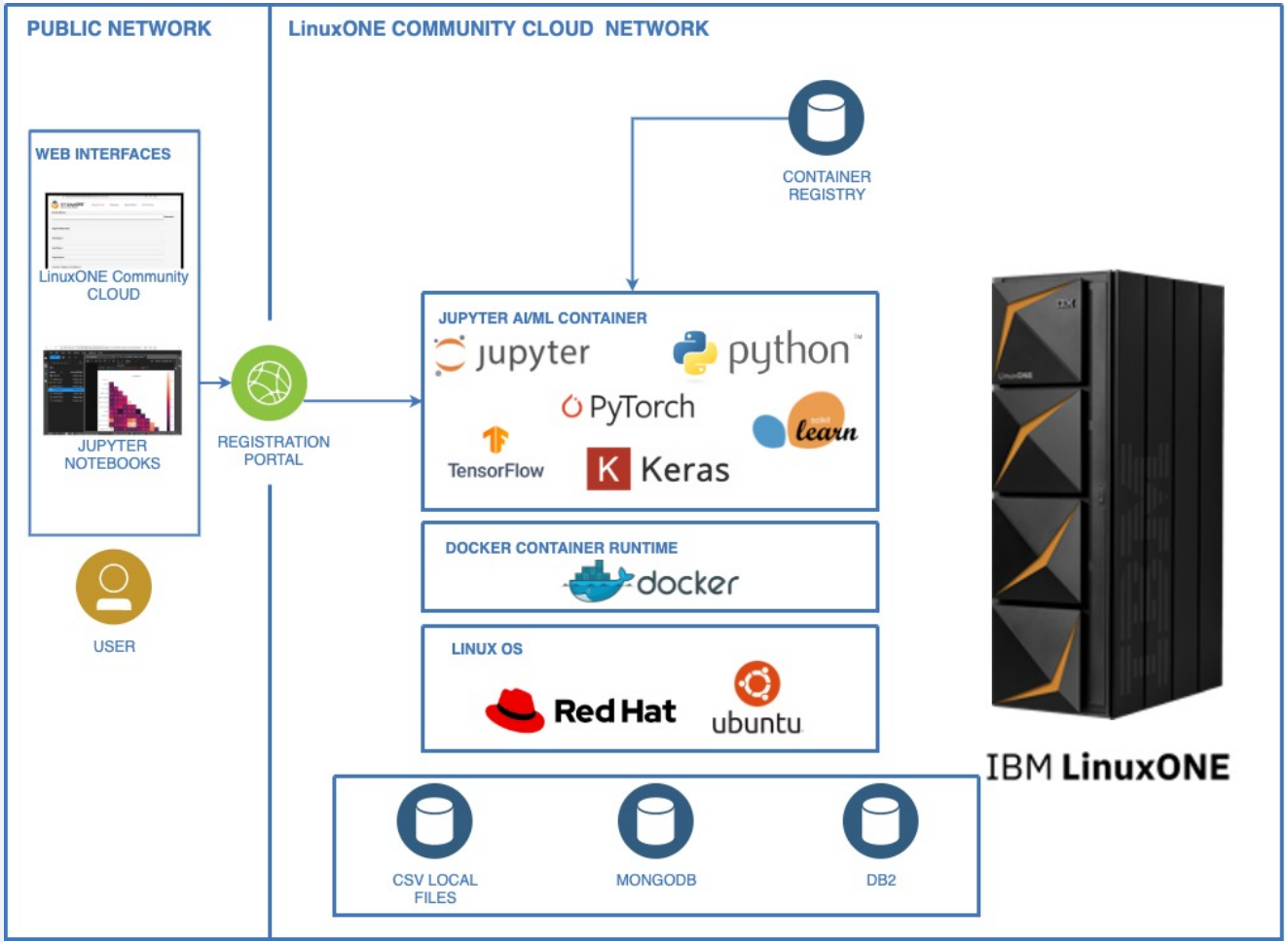
Explore AI/ML using the LinuxONE Machine Learning Lab

—

Sep 22, 2022

Dr. Alex Osadchyy
Solutions Architect, IBM Ecosystem, ISVs, zSystems

AI/ML LAB Environment - For this Event!!!



