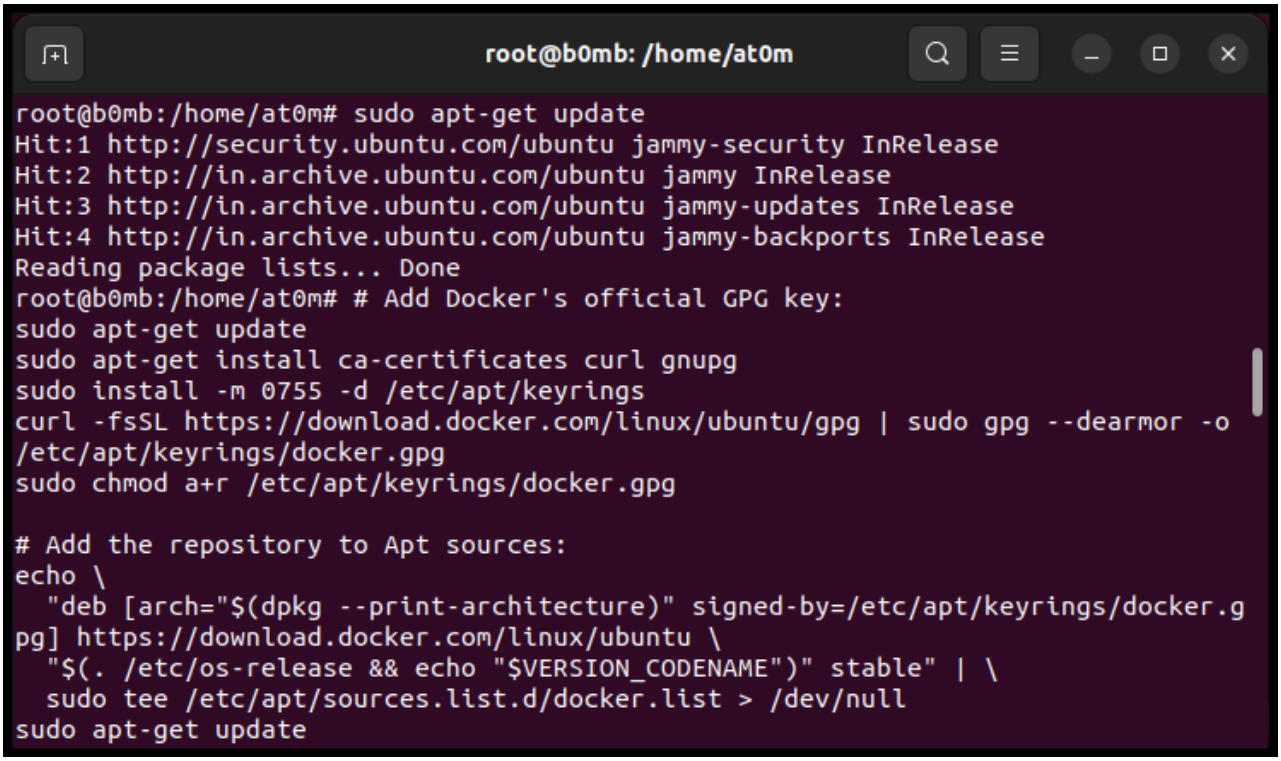
Name: Kailash Parshad

GitHub: - https://github.com/at0m-b0mb/
LinkedIn: - https://www.linkedin.com/in/kailash-parshad/

- *Project Deploying Kubeflow pipeline locally.*

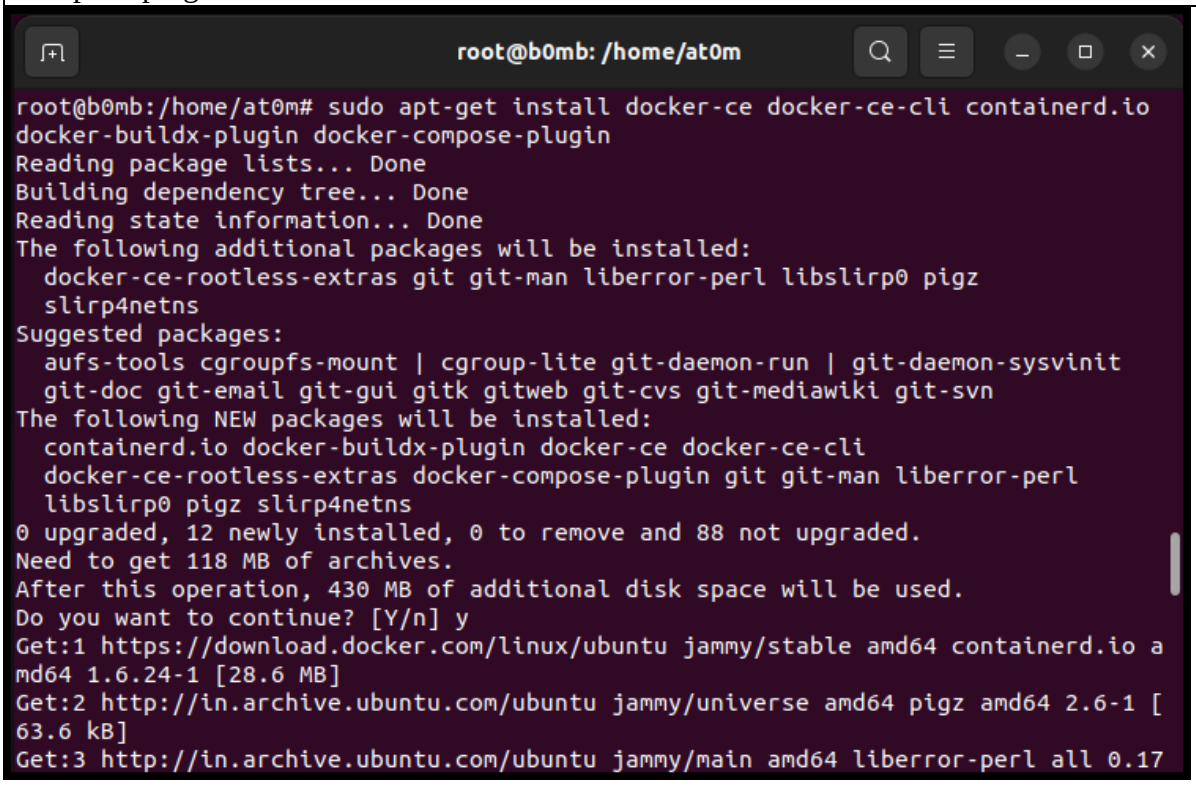
  - ***Installing Docker on Ubuntu: -***

```
# Add Docker's official GPG key:
sudo apt-get update
sudo apt-get install ca-certificates curl gnupg
sudo install -m 0755 -d /etc/apt/keyrings
curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo gpg --dearmor -o
/etc/apt/keyrings/docker.gpg
sudo chmod a+r /etc/apt/keyrings/docker.gpg

# Add the repository to Apt sources:
echo \
  "deb [arch="$(dpkg --print-architecture)" signed-by=/etc/apt/keyrings/docker.gpg]
https://download.docker.com/linux/ubuntu \
  "$(. /etc/os-release && echo "$VERSION_CODENAME")" stable" | \
  sudo tee /etc/apt/sources.list.d/docker.list > /dev/null
sudo apt-get update
```



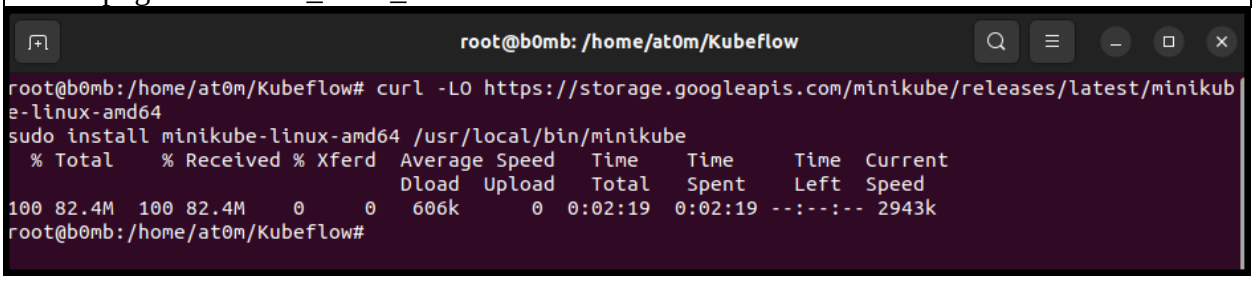
- ***Install the Docker Package: -***

sudo apt-get install docker-ce docker-ce-cli containerd.io docker-buildx-plugin docker-compose-plugin



