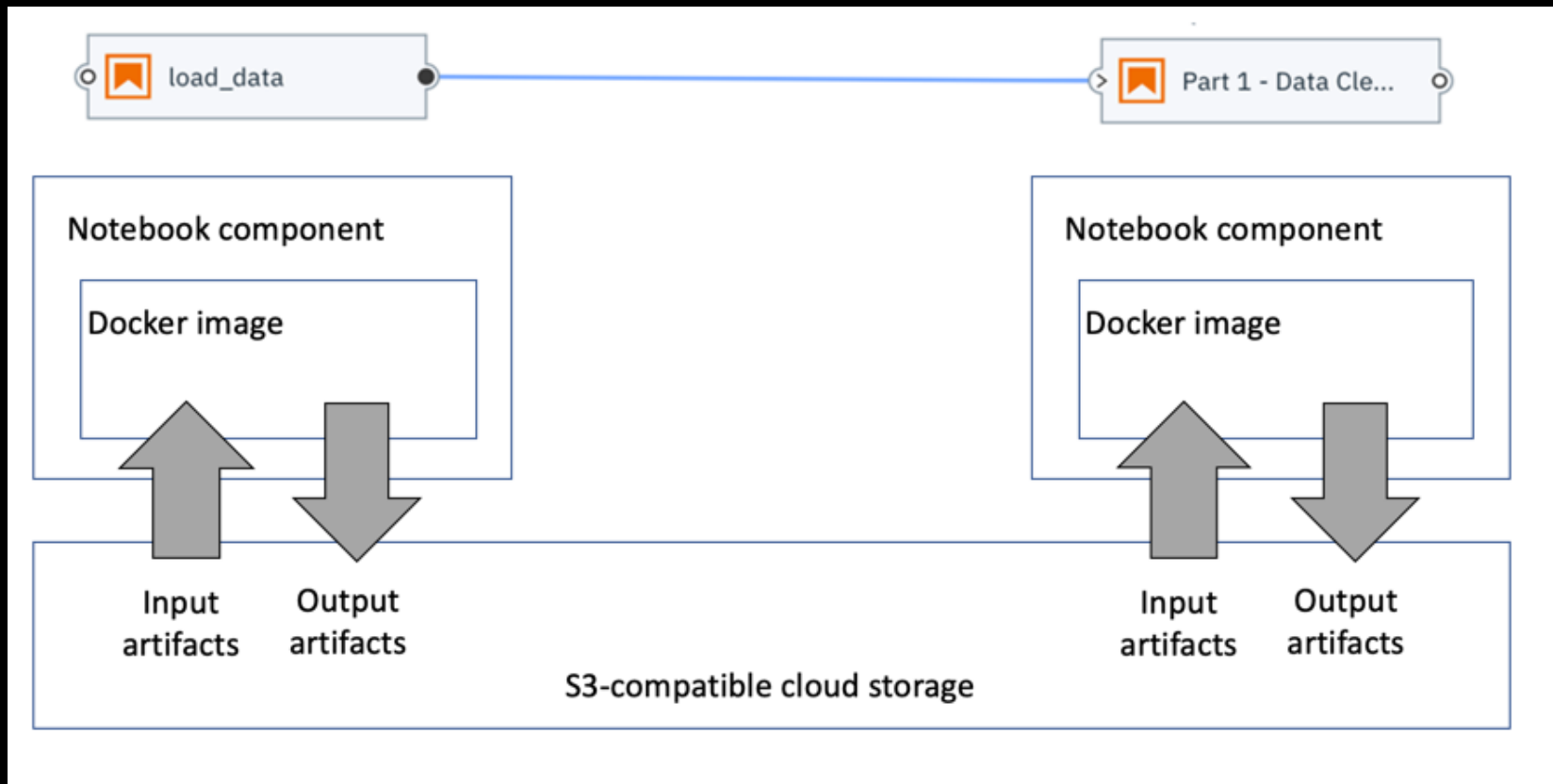
(No selection) ▼

Filter...

- Python
- Java
- R
- Scala

Save & Close

Configuring notebook nodes



If you want to run your pipeline on KubeFlow...