Get a Linux VM in IBM LinuxONE Community Cloud

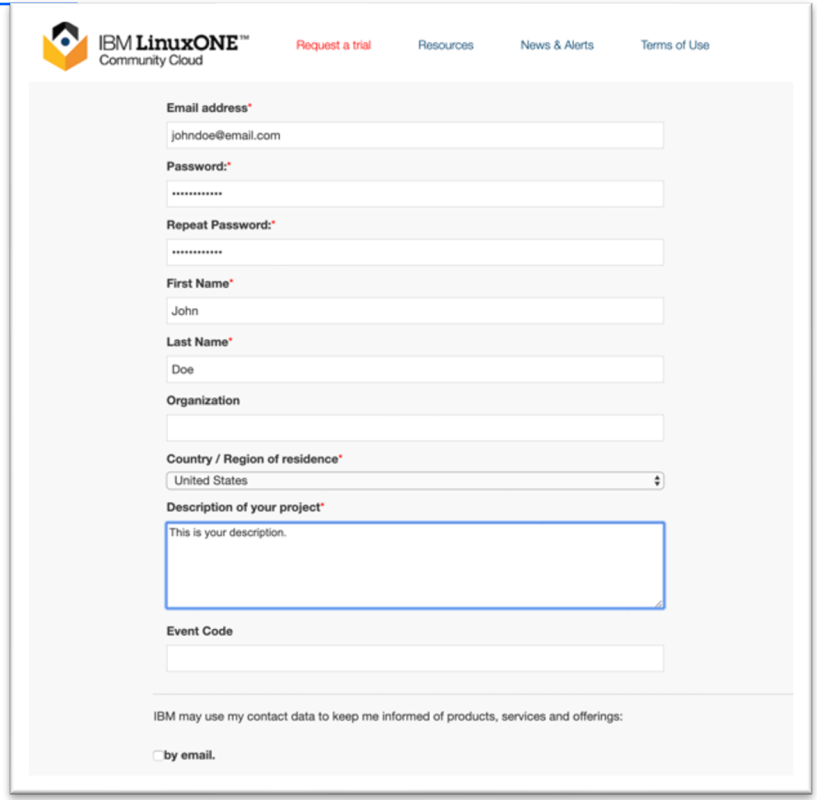
(Ref detail instructions in github)

https://github.com/linuxone-community-cloud/jupyter-lab-ml

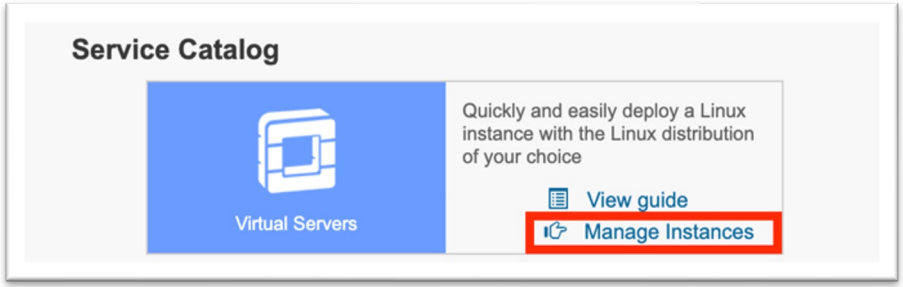
1. Register in LinuxONE Community Cloud

- Event code: SSTARSL1CC

- Instructions: https://ibm.biz/BdPcL8



2. Create your Ubuntu 20.04 instance



Bring up Jupyter Lab container

3. Open a secure shell connection and install docker runtime

```
ssh -i <your_key>.pem linux1@148.100.xx.xx
curl -fsSL https://get.docker.com -o get-docker.sh && sudo sh get-docker.sh
sudo usermod -aG docker $USER; newgrp docker
sudo systemctl start docker
exec bash   # or exit and reconnect via ssh
```

4. Start Jupyter Lab container on the port 38888

```
docker login -u l1cc registry.linuxone.cloud.marist.edu
```

Password: LinuxONE (0 is zero).

```
docker run -p 38888:8888 --name notebook -v /home/linux1/jupyter:/home/jovyan/shared \
-d registry.linuxone.cloud.marist.edu/jupyterlab-image-s390x:latest jupyter lab --ServerApp.token='L1Hackathon'
```