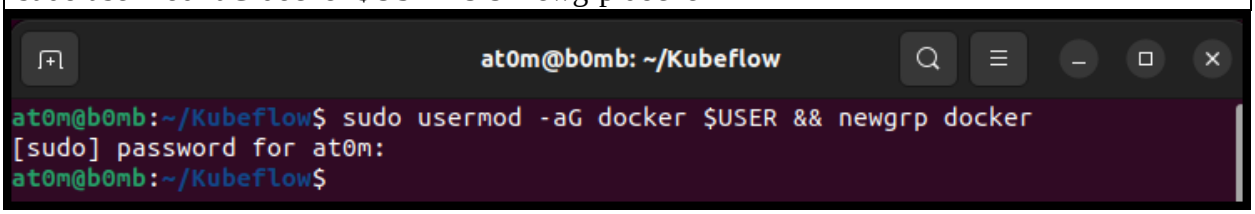
- ***<u>We will download and install minikube: -</u>***

```
curl -LO https://storage.googleapis.com/minikube/releases/latest/minikube_latest_amd64.deb
sudo dpkg -i minikube_latest_amd64.deb
```

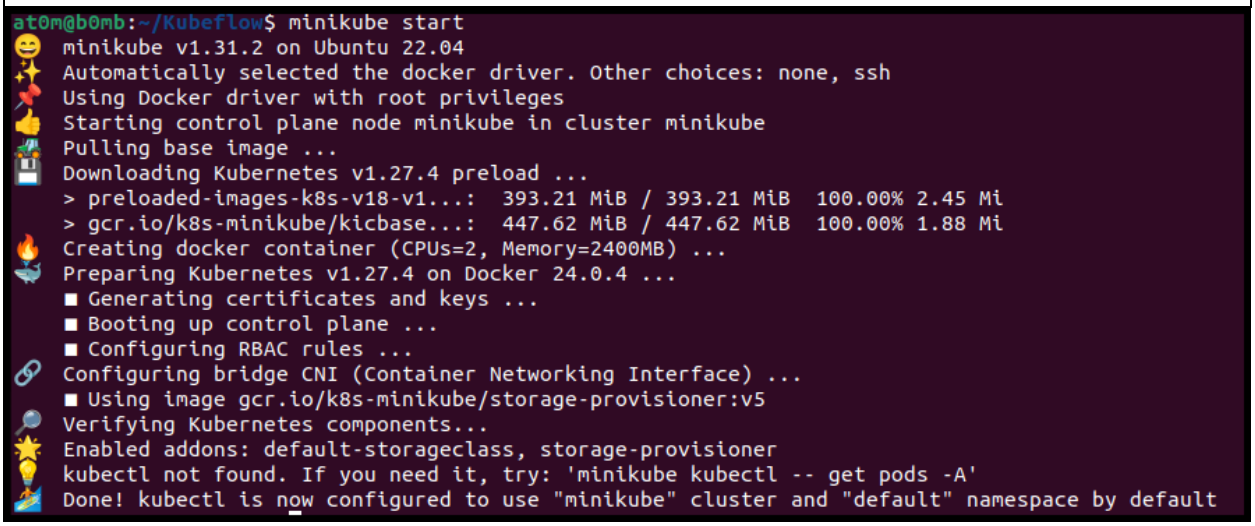


- ***<u>Starting our Cluster: -</u>***

```
sudo usermod -aG docker $USER && newgrp docker
```




```
minikube start
```



- ***<u>Interacting with our Cluster: -</u>***

```
kubectl get po -A
```

```
at0m@b0mb:~/Kubeflow$ kubectl get po -A
NAMESPACE     NAME                                 READY   STATUS    RESTARTS        AGE
kube-system   coredns-5d78c9869d-tstnc             1/1     Running   1 (7m6s ago)    76m
kube-system   etcd-minikube                        1/1     Running   1 (7m5s ago)    76m
kube-system   kube-apiserver-minikube              1/1     Running   1 (68s ago)     76m
kube-system   kube-controller-manager-minikube     1/1     Running   1 (7m5s ago)    76m
kube-system   kube-proxy-f6575                     1/1     Running   1 (7m12s ago)   76m
kube-system   kube-scheduler-minikube              1/1     Running   1 (7m5s ago)    76m
kube-system   storage-provisioner                  1/1     Running   3 (28s ago)     76m
```

- ***Deploying Kubeflow Pipelines: -***

export PIPELINE_VERSION=2.0.2
kubectl apply -k "github.com/kubeflow/pipelines/manifests/kustomize/cluster-scoped-resources?ref=$PIPELINE_VERSION"
kubectl wait --for condition=established --timeout=60s crd/applications.app.k8s.io
kubectl apply -k "github.com/kubeflow/pipelines/manifests/kustomize/env/platform-agnostic-pns?ref=$PIPELINE_VERSION"

```
at0m@b0mb:~$ export PIPELINE_VERSION=2.0.2
```

```
at0m@b0mb:~$ kubectl apply -k "github.com/kubeflow/pipelines/manifests/kustomize/cluster-scope
d-resources?ref=$PIPELINE_VERSION"
# Warning: 'bases' is deprecated. Please use 'resources' instead. Run 'kustomize edit fix' to
update your Kustomization automatically.
# Warning: 'vars' is deprecated. Please use 'replacements' instead. [EXPERIMENTAL] Run 'kustom
ize edit fix' to update your Kustomization automatically.
# Warning: 'bases' is deprecated. Please use 'resources' instead. Run 'kustomize edit fix' to
update your Kustomization automatically.
namespace/kubeflow created
customresourcedefinition.apiextensions.k8s.io/applications.app.k8s.io created
customresourcedefinition.apiextensions.k8s.io/clusterworkflowtemplates.argoproj.io created
customresourcedefinition.apiextensions.k8s.io/cronworkflows.argoproj.io created
customresourcedefinition.apiextensions.k8s.io/scheduledworkflows.kubeflow.org created
customresourcedefinition.apiextensions.k8s.io/viewers.kubeflow.org created
customresourcedefinition.apiextensions.k8s.io/workfloweventbindings.argoproj.io created
customresourcedefinition.apiextensions.k8s.io/workflows.argoproj.io created
customresourcedefinition.apiextensions.k8s.io/workflowtaskresults.argoproj.io created
customresourcedefinition.apiextensions.k8s.io/workflowtasksets.argoproj.io created
customresourcedefinition.apiextensions.k8s.io/workflowtemplates.argoproj.io created
serviceaccount/kubeflow-pipelines-cache-deployer-sa created
clusterrole.rbac.authorization.k8s.io/kubeflow-pipelines-cache-deployer-clusterrole created
clusterrolebinding.rbac.authorization.k8s.io/kubeflow-pipelines-cache-deployer-clusterrolebind
ing created
```

```
at0m@b0mb:~$ kubectl wait --for condition=established --timeout=60s crd/applications.app.k8s.i
o
customresourcedefinition.apiextensions.k8s.io/applications.app.k8s.io condition met
```

```
at0m@b0mb:~$ kubectl apply -k "github.com/kubeflow/pipelines/manifests/kustomize/env/platform-agnostic-pns?ref=$PIPELI
NE_VERSION"
secret/mysql-secret created
service/cache-server created
service/metadata-envoy-service created
service/metadata-grpc-service created
service/minio-service created
service/ml-pipeline created
service/ml-pipeline-ui created
service/ml-pipeline-visualizationserver created
service/mysql created
service/workflow-controller-metrics created
priorityclass.scheduling.k8s.io/workflow-controller created
persistentvolumeclaim/minio-pvc created
persistentvolumeclaim/mysql-pv-claim created
deployment.apps/cache-deployer-deployment created
deployment.apps/cache-server created
deployment.apps/metadata-envoy-deployment created
deployment.apps/metadata-grpc-deployment created
deployment.apps/metadata-writer created
deployment.apps/minio created
deployment.apps/ml-pipeline created
deployment.apps/ml-pipeline-persistenceagent created
deployment.apps/ml-pipeline-scheduledworkflow created
deployment.apps/ml-pipeline-ui created
deployment.apps/ml-pipeline-viewer-crd created
```