Configure a Kubeflow Pipeline runtime

Launcher

analyze_NOAA_weather_da

My Kubeflow Pipeline - yukl

Edit "My Kubeflow Pipeline - yukked1"

Name (required)	Description (optional)
<input type="text" value="My Kubeflow Pipeline -"/>	<input type="text"/>
Kubeflow Pipelines API Endpoint (required)	Cloud Object Storage Endpoint (required)
<input type="text" value="http://"/>	<input type="text" value="http://"/>
Cloud Object Storage Username (required)	Cloud Object Storage Password (required)
<input type="text" value=""/>	<input type="password" value="....."/>
Cloud Object Storage Bucket Name (required)	
<input type="text" value="yiwenli"/>	

Save & Close



Kubeflow

Manage runtime config using CLI

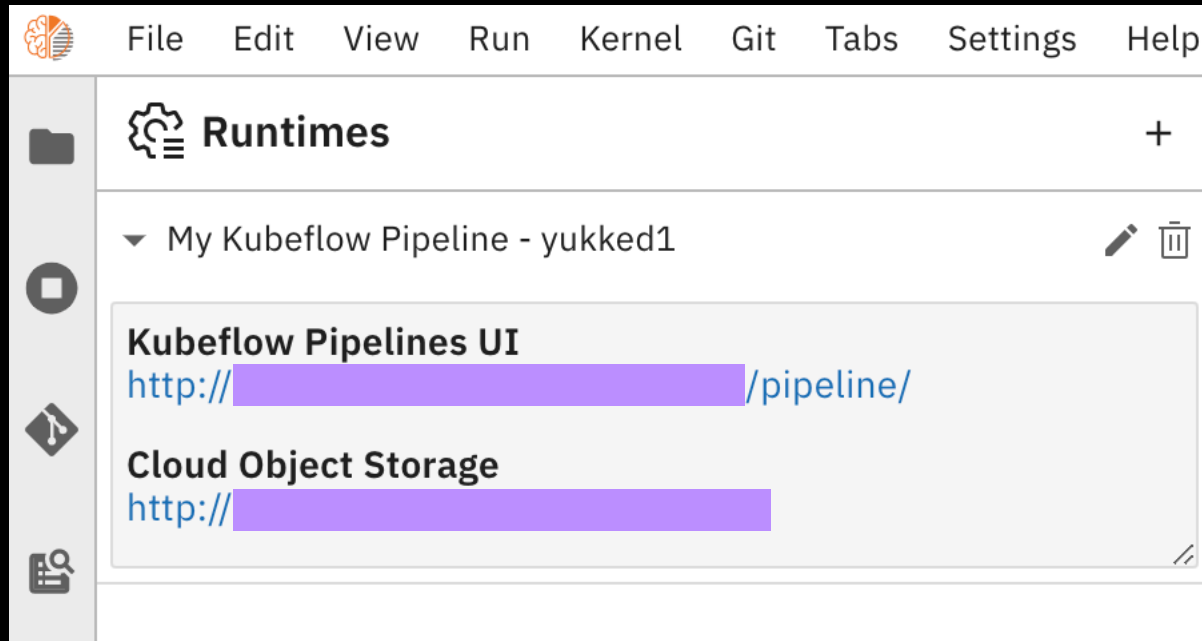
```
$ elyra-metadata install runtimes --schema_name=kfp \
--name=kfp_dev_instance \
--display_name="KFP dev instance" \
--api_endpoint=http://.../pipeline \
--cos_endpoint=http://... \
--cos_username=... \
--cos_password=... \
--cos_bucket=...

$ elyra-metadata list runtimes
Available metadata instances for runtimes (includes invalid):
Schema Instance      Resource
-----
kfp      kfp_dev_instance /Users/.../kfp_dev_instance.json
```

```
$ elyra-metadata list runtimes --json
[
  {
    "name": "kfp_dev_instance",
    "display_name": "KFP dev instance",
    "metadata": {
      "api_endpoint": "http://.../pipeline",
      "cos_endpoint": "http://...",
      "description": "...",
      "cos_username": "...",
      "cos_password": "...",
      "cos_bucket": "..."
    },
    "schema_name": "kfp",
    "resource": "/Users/.../kfp_dev_instance.json"
  }
]

$ elyra-metadata remove runtimes --name=kfp_dev_instance
Metadata instance '...' removed from namespace 'runtimes'.
```

Monitor a notebook run



Let's view the results on KubeFlow!

