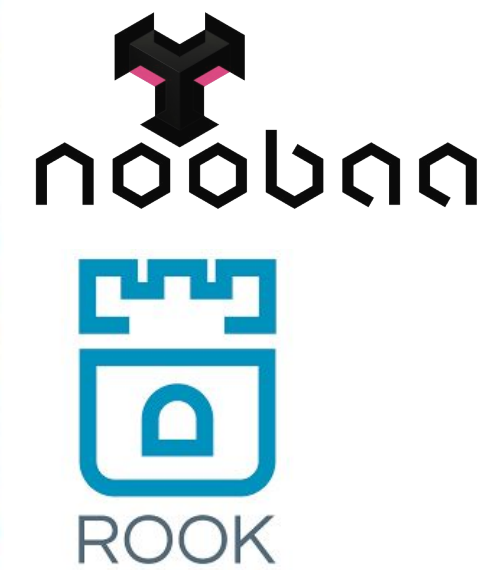
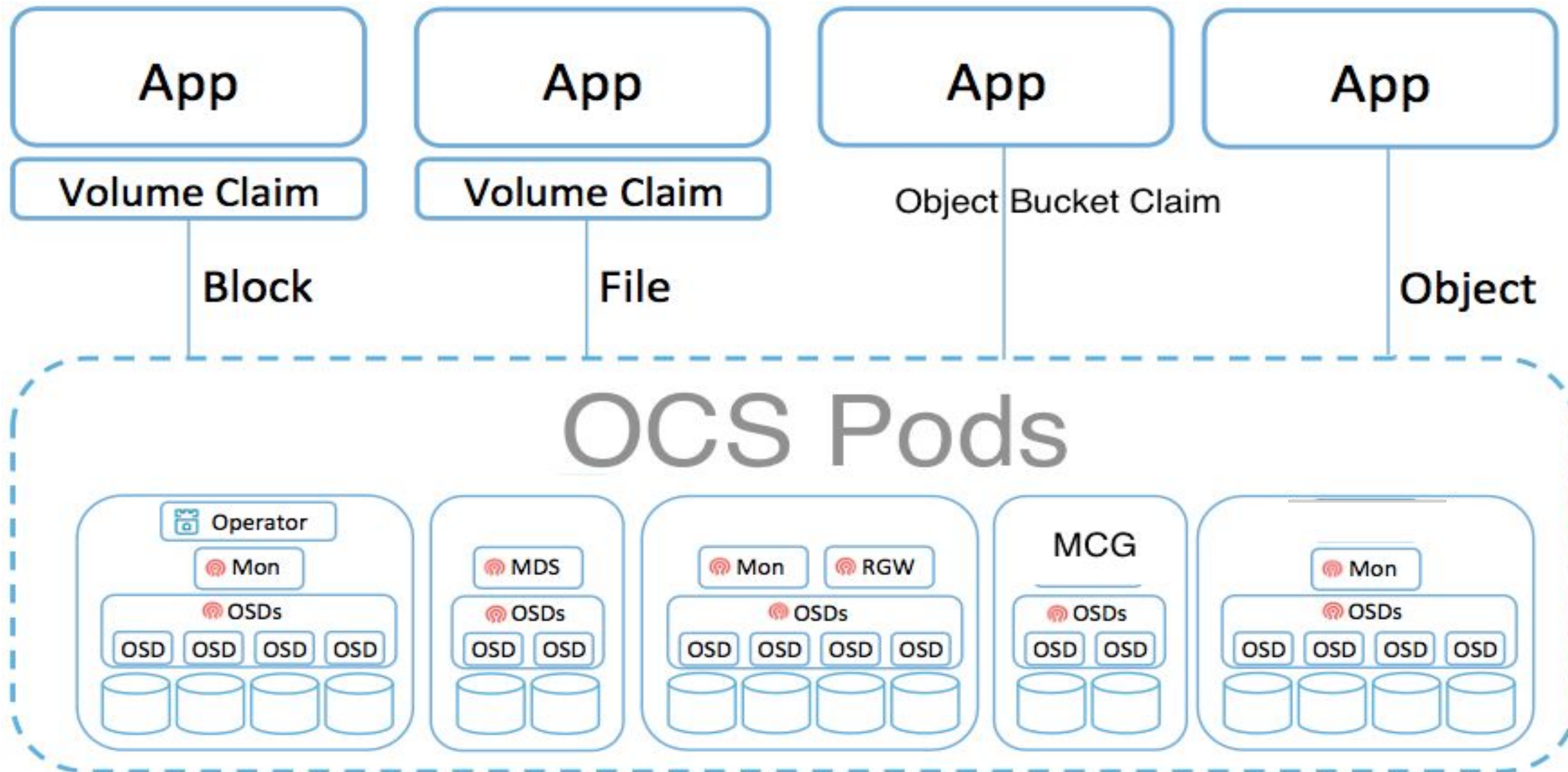