- ***<u>Deploying Kubeflow Pipelines takes time (>20 mins): -</u>***

```
at0m@b0mb:~$ kubectl get po -A
NAMESPACE     NAME                                        READY   STATUS              RESTARTS         AGE
kube-system   coredns-5d78c9869d-tstnc                    1/1     Running             1 (7m50s ago)    14h
kube-system   etcd-minikube                               1/1     Running             1 (7m50s ago)    14h
kube-system   kube-apiserver-minikube                     1/1     Running             1 (7m50s ago)    14h
kube-system   kube-controller-manager-minikube            1/1     Running             1 (7m50s ago)    14h
kube-system   kube-proxy-f6575                            1/1     Running             1 (7m50s ago)    14h
kube-system   kube-scheduler-minikube                     1/1     Running             1 (7m50s ago)    14h
kube-system   storage-provisioner                         1/1     Running             3 (7m15s ago)    14h
kubeflow      cache-deployer-deployment-59c7d8c975-tbs2h  0/1     ContainerCreating   0                57s
kubeflow      cache-server-f866cbcb5-wq4kd                0/1     ContainerCreating   0                57s
kubeflow      metadata-envoy-deployment-58f4d869b6-c2n8q  1/1     Running             0                57s
kubeflow      metadata-grpc-deployment-c568bd446-j9w59    0/1     ContainerCreating   0                57s
kubeflow      metadata-writer-8bd95866c-k9rlh             0/1     ContainerCreating   0                57s
kubeflow      minio-55464b6ddb-8zg9j                      0/1     ContainerCreating   0                57s
kubeflow      ml-pipeline-768c46f69c-whtrv                0/1     ContainerCreating   0                57s
kubeflow      ml-pipeline-persistenceagent-79699ccd96-j5h9v  0/1  ContainerCreating   0                57s
kubeflow      ml-pipeline-scheduledworkflow-86458cfcdb-6v856  0/1 ContainerCreating   0                56s
kubeflow      ml-pipeline-ui-788dd7c4d8-qkfzs             0/1     ContainerCreating   0                56s
kubeflow      ml-pipeline-viewer-crd-774bcdddc-zl6m6      0/1     ContainerCreating   0                56s
kubeflow      ml-pipeline-visualizationserver-8455bd6dbf-rk28t  0/1  ContainerCreating  0                56s
kubeflow      mysql-7d8b8ff4f4-jpgf5                      0/1     ContainerCreating   0                56s
kubeflow      workflow-controller-589ff7c479-nhd2x        0/1     ContainerCreating   0                56s
```

- ***<u>You will get CrashLoopBackOff error in metadata-grpc-deployment: -</u>***

```
at0m@b0mb:~$ kubectl get po -A
NAMESPACE     NAME                                               READY   STATUS              RESTARTS       AGE
kube-system   coredns-5d78c9869d-tstnc                           1/1     Running             1 (23m ago)    14h
kube-system   etcd-minikube                                      1/1     Running             1 (23m ago)    14h
kube-system   kube-apiserver-minikube                            1/1     Running             1 (23m ago)    14h
kube-system   kube-controller-manager-minikube                   1/1     Running             1 (23m ago)    14h
kube-system   kube-proxy-f6575                                   1/1     Running             1 (23m ago)    14h
kube-system   kube-scheduler-minikube                            1/1     Running             1 (23m ago)    14h
kube-system   storage-provisioner                                1/1     Running             3 (23m ago)    14h
kubeflow      cache-deployer-deployment-59c7d8c975-tbs2h         1/1     Running             0              16m
kubeflow      cache-server-f866cbcb5-wq4kd                       0/1     ContainerCreating   0              16m
kubeflow      metadata-envoy-deployment-58f4d869b6-c2n8q         1/1     Running             0              16m
kubeflow      metadata-grpc-deployment-c568bd446-j9w59           0/1     CrashLoopBackOff    6 (108s ago)   16m
kubeflow      metadata-writer-8bd95866c-k9rlh                    1/1     Running             0              16m
kubeflow      minio-55464b6ddb-8zg9j                             1/1     Running             0              16m
kubeflow      ml-pipeline-768c46f69c-whtrv                       0/1     Running             4 (47s ago)    16m
kubeflow      ml-pipeline-persistenceagent-79699ccd96-j5h9v      1/1     Running             0              16m
kubeflow      ml-pipeline-scheduledworkflow-86458cfcdb-6v856     1/1     Running             0              16m
kubeflow      ml-pipeline-ui-788dd7c4d8-qkfzs                    0/1     ContainerCreating   0              16m
kubeflow      ml-pipeline-viewer-crd-774bcdddc-zl6m6             0/1     ContainerCreating   0              16m
kubeflow      ml-pipeline-visualizationserver-8455bd6dbf-rk28t   0/1     ContainerCreating   0              16m
kubeflow      mysql-7d8b8ff4f4-jpgf5                             0/1     ContainerCreating   0              16m
kubeflow      workflow-controller-589ff7c479-nhd2x               0/1     ContainerCreating   0              16m
```

```
at0m@b0mb:~$ kubectl logs metadata-grpc-deployment-c568bd446-j9w59 -n kubeflow
WARNING: Logging before InitGoogleLogging() is written to STDERR
E1016 06:05:28.479928     1 mysql_metadata_source.cc:174] MySQL database was not initialized. Please ensure your MySQL
 server is running. Also, this error might be caused by starting from MySQL 8.0, mysql_native_password used by MLMD is
 not supported as a default for authentication plugin. Please follow <https://dev.mysql.com/blog-archive/upgrading-to-
mysql-8-0-default-authentication-plugin-considerations/>to fix this issue.
F1016 06:05:28.483739     1 metadata_store_server_main.cc:555] Check failed: absl::OkStatus() == status (OK vs. INTERN
AL: mysql_real_connect failed: errno: , error:  [mysql-error-info=''']) MetadataStore cannot be created with the given
connection config.
*** Check failure stack trace: ***
```

```
at0m@b0mb:~$ kubectl get po -A
NAMESPACE     NAME                                               READY   STATUS              RESTARTS       AGE
kube-system   coredns-5d78c9869d-tstnc                           1/1     Running             1 (31m ago)    14h
kube-system   etcd-minikube                                      1/1     Running             1 (31m ago)    14h
kube-system   kube-apiserver-minikube                            1/1     Running             1 (31m ago)    14h
kube-system   kube-controller-manager-minikube                   1/1     Running             1 (31m ago)    14h
kube-system   kube-proxy-f6575                                   1/1     Running             1 (31m ago)    14h
kube-system   kube-scheduler-minikube                            1/1     Running             1 (31m ago)    14h
kube-system   storage-provisioner                                1/1     Running             3 (30m ago)    14h
kubeflow      cache-deployer-deployment-59c7d8c975-tbs2h         1/1     Running             0              24m
kubeflow      cache-server-f866cbcb5-wq4kd                       0/1     ContainerCreating   0              24m
kubeflow      metadata-envoy-deployment-58f4d869b6-c2n8q         1/1     Running             0              24m
kubeflow      metadata-grpc-deployment-c568bd446-j9w59           0/1     CrashLoopBackOff    7 (4m21s ago)  24m
kubeflow      metadata-writer-8bd95866c-k9rlh                    1/1     Running             2 (2m50s ago)  24m
kubeflow      minio-55464b6ddb-8zg9j                             1/1     Running             0              24m
kubeflow      ml-pipeline-768c46f69c-whtrv                       0/1     Running             7 (3m53s ago)  24m
kubeflow      ml-pipeline-persistenceagent-79699ccd96-j5h9v      1/1     Running             3 (39s ago)    24m
kubeflow      ml-pipeline-scheduledworkflow-86458cfcdb-6v856     1/1     Running             0              24m
kubeflow      ml-pipeline-ui-788dd7c4d8-qkfzs                    1/1     Running             0              24m
kubeflow      ml-pipeline-viewer-crd-774bcdddc-zl6m6             1/1     Running             0              24m
kubeflow      ml-pipeline-visualizationserver-8455bd6dbf-rk28t   0/1     ContainerCreating   0              24m
kubeflow      mysql-7d8b8ff4f4-jpgf5                             0/1     ContainerCreating   0              24m
kubeflow      workflow-controller-589ff7c479-nhd2x               0/1     ContainerCreating   0              24m
```