 Pipelines

 Experiments

 Artifacts

 Executions

 Archive

 Documentation 

 Github Repo 

 AI Hub Samples 

Experiments

[+ Create run](#)

[+ Create experiment](#)

[Compare runs](#)

[Clone run](#)

All experiments

All runs

Filter experiments



Experiment name	Description	Last 5 runs
▶ aaaa-0825120733		✓
▶ 11_test2-0825120320		✓
▶ 11_validation-0825112943		✓
▶ yuwen-0824164826		✓
▶ yuwen-0824141355		✓
▶ aaa2-0824114351		✓
▶ aaa1-0824114116		!
▶ aaa-0824113956		!
▶ aaa-0824104836		✓
▶ Default	All runs created without specifying an experiment will be grouped here.	

Pipelines

Experiments

Artifacts

Executions

Archive

Documentation

Github Repo

AI Hub Samples

Build commit: 9c16e12

Report an Issue

Experiments > yiwen-0824164826

0824164826

RetryClone runTerminateArchive

GraphRun outputConfig

load-data

part-1-data-clean...

part-2-data-analy...

part-3-time-serie...

Runtime execution graph. Only steps that are currently running or have

lambda-k4w7d-3201965441

ArtifactsInput/OutputML MetadataVolumesManifestLogsPodEvents

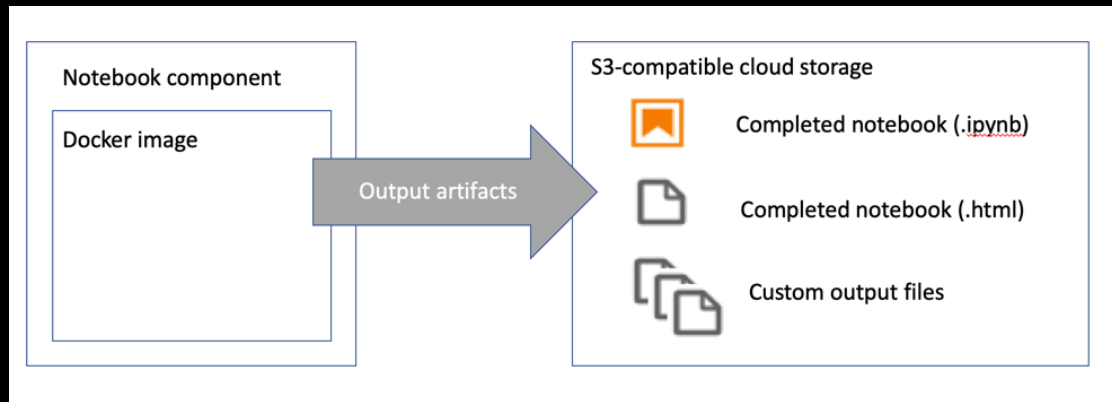
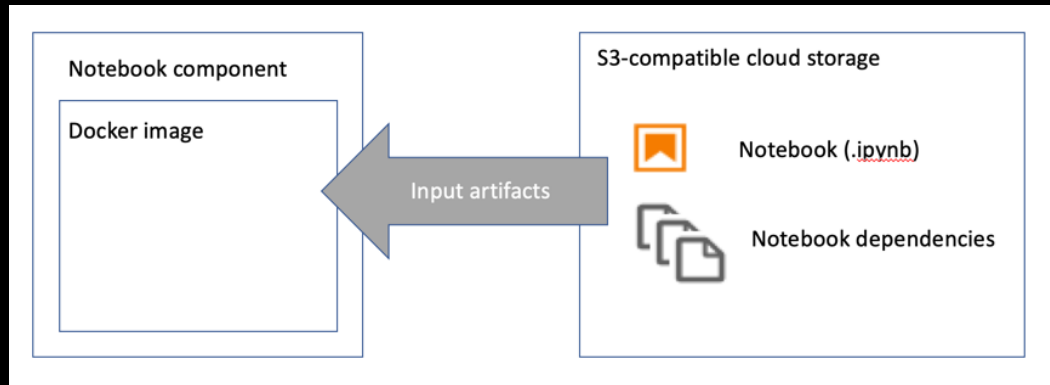
testpath==0.4.4
textwrap3==0.9.2
toml==0.10.1
tornado==6.0.4
tqdm==4.48.2
traitlets==4.3.3
typed-ast==1.4.1
urllib3==1.25.9
wcwidth==0.2.5
webencodings==0.5.1
zipp==3.1.0
/
load_data.ipynb
tar: Removing leading `/' from member names
Package not found. Installing ipykernel package with version 5.3.0...
Package not found. Installing ipython package with version 7.15.0...
Package not found. Installing ipython-genutils package with version 0.2.0...
Package not found. Installing jupyter-client package with version 6.1.6...