# Open Data Hub with OpenShift Container Storage



# Ceph on Openshift controlled by R00K



**Anticipating the question: source code is available!**

<https://github.com/guimou/dataprep-code/odh-ocs>

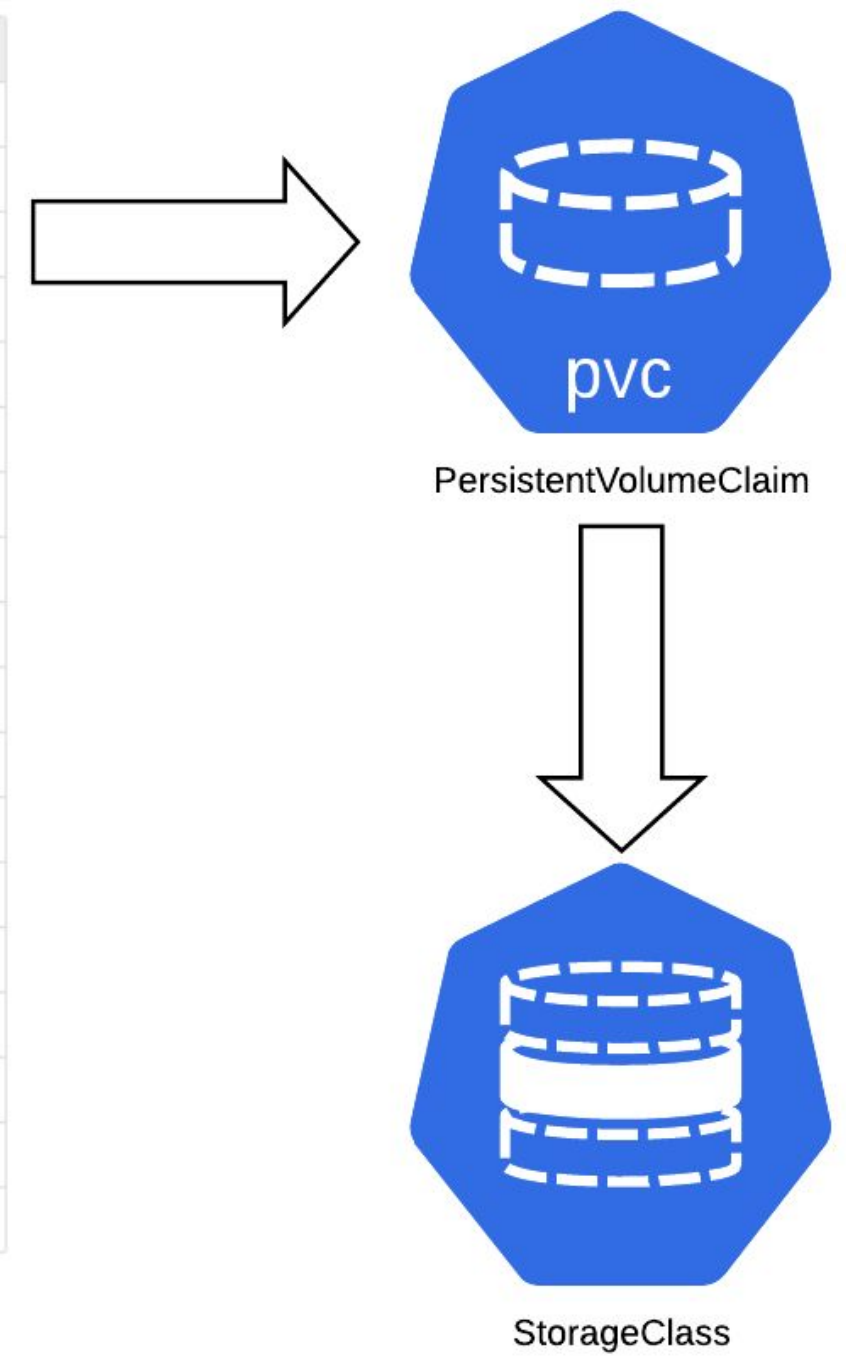
# Open Data Hub with OpenShift Container Storage