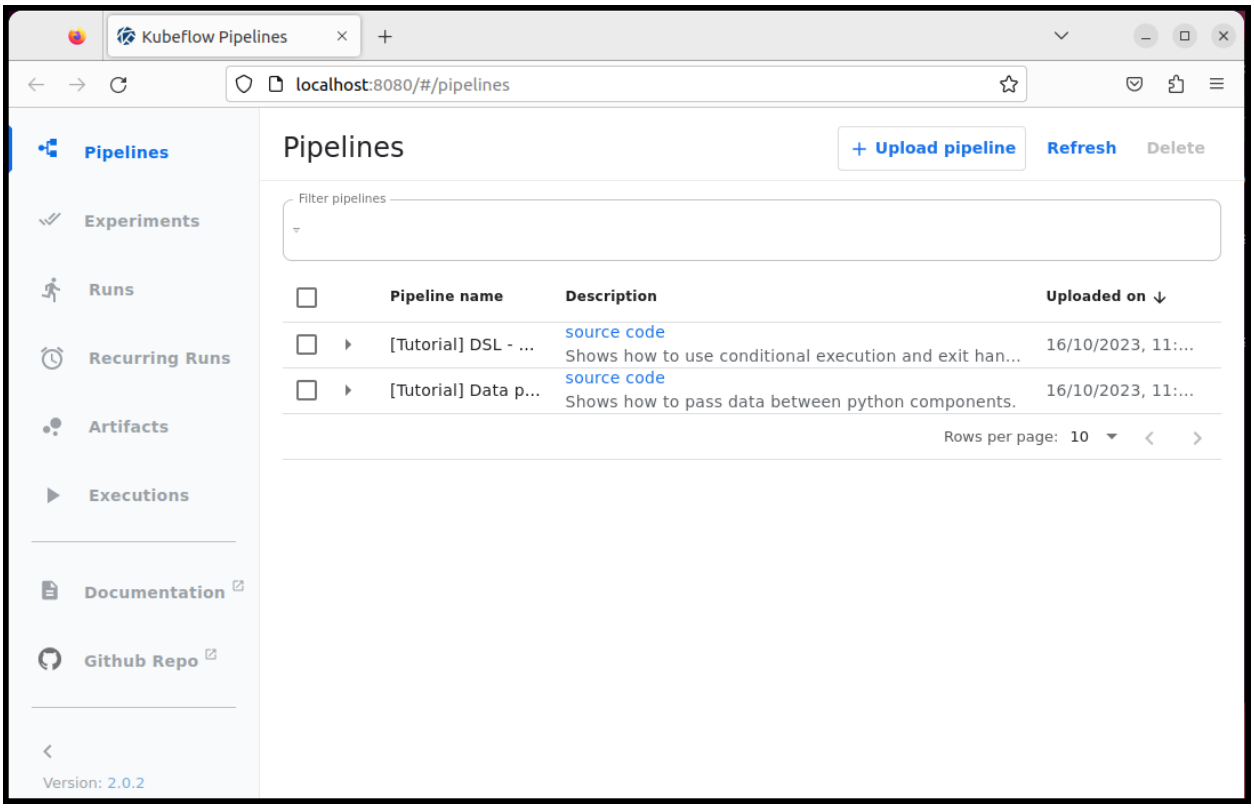
```
at0m@b0mb:~$ kubectl get po -A
NAMESPACE     NAME                                               READY   STATUS    RESTARTS        AGE
kube-system   coredns-5d78c9869d-tstnc                           1/1     Running   1 (48m ago)     15h
kube-system   etcd-minikube                                      1/1     Running   1 (48m ago)     15h
kube-system   kube-apiserver-minikube                            1/1     Running   1 (48m ago)     15h
kube-system   kube-controller-manager-minikube                   1/1     Running   1 (48m ago)     15h
kube-system   kube-proxy-f6575                                   1/1     Running   1 (48m ago)     15h
kube-system   kube-scheduler-minikube                            1/1     Running   1 (48m ago)     15h
kube-system   storage-provisioner                                1/1     Running   3 (48m ago)     15h
kubeflow      cache-deployer-deployment-59c7d8c975-tbs2h         1/1     Running   0               41m
kubeflow      cache-server-f866cbcb5-wq4kd                       1/1     Running   0               41m
kubeflow      metadata-envoy-deployment-58f4d869b6-c2n8q         1/1     Running   0               41m
kubeflow      metadata-grpc-deployment-c568bd446-j9w59           1/1     Running   10 (11m ago)    41m
kubeflow      metadata-writer-8bd95866c-k9rlh                    1/1     Running   6 (6m32s ago)   41m
kubeflow      minio-55464b6ddb-8zg9j                             1/1     Running   0               41m
kubeflow      ml-pipeline-768c46f69c-whtrv                       1/1     Running   9 (10m ago)     41m
kubeflow      ml-pipeline-persistenceagent-79699ccd96-j5h9v      1/1     Running   5 (11m ago)     41m
kubeflow      ml-pipeline-scheduledworkflow-86458cfcdb-6v856     1/1     Running   0               41m
kubeflow      ml-pipeline-ui-788dd7c4d8-qkfzs                    1/1     Running   0               41m
kubeflow      ml-pipeline-viewer-crd-774bcdddc-zl6m6             1/1     Running   0               41m
kubeflow      ml-pipeline-visualizationserver-8455bd6dbf-rk28t   1/1     Running   0               41m
kubeflow      mysql-7d8b8ff4f4-jpgf5                             1/1     Running   0               41m
kubeflow      workflow-controller-589ff7c479-nhd2x               1/1     Running   0               41m
at0m@b0mb:~$
```

- ***<u>Verifying that the Kubeflow Pipelines UI is accessible by port-forwarding: -</u>***

```
kubectl port-forward -n kubeflow svc/ml-pipeline-ui 8080:80
```

```
at0m@b0mb: ~
at0m@b0mb:~$ kubectl port-forward -n kubeflow svc/ml-pipeline-ui 8080:80
Forwarding from 127.0.0.1:8080 -> 3000
Forwarding from [::1]:8080 -> 3000
```

```
at0m@b0mb: ~
at0m@b0mb: ~                                    at0m@b0mb: ~
at0m@b0mb:~$ firefox http://localhost:8080/
Gtk-Message: 15:19:58.123: Not loading module "atk-bridge": The functionality is provided by GTK natively. Please try
to not load it.

(firefox:161014): Gtk-WARNING **: 15:19:58.304: GTK+ module /snap/firefox/3252/gnome-platform/usr/lib/gtk-2.0/modules/
libcanberra-gtk-module.so cannot be loaded.
GTK+ 2.x symbols detected. Using GTK+ 2.x and GTK+ 3 in the same process is not supported.
Gtk-Message: 15:19:58.304: Failed to load module "canberra-gtk-module"

(firefox:161014): Gtk-WARNING **: 15:19:58.309: GTK+ module /snap/firefox/3252/gnome-platform/usr/lib/gtk-2.0/modules/
libcanberra-gtk-module.so cannot be loaded.
GTK+ 2.x symbols detected. Using GTK+ 2.x and GTK+ 3 in the same process is not supported.
Gtk-Message: 15:19:58.309: Failed to load module "canberra-gtk-module"
```