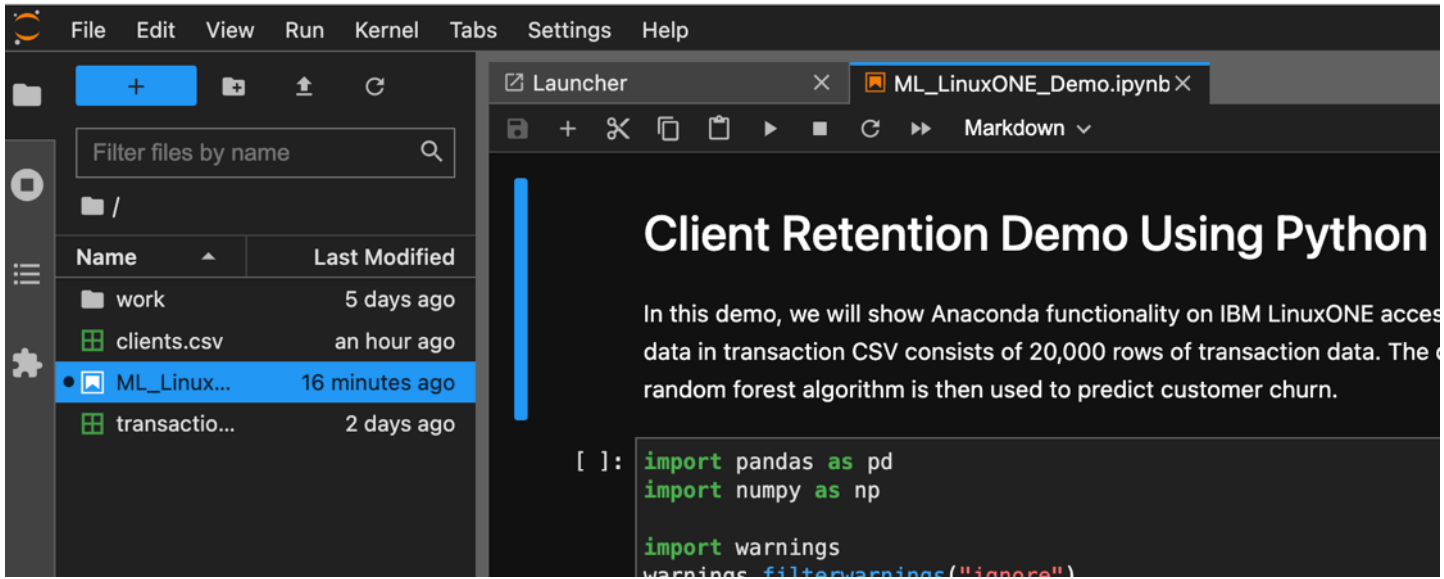
Open Jupyter Lab IDE and run a sample notebook