Package not found. Installing jupyter-core package with version 4.6.3...
Package not found. Installing minio package with version 5.0.10...
Package not found. Installing nbclient package with version 0.4.1...
Package not found. Installing nbconvert package with version 5.6.1...
Package not found. Installing nbformat package with version 5.0.7...
Package not found. Installing papermill package with version 2.1.2...
Package not found. Installing prompt-toolkit package with version 3.0.5...
Package not found. Installing pyzmq package with version 19.0.1...
Package not found. Installing requests package with version 2.23.0...
Package not found. Installing tornado package with version 6.0.4...
Package not found. Installing traitlets package with version 4.3.3...
Package not found. Installing urllib3 package with version 1.25.9...
Package Installation Complete....
Parsing Arguments.....
Get file load_data-c88d9c0b-a5d5-45ab-88d6-eb6ce24ffdbb.tar.gz from bucket yiwenli
Processing dependencies.....
TAR Archive pulled from Object Storage.
Unpacking.....
Unpacking Complete.
Executing notebook through Papermill: load_data.ipynb ==> load_data-output.ipynb
Executing: 0% | 0/8 [00:00<?, ?cell/s]Executing: 12% | 1/8 [00:00<00:06, 1.14cell/s]Executing: 75% |
Converting from ipynb to html....
Uploading Result Notebook back to Object Storage
Uploading file load_data-output.ipynb as load_data.ipynb to bucket yiwenli
Uploading file load_data.html as load_data.html to bucket yiwenli
Processing outputs.....
Uploading file data/noaa-weather-data-jfk-airport/jfk_weather.csv as data/noaa-weather-data-jfk-airport/jfk_weather.csv to bucket yiwe
Upload Complete.
257