- ✓ ***Kubeflow Pipeline: IRIS Classifier Model: -***
  - ***Installing Python Packages and dependencies: -***

```
at0m@b0mb:~/Kubeflow$ /bin/python3 -m pip install ipykernel -U --user --force-reinstal
l
Collecting ipykernel
  Downloading ipykernel-6.25.2-py3-none-any.whl (154 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 154.2/154.2 KB 385.2 kB/s eta 0:00:00
Collecting debugpy>=1.6.5
  Downloading debugpy-1.8.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
 (3.3 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.3/3.3 MB 388.6 kB/s eta 0:00:06
```

```
at0m@b0mb:~/Kubeflow$ pip install kfp==1.8.18
Defaulting to user installation because normal site-packages is not writeable
Collecting kfp==1.8.18
  Downloading kfp-1.8.18.tar.gz (304 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 304.8/304.8 KB 220.6 kB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Collecting Deprecated<2,>=1.2.7
  Downloading Deprecated-1.2.14-py2.py3-none-any.whl (9.6 kB)
Requirement already satisfied: PyYAML<6,>=5.3 in /usr/lib/python3/dist-packages (from kfp==
1.8.18) (5.4.1)
Collecting absl-py<2,>=0.9
  Downloading absl_py-1.4.0-py3-none-any.whl (126 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 126.5/126.5 KB 162.6 kB/s eta 0:00:00
Requirement already satisfied: click<9,>=7.1.2 in /usr/lib/python3/dist-packages (from kfp=
=1.8.18) (8.0.3)
```

```
at0m@b0mb:~/Kubeflow$ pip install numpy==1.21.0
Defaulting to user installation because normal site-packages is not writeable
Collecting numpy==1.21.0
  Downloading numpy-1.21.0.zip (10.3 MB)
     ▪ 0.1/10.3 MB 44.2 kB/s eta 0:03:49
```

```
at0m@b0mb:~/Kubeflow$ pip install pandas==1.2.4
Defaulting to user installation because normal site-packages is not writeable
Collecting pandas==1.2.4
  Downloading pandas-1.2.4.tar.gz (5.5 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 5.5/5.5 MB 1.9 MB/s eta 0:00:00
```

```
at0m@b0mb:~/Kubeflow$ pip install scikit-learn==0.24.2
Defaulting to user installation because normal site-packages is not writeable
Collecting scikit-learn==0.24.2
  Downloading scikit-learn-0.24.2.tar.gz (7.5 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.0/7.5 MB 280.2 kB/s eta 0:00:20
```

- ***Running the code in Jupyter Notebook: -***
  Link: - https://github.com/at0m-b0mb/KubeFlow-Pipeline-IRIS-Classifier-Demo/blob/main/KubeFlow_Pipeline_IRIS_Classifier_Kailash.ipynb

```
import kfp
import kfp.components as comp
import requests
import kfp.dsl as dsl
```

[1] ✓ 0.7s Python

```
!pip show kfp
```

[2] ✓ 1.1s Python

```
Name: kfp
Version: 1.8.18
Summary: KubeFlow Pipelines SDK
Home-page: https://github.com/kubeflow/pipelines
Author: The Kubeflow Authors
Author-email:
License: UNKNOWN
Location: /home/at0m/.local/lib/python3.10/site-packages
Requires: absl-py, click, cloudpickle, Deprecated, docstring-parser, fire, google-api-core, google-api-pyth
Required-by:
```

```python
def prepare_data():
    import pandas as pd
    print("---- Inside prepare_data component ----")
    # Load dataset
    df = pd.read_csv("https://raw.githubusercontent.com/at0m-b0mb/KubeFlow-Pi
    df = df.dropna()
    df.to_csv(f'data/final_df.csv', index=False)
    print("\n ---- data csv is saved to PV location /data/final_df.csv ----")
```

[21] ✓ 0.0s Python

```python
def train_test_split():
    import pandas as pd
    import numpy as np
    from sklearn.model_selection import train_test_split
    print("---- Inside train_test_split component ----")
    final_data = pd.read_csv(f'data/final_df.csv')
    target_column = 'class'
    X = final_data.loc[:, final_data.columns != target_column]
    y = final_data.loc[:, final_data.columns == target_column]

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,stratify = y, random_state=47)

    np.save(f'data/X_train.npy', X_train)
    np.save(f'data/X_test.npy', X_test)
    np.save(f'data/y_train.npy', y_train)
    np.save(f'data/y_test.npy', y_test)

    print("\n---- X_train ----")
    print("\n")
    print(X_train)

    print("\n---- X_test ----")
    print("\n")
    print(X_test)

    print("\n---- y_train ----")
    print("\n")
    print(y_train)

    print("\n---- y_test ----")
    print("\n")
    print(y_test)
```

[4] ✓ 0.0s Python

```
def training_basic_classifier():
    import pandas as pd
    import numpy as np
    from sklearn.linear_model import LogisticRegression

    print("---- Inside training_basic_classifier component ----")

    X_train = np.load(f'data/X_train.npy',allow_pickle=True)
    y_train = np.load(f'data/y_train.npy',allow_pickle=True)

    classifier = LogisticRegression(max_iter=500)
    classifier.fit(X_train,y_train)
    import pickle
    with open(f'data/model.pkl', 'wb') as f:
        pickle.dump(classifier, f)

    print("\n logistic regression classifier is trained on iris data and saved to PV location /data/model.pkl ---
```

[5] ✓ 0.0s Python

```
def predict_on_test_data():
    import pandas as pd
    import numpy as np
    import pickle
    print("---- Inside predict_on_test_data component ----")
    with open(f'data/model.pkl','rb') as f:
        logistic_reg_model = pickle.load(f)
    X_test = np.load(f'data/X_test.npy',allow_pickle=True)
    y_pred = logistic_reg_model.predict(X_test)
    np.save(f'data/y_pred.npy', y_pred)

    print("\n---- Predicted classes ----")
    print("\n")
    print(y_pred)
```

[6] ✓ 0.0s Python

```
def predict_prob_on_test_data():
    import pandas as pd
    import numpy as np
    import pickle
    print("---- Inside predict_prob_on_test_data component ----")
    with open(f'data/model.pkl','rb') as f:
        logistic_reg_model = pickle.load(f)
    X_test = np.load(f'data/X_test.npy',allow_pickle=True)
    y_pred_prob = logistic_reg_model.predict_proba(X_test)
    np.save(f'data/y_pred_prob.npy', y_pred_prob)

    print("\n---- Predicted Probabilities ----")
    print("\n")
    print(y_pred_prob)
```

[7] ✓ 0.0s Python

```
def get_metrics():
    import pandas as pd
    import numpy as np
    from sklearn.metrics import accuracy_score,precision_score,recall_score,log_loss
    from sklearn import metrics
    print("---- Inside get_metrics component ----")
    y_test = np.load(f'data/y_test.npy',allow_pickle=True)
    y_pred = np.load(f'data/y_pred.npy',allow_pickle=True)
    y_pred_prob = np.load(f'data/y_pred_prob.npy',allow_pickle=True)
    acc = accuracy_score(y_test, y_pred)
    prec = precision_score(y_test, y_pred,average='micro')
    recall = recall_score(y_test, y_pred,average='micro')
    entropy = log_loss(y_test, y_pred_prob)

    y_test = np.load(f'data/y_test.npy',allow_pickle=True)
    y_pred = np.load(f'data/y_pred.npy',allow_pickle=True)
    print(metrics.classification_report(y_test, y_pred))

    print("\n Model Metrics:", {'accuracy': round(acc, 2), 'precision': round(prec, 2), 'recall': round(recall, 2
```

[8] ✓ 0.0s Python

- ***<u>Kubeflow Pipeline Creation work starts here: -</u>***

```
create_step_prepare_data = kfp.components.create_component_from_func(
    func=prepare_data,
    base_image='python:3.10',
    packages_to_install=['pandas==2.1.1','numpy==1.26.1']
)
```

[9] ✓ 0.0s Python

```
create_step_train_test_split = kfp.components.create_component_from_func(
    func=train_test_split,
    base_image='python:3.10',
    packages_to_install=['pandas==2.1.1','numpy==1.26.1','scikit-learn==1.3.1']
)
```

[10] ✓ 0.0s Python

```
create_step_training_basic_classifier = kfp.components.create_component_from_func(
    func=training_basic_classifier,
    base_image='python:3.10',
    packages_to_install=['pandas==2.1.1','numpy==1.26.1','scikit-learn==1.3.1']
)
```

[11] ✓ 0.0s Python

```
create_step_predict_on_test_data = kfp.components.create_component_from_func(
    func=predict_on_test_data,
    base_image='python:3.10',
    packages_to_install=['pandas==2.1.1','numpy==1.26.1','scikit-learn==1.3.1']
)
```