5. Open Jupyter Lab in the Browser using the public IP address of your instance

   URL: http://148.100.X.X:38888

   Password: L1Hackathon

6. Run Demo notebook ML_LinuxONE_Demo.ipynb



List of Pre-installed Packages

absl-py==1.0.0 aiohttp==3.8.1 aiohttp-cors==0.7.0 aiosignal==1.2.0
alembic==1.8.1 altair==4.2.0 anyio==3.6.1 apistar==0.5.41 argon2-
cffi==21.3.0 argon2-cffi-bindings==21.2.0 asttokens==2.0.7
astunparse==1.6.3 async-timeout==4.0.2 attrs==22.1.0 Babel==2.10.3
backcall==0.2.0 beautifulsoup4==4.10.0 BentoML==0.13.0 bleach==5.0.1
bokeh==2.4.2 boto3==1.24.49 botocore==1.27.49 Bottleneck==1.3.4
cachetools==5.0.0 Cerberus==1.3.4 certifi==2021.10.8 cffi==1.15.1
chardet==5.0.0 charset-normalizer==2.0.12 click==8.1.3
cloudpickle==2.0.0 configparser==5.2.0 contextlib2==21.6.0
cycler==0.11.0 Cython==0.29.28 dask==2022.2.1 dbus-python==1.2.16
debugpy==1.6.2 decorator==5.1.1 deepmerge==1.0.1 defusedxml==0.7.1
dependency-injector==4.40.0 dill==0.3.4 distro==1.7.0 docker==5.0.3
entrypoints==0.4 executing==0.9.1 fastjsonschema==2.16.1 Flask==2.2.2
flatbuffers==1.12 fonttools==4.34.4 frozenlist==1.3.1 fsspec==2022.7.1
gast==0.4.0 google-auth==2.6.6 google-auth-oauthlib==0.4.6 google-
pasta==0.2.0 grpcio==1.44.0 gunicorn==20.1.0 h5py==3.6.0
humanfriendly==10.0 idna==3.3 imageio==2.21.1 importlib-
metadata==4.11.3 importlib-resources==5.9.0 ipykernel==6.15.1
ipympl==0.8.8 ipython==8.4.0 ipython-genutils==0.2.0 ipywidgets==7.6.5
itsdangerous==2.1.2 jedi==0.18.1 Jinja2==3.1.2 jmespath==1.0.1
joblib==1.1.0 json5==0.9.9 jsonschema==4.9.1 jupyter-client==7.3.4
jupyter-core==4.11.1 jupyter-server==1.13.5 jupyterlab==3.3.0
jupyterlab-pygments==0.2.2 jupyterlab-server==2.15.0 jupyterlab-
widgets==1.1.1 keras==2.7.0rc0 Keras-Preprocessing==1.1.2
kiwisolver==1.4.4 libclang==14.0.1 llvmlite==0.36.0 locket==1.0.0
lxml==4.9.1 Mako==1.2.1 Markdown==3.3.6 MarkupSafe==2.1.1
matplotlib==3.5.1 matplotlib-inline==0.1.3 mistune==0.8.4 mpmath==1.2.1
multidict==6.0.2 nbclassic==0.4.3 nbclient==0.6.6 nbconvert==6.5.1
nbformat==5.4.0

nest-asyncio==1.5.5 networkx==2.8.5 nose==1.3.7 notebook==6.4.12 notebook-
shim==0.1.0 numba==0.53.1 numexpr==2.8.1 numpy==1.19.5 oauthlib==3.2.0
onnx==1.12.0 onnxconverter-common==1.12.1 opt-einsum==3.3.0 packaging==21.3
pandas==1.4.3 pandocfilters==1.5.0 parso==0.8.3 partd==1.2.0 patsy==0.5.2
pexpect==4.8.0 pickleshare==0.7.5 Pillow==9.2.0 pip==21.3.1
pkgutil_resolve_name==1.3.10 portpicker==1.5.0 prometheus-client==0.14.1
prompt-toolkit==3.0.30 protobuf==3.19.4 psutil==5.9.0 ptyprocess==0.7.0
pure-eval==0.2.2 pyasn1==0.4.8 pyasn1-modules==0.2.8 pycparser==2.21
Pygments==2.12.0 PyGObject==3.36.0 pyparsing==3.0.9 pyrsistent==0.18.1
python-dateutil==2.8.2 python-json-logger==2.0.4 pytz==2022.1
PyWavelets==1.3.0 PyYAML==5.3.1 pyzmq==23.2.0 requests==2.27.1 requests-
oauthlib==1.3.1 rsa==4.8 ruamel.yaml==0.17.21 ruamel.yaml.clib==0.2.6
s3transfer==0.6.0 schema==0.7.5 scikit-build==0.15.0 scikit-image==0.19.2
scikit-learn==1.1.1 scipy==1.8.0 seaborn==0.11.2 Send2Trash==1.8.0
setuptools==45.2.0 six==1.16.0 skl2onnx==1.12 snapml==1.9.1 sniffio==1.2.0
soupsieve==2.3.2.post1 SQLAlchemy==1.3.24 SQLAlchemy-Utils==0.36.5 stack-
data==0.3.0 statsmodels==0.13.2 sympy==1.9 tabulate==0.8.10
tensorboard==2.9.0 tensorboard-data-server==0.6.1 tensorboard-plugin-
wit==1.8.1 tensorflow==2.7.0 tensorflow-estimator==2.7.0rc0
termcolor==1.1.0 terminado==0.15.0 tf2onnx==1.12.0 tfserve==0.3
threadpoolctl==3.1.0 tifffile==2022.8.8 tinycss2==1.1.1 toolz==0.12.0
torch==1.8.0a0+56b43f4 tornado==6.2 traitlets==5.3.0
typing_extensions==4.2.0 urllib3==1.25.11 wcwidth==0.2.5
webencodings==0.5.1 websocket-client==1.3.3 Werkzeug==2.2.2 wheel==0.34.2
whitenoise==6.2.0 widgetsnbextension==3.5.2 wrapt==1.14.0 xgboost==1.6.1
xlrd==2.0.1 yarl==1.8.1 zipp==3.8.0

FAQs

1. Docker gives permission error. How to resolve it?

e.g. docker: Got permission denied while trying to connect to the Docker daemon socket at unix:///var/run/docker.sock

After a fresh installation of docker, it configures permissions and environment. For some settings to take an effect, you need to reload the shell (exec bash) or reestablish the ssh session.