© 2020 IBM Corporation













http://bit.ly/0901-ELYRA

19

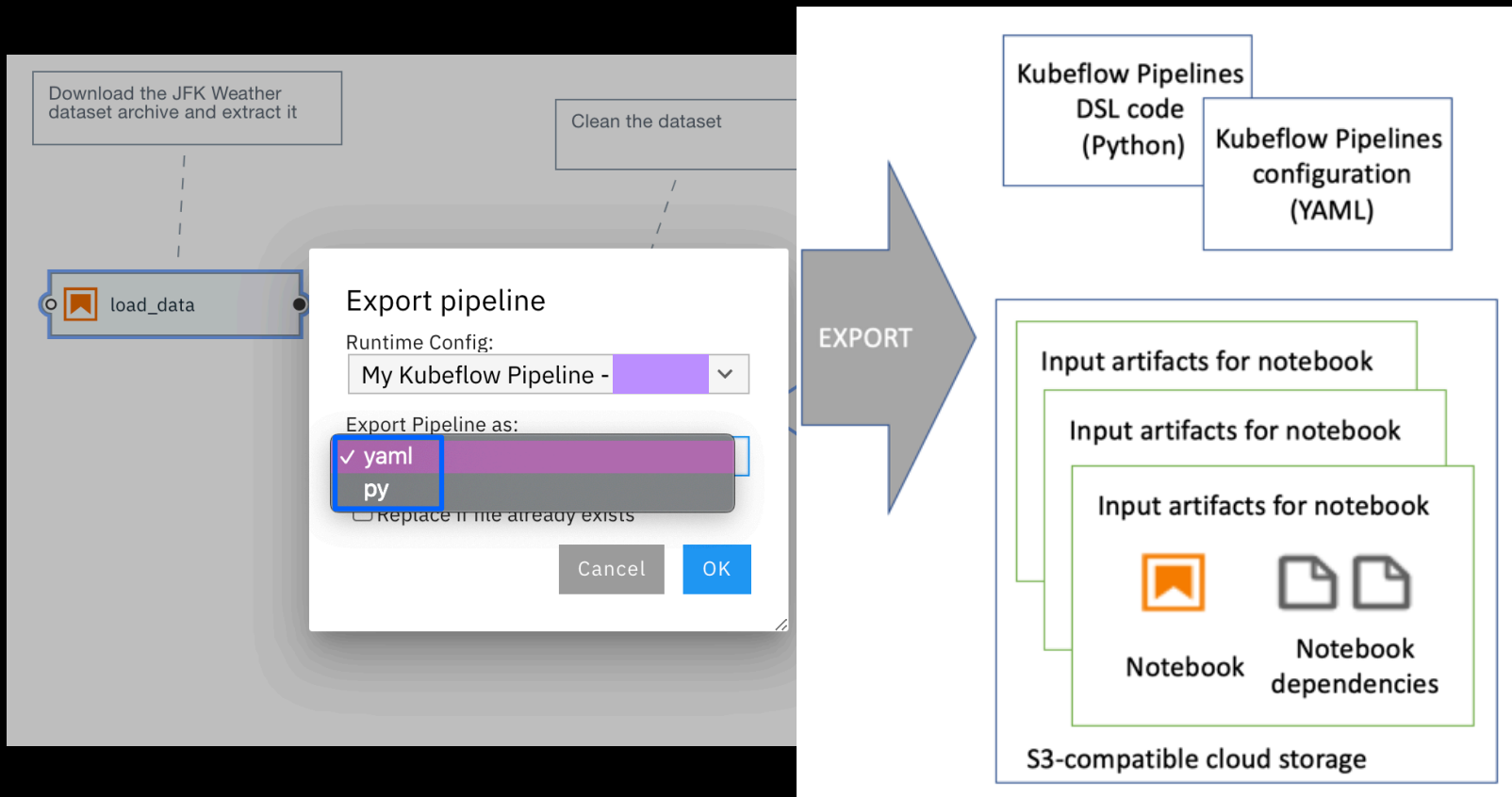
How KubeFlow pipeline works?



Pipeline outputs

Name	Size	Last Modified	
data/			
 Part 3 - Time Series Forecasting.html	564.73 KB	Aug 24, 2020 4:50 PM	...
 Part 3 - Time Series Forecasting.ipynb	956.79 KB	Aug 24, 2020 4:50 PM	...
 Part 2 - Data Analysis.html	3.62 MB	Aug 24, 2020 4:50 PM	...
 Part 2 - Data Analysis.ipynb	3.81 MB	Aug 24, 2020 4:50 PM	...
 Part 1 - Data Cleaning.html	348.84 KB	Aug 24, 2020 4:49 PM	...
 Part 1 - Data Cleaning.ipynb	110.26 KB	Aug 24, 2020 4:49 PM	...
 load_data.html	274.55 KB	Aug 24, 2020 4:49 PM	...
 load_data.ipynb	7.20 KB	Aug 24, 2020 4:49 PM	...
 Part 3 - Time Series Forecasting-b00e4654-a2b0-417c-8f93-8a03bec95945.tar.gz	9.59 KB	Aug 24, 2020 4:48 PM	...
 Part 2 - Data Analysis-982e672a-4ae5-4608-bcb0-ce309868415a.tar.gz	4.62 KB	Aug 24, 2020 4:48 PM	...
 Part 1 - Data Cleaning-e07e1b7f-568b-4bc3-9fc6-da372fd58daf.tar.gz	7.48 KB	Aug 24, 2020 4:48 PM	...
 load_data-c88d9c0b-a5d5-45ab-88d6-eb6ce24ffdbb.tar.gz	1.69 KB	Aug 24, 2020 4:48 PM	...

Export a pipeline



Ways to run pipelines

Try JupyterLab and Elyra on pre-built docker image:

```
$ docker run -it -p 8888:8888 elyra/elyra:1.0.0 jupyter lab --debug
```

Or you can use Binder:

<https://mybinder.org/v2/gh/elyra-ai/elyra/v1.0.0?urlpath=lab/tree/binder-demo>

Or you can install Elyra:

<https://github.com/elyra-ai/elyra#installation>

Get involved with Elyra!

- Open an enhancement request
- Open an issue
- Contributing to Elyra!

Curious about the notebooks in the demo?

More is available on Data Asset eXchange!