[12] ✓ 0.0s Python

```
create_step_predict_prob_on_test_data = kfp.components.create_component_from_func(
    func=predict_prob_on_test_data,
    base_image='python:3.10',
    packages_to_install=['pandas==2.1.1','numpy==1.26.1','scikit-learn==1.3.1']
)
```

[13] ✓ 0.0s Python

```
create_step_get_metrics = kfp.components.create_component_from_func(
    func=get_metrics,
    base_image='python:3.10',
    packages_to_install=['pandas==2.1.1','numpy==1.26.1','scikit-learn==1.3.1']
)
```

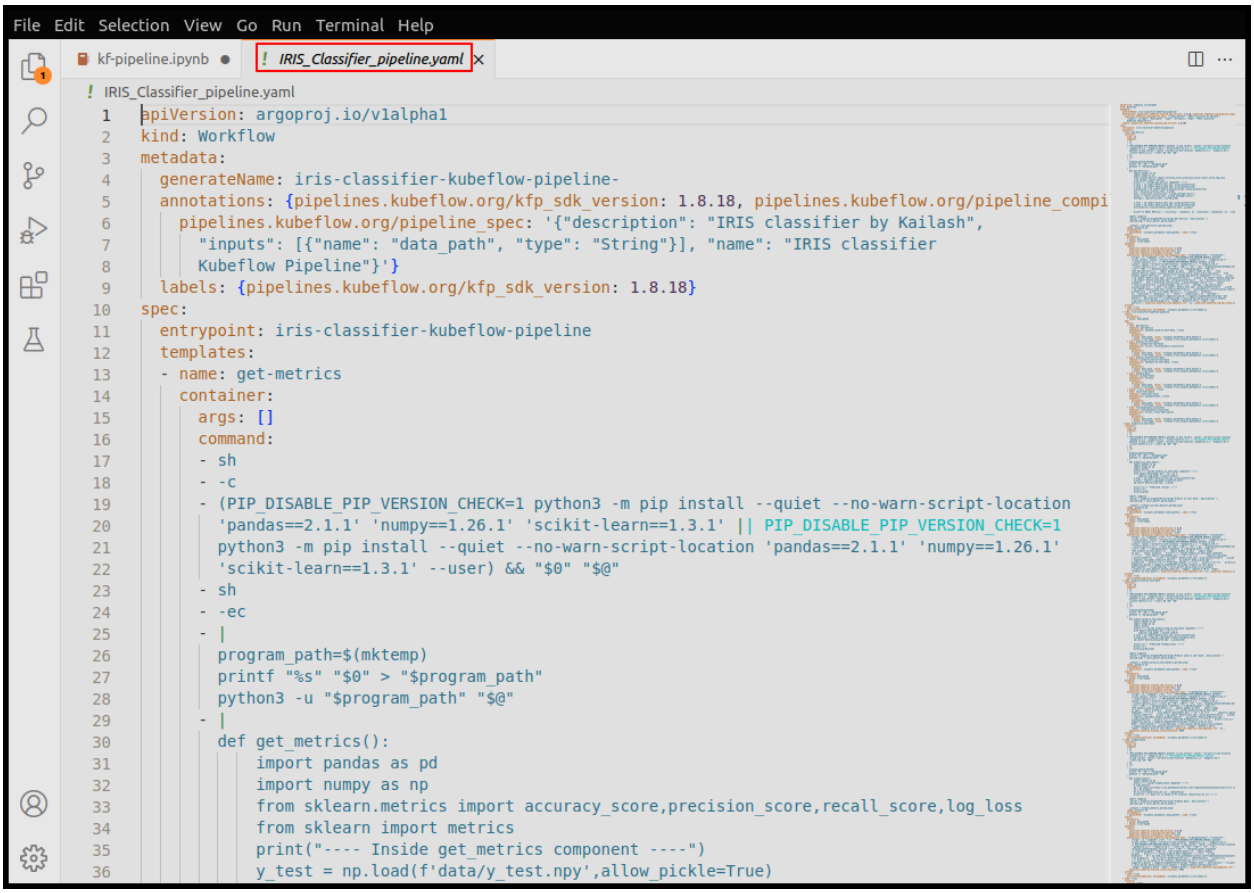
[14] ✓ 0.0s Python

```
# Define the pipeline
@dsl.pipeline(
   name='IRIS classifier Kubeflow Pipeline',
   description='IRIS classifier by Kailash'
)
# Define parameters to be fed into pipeline
def iris_classifier_pipeline(data_path: str):
    vop = dsl.VolumeOp(
    name="t-vol",
    resource_name="t-vol",
    size="1Gi",
    modes=dsl.VOLUME_MODE_RWO)

    prepare_data_task = create_step_prepare_data().add_pvolumes({data_path: vop.volume})
    train_test_split = create_step_train_test_split().add_pvolumes({data_path: vop.volume}).after(prepare_data_ta
    classifier_training = create_step_training_basic_classifier().add_pvolumes({data_path: vop.volume}).after(tra
    log_predicted_class = create_step_predict_on_test_data().add_pvolumes({data_path: vop.volume}).after(classifi
    log_predicted_probabilities = create_step_predict_prob_on_test_data().add_pvolumes({data_path: vop.volume}).a
    log_metrics_task = create_step_get_metrics().add_pvolumes({data_path: vop.volume}).after(log_predicted_probab


    prepare_data_task.execution_options.caching_strategy.max_cache_staleness = "P0D"
    train_test_split.execution_options.caching_strategy.max_cache_staleness = "P0D"
    classifier_training.execution_options.caching_strategy.max_cache_staleness = "P0D"
    log_predicted_class.execution_options.caching_strategy.max_cache_staleness = "P0D"
    log_predicted_probabilities.execution_options.caching_strategy.max_cache_staleness = "P0D"
    log_metrics_task.execution_options.caching_strategy.max_cache_staleness = "P0D"
```

[15] ✓ 0.0s Python

```
kfp.compiler.Compiler().compile(
    pipeline_func=iris_classifier_pipeline,
    package_path='IRIS_Classifier_pipeline.yaml')
```

[16] ✓ 0.0s Python



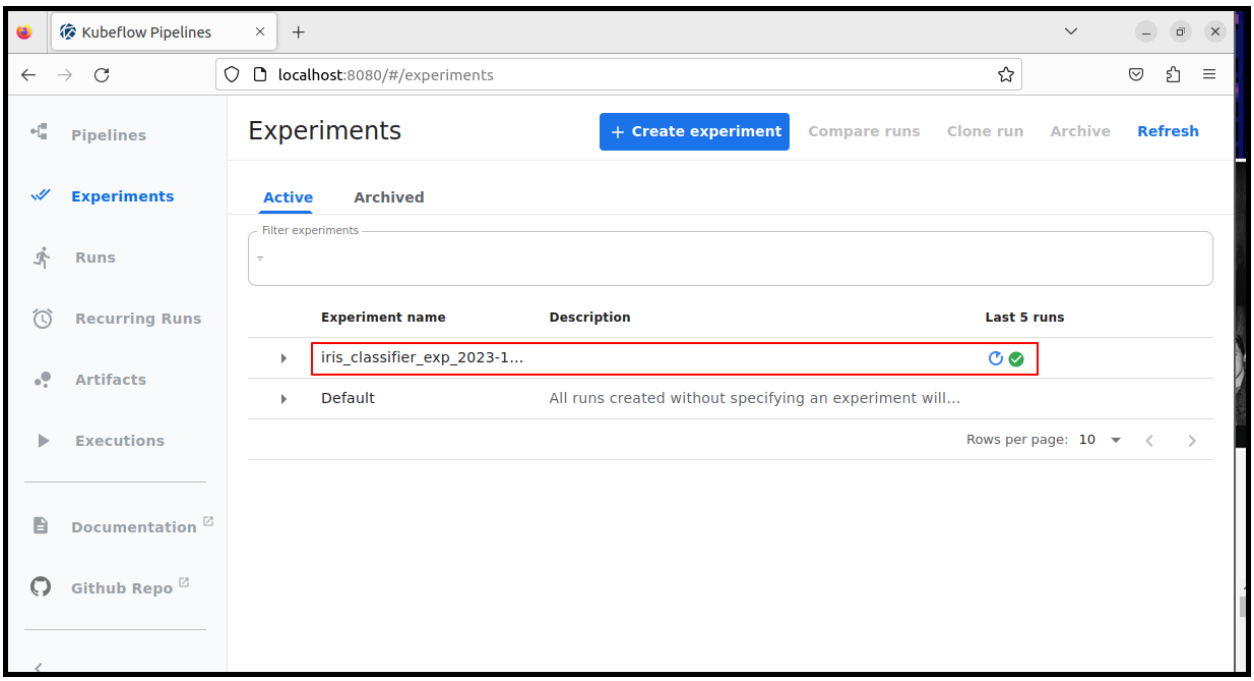
- ***<u>The Compiled YAML file is also attached: -</u>***