2. Jupyter Lab gives permission error for the ”shared” folder

The folder is mounted from the host. Need to make sure that the user with uid 1000 has write access to it. E.g.Run this command on the Linux instance

sudo chown -R 1000:1000 /home/linux1/jupyter

FAQs (cont.)

3. Cannot get my jupyter lab container working. How to re-deploy it?

```
docker logs notebook
docker rm -f notebook
docker run …..
```

4. How can I get a list of files uploaded to IBM LinuxONE environment?

In Jupiter notebook, type:

```
import glob
print(glob.glob("*.csv"))
```

5. How can I install a missing package from pypi?

```
import sys
!{sys.executable} -m pip install tensorflow_datasets
```

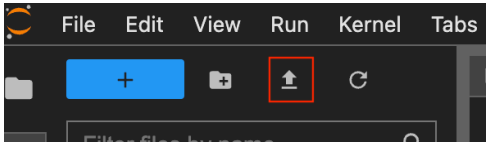
FAQs (cont.)

6. How can I upload my data and notebook to IBM LinuxONE environment

Upload files in the Jupyter Lab interface.

CSV files are supported.

Max file size is 200MB.



7. Where can I see the list of available packages?

https://github.com/linuxone-community-cloud/jupyter-lab-ml/blob/main/packages.txt

Or in Jupiter notebook, type:

```
import pkg_resources

for i in pkg_resources.working_set:

    print(i)
```

FAQs (cont.)

8. I ran out of disk space on my Linux Virtual Server. How can I free it up ?

You can clean up unused docker images to free up space:

```
docker images -a

docker rmi $(docker images –qa)
```

9. I have created the Ubuntu instance, but facing issue in opening ssh and installing docker runtime.

Make sure to follow the required steps. Re-login to ssh.

```
ssh -i <your_key>.pem linux1@148.100.xx.xx

curl -fsSL https://get.docker.com -o get-docker.sh && sudo sh get-docker.sh

sudo usermod -aG docker $USER; newgrp docker

sudo systemctl start docker

exec bash   # or exit and reconnect via ssh
```

FAQs (cont.)

10. I receive memory error, when we try to do some big computation.

Use the following commands in ssh shell to validate the memory usage:

```
docker stats

free -h
```

Suggestions on where to source data

kaggle.com - public datasets for machine learning

data.gov - government organizations provide all of its data to public

US Bureau of labor statistics.
- Unemployment rates
- Compensation
- Price index

Federal Reserve
- Household debt
- Consumer finances
- Lending rates

Public datasets in the cloud

Universities
- UCI
- UC Davis

Covid:
- CDC data (includes vaccine info) - https://covid.cdc.gov/covid-data-tracker/
- Vaccination in other countries - https://ourworldindata.org/covid-vaccinations
- ESRI data on hospital resources

Medical:
Center for medicaid and medicare - https://www.cms.gov/OpenPayments/Explore-the-Data/Dataset-Downloads
Mental Health:
https://www.cdc.gov/mentalhealth/data_publications/index.htm

Criteria for the ML on IBM LinuxONE Challenge

Creativity

Novelty

Practicality

Leveraging IBM LinuxONE for

Machine Learning

Analytics

Data

Running /validating on IBM LinuxONE

Parts

Entire model

etc

Thank you !

Important Links:

Registration in LinuxONE Community Cloud
- Event code (put in project description): SSTARSL1CC
- Instructions: https://ibm.biz/BdPcL8

Accessing VM with Jupyter Lab ML
https://github.com/linuxone-community-cloud/jupyter-lab-ml
- Docker repo credentials: u: l1cc p: LinuxONE (0 is zero).
- Jupyter Lab token: L1Hackathon