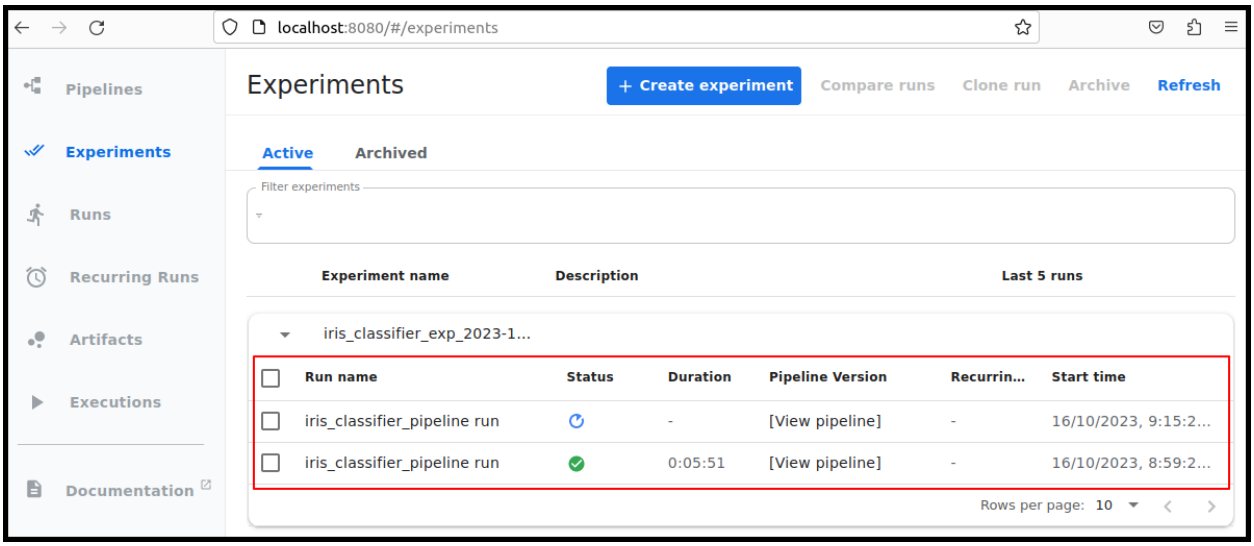
```
DATA_PATH = '/data'

import datetime
print(datetime.datetime.now().date())


pipeline_func = iris_classifier_pipeline
experiment_name = 'iris_classifier_exp' +"_"+ str(datetime.datetime.now().date())
run_name = pipeline_func.__name__ + ' run'
namespace = "kubeflow"

arguments = {"data_path":DATA_PATH}

kfp.compiler.Compiler().compile(pipeline_func,
  '{}.zip'.format(experiment_name))

run_result = client.create_run_from_pipeline_func(pipeline_func,
                                                  experiment_name=experiment_name,
                                                  run_name=run_name,
                                                  arguments=arguments)
```

- ***<u>If there are secrets: -</u>***

```
# from kubernetes import client as k8s_client
# pipeline_conf = kfp.dsl.PipelineConf()
# pipeline_conf.set_image_pull_secrets([k8s_client.V1ObjectReference(namespace='kubeflow',
#                                                                    name="secret")])
# pipeline_conf.set_image_pull_policy("IfNotPresent")

# Compile pipeline to generate compressed YAML definition of the pipeline.
# kfp.compiler.Compiler().compile(pipeline_func,
#   '{}.zip'.format(experiment_name))

# Submit pipeline directly from pipeline function
# run_result = client.create_run_from_pipeline_func(pipeline_func,
#                                                   experiment_name=experiment_name,
#                                                   run_name=run_name,
#                                                   arguments=arguments,
#                                                   namespace = namespace,
#                                                   pipeline_conf=pipeline_conf)
```

- ***<u>Now Let's Check Kubeflow UI for more information about the runs: -</u>***


Kubeflow Pipelines
localhost:8080/#/experiments
Pipelines
Experiments
Runs
Recurring Runs
Artifacts
Executions
Documentation
Github Repo
Experiments
+ Create experiment
Compare runs
Clone run
Archive
Refresh
Active
Archived
Filter experiments
Experiment name
Description
Last 5 runs
iris_classifier_exp_2023-1...
Default
All runs created without specifying an experiment will...
Rows per page: 10


localhost:8080/#/experiments
Pipelines
Experiments
Runs
Recurring Runs
Artifacts
Executions
Documentation
Experiments
+ Create experiment
Compare runs
Clone run
Archive
Refresh
Active
Archived
Filter experiments
Experiment name
Description
Last 5 runs
iris_classifier_exp_2023-1...
Run name
Status
Duration
Pipeline Version
Recurrin...
Start time
iris_classifier_pipeline run
-
[View pipeline]
-
16/10/2023, 9:15:2...
iris_classifier_pipeline run
0:05:51
[View pipeline]
-
16/10/2023, 8:59:2...
Rows per page: 10

iris_classifier_pipeline run

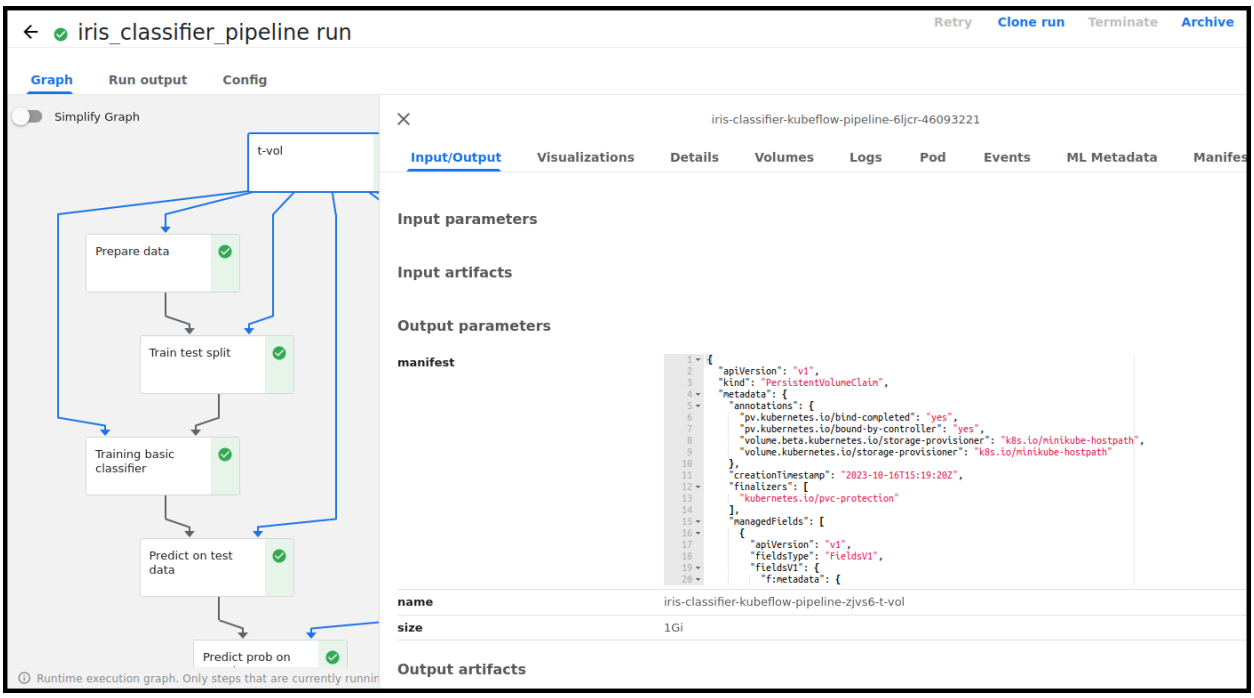
Graph | Run output | Config

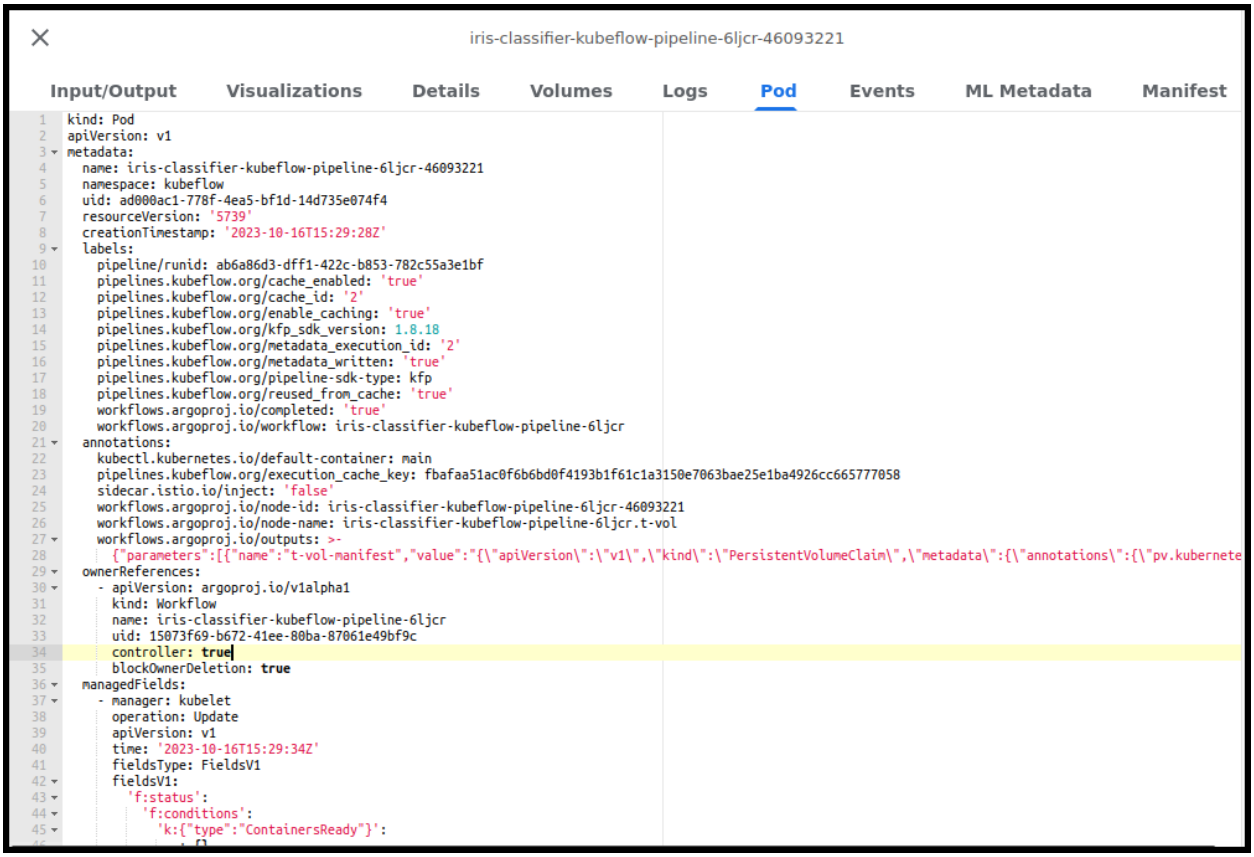
Run details

Run ID	ab6a86d3-dff1-422c-b853-782c55a3e1bf
Workflow name	iris-classifier-kubeflow-pipeline-6ljcr
Status	Succeeded
Description	
Created at	16/10/2023, 8:59:28 pm
Started at	16/10/2023, 8:59:28 pm
Finished at	16/10/2023, 9:05:19 pm
Duration	0:05:51

Run parameters

data_path	/data


iris_classifier_pipeline run
Retry
Clone run
Terminate
Archive
Graph
Run output
Config
Simplify Graph
t-vol
Prepare data
Train test split
Training basic classifier
Predict on test data
Predict prob on
Runtime execution graph. Only steps that are currently runnin
iris-classifier-kubeflow-pipeline-6ljcr-46093221
Input/Output
Visualizations
Details
Volumes
Logs
Pod
Events
ML Metadata
Manifes
Input parameters
Input artifacts
Output parameters
manifest
{
"apiVersion": "v1",
"kind": "PersistentVolumeClaim",
"metadata": {
"annotations": {
"pv.kubernetes.io/bind-completed": "yes",
"pv.kubernetes.io/bound-by-controller": "yes",
"volume.beta.kubernetes.io/storage-provisioner": "k8s.io/minikube-hostpath",
"volume.kubernetes.io/storage-provisioner": "k8s.io/minikube-hostpath"
},
"creationTimestamp": "2023-10-16T15:19:20Z",
"finalizers": [
"kubernetes.io/pvc-protection"
],
"managedFields": [
{
"apiVersion": "v1",
"fieldsType": "FieldsV1",
"fieldsV1": {
"f:metadata": {
name
iris-classifier-kubeflow-pipeline-zjvs6-t-vol
size
1Gi
Output artifacts


iris-classifier-kubeflow-pipeline-6ljcr-46093221
Input/Output
Visualizations
Details
Volumes
Logs
Pod
Events
ML Metadata
Manifest
kind: Pod
apiVersion: v1
metadata:
name: iris-classifier-kubeflow-pipeline-6ljcr-46093221
namespace: kubeflow
uid: ad000ac1-778f-4ea5-bf1d-14d735e074f4
resourceVersion: '5739'
creationTimestamp: '2023-10-16T15:29:28Z'
labels:
pipeline/runid: ab6a86d3-dff1-422c-b853-782c55a3e1bf
pipelines.kubeflow.org/cache_enabled: 'true'
pipelines.kubeflow.org/cache_id: '2'
pipelines.kubeflow.org/enable_caching: 'true'
pipelines.kubeflow.org/kfp_sdk_version: 1.8.18
pipelines.kubeflow.org/metadata_execution_id: '2'
pipelines.kubeflow.org/metadata_written: 'true'
pipelines.kubeflow.org/pipeline-sdk-type: kfp
pipelines.kubeflow.org/reused_from_cache: 'true'
workflows.argoproj.io/completed: 'true'
workflows.argoproj.io/workflow: iris-classifier-kubeflow-pipeline-6ljcr
annotations:
kubectl.kubernetes.io/default-container: main
pipelines.kubeflow.org/execution_cache_key: fbafaa51ac0f6b6bd0f4193b1f61c1a3150e7063bae25e1ba4926cc665777058
sidecar.istio.io/inject: 'false'
workflows.argoproj.io/node-id: iris-classifier-kubeflow-pipeline-6ljcr-46093221
workflows.argoproj.io/node-name: iris-classifier-kubeflow-pipeline-6ljcr.t-vol
workflows.argoproj.io/outputs: >-
{"parameters":[{"name":"t-vol-manifest","value":"{\"apiVersion\":\"v1\",\"kind\":\"PersistentVolumeClaim\",\"metadata\":{\"annotations\":{\"pv.kubernete
ownerReferences:
- apiVersion: argoproj.io/v1alpha1
kind: Workflow
name: iris-classifier-kubeflow-pipeline-6ljcr
uid: 15073f69-b672-41ee-80ba-87061e49bf9c
controller: true
blockOwnerDeletion: true
managedFields:
- manager: kubelet
operation: Update
apiVersion: v1
time: '2023-10-16T15:29:34Z'
fieldsType: FieldsV1
fieldsV1:
'f:status':
'f:conditions':
'k:{"type":"ContainersReady"}':

iris_classifier_pipeline run

Graph | Run output | Config