jupyter Logout Control Panel

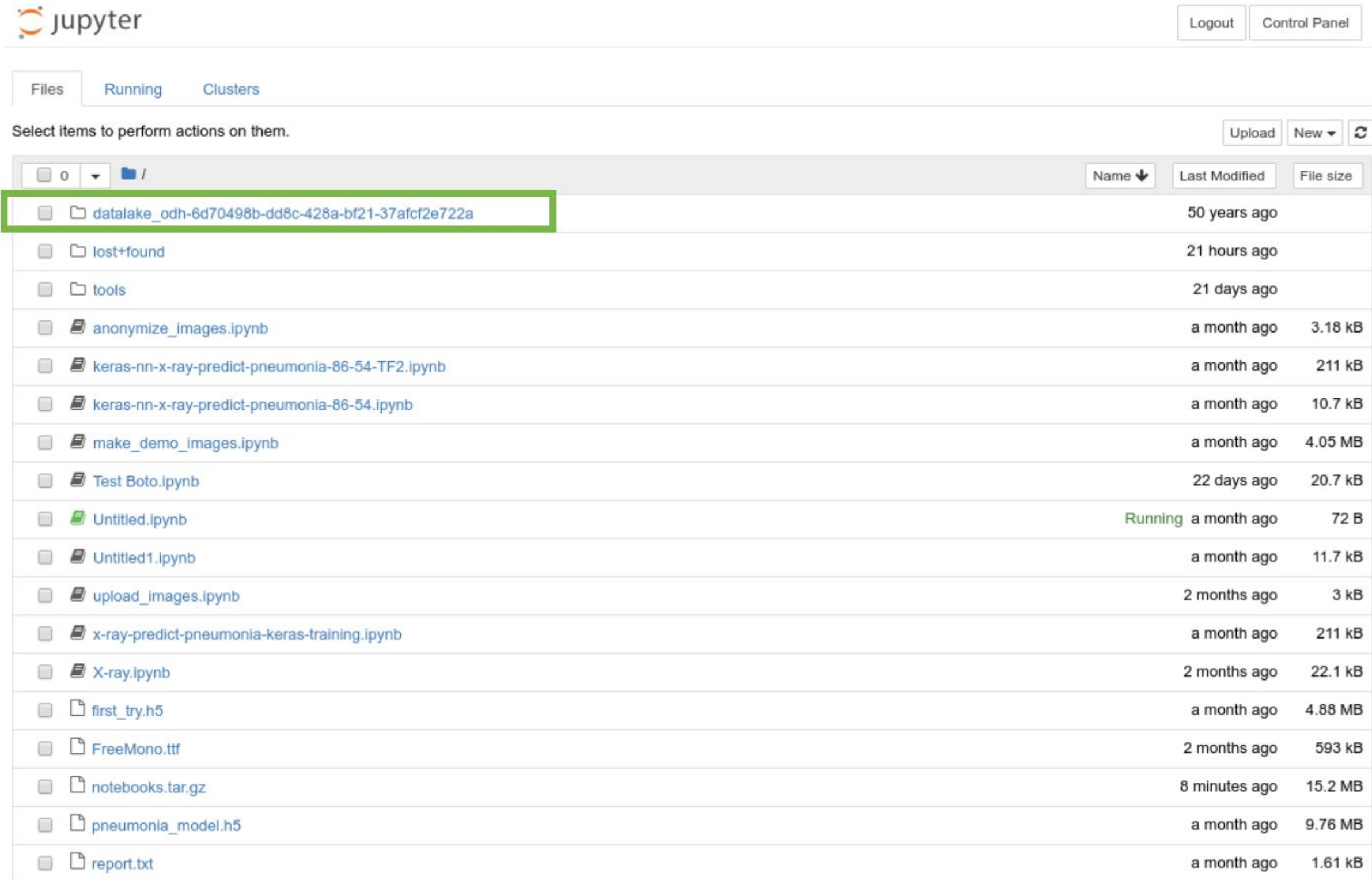
Files Running Clusters

Select items to perform actions on them. Upload New Refresh

Name	Last Modified	File size
datalake_odh-6d70498b-dd8c-428a-bf21-37afcf2e722a	50 years ago	
lost+found	21 hours ago	
tools	21 days ago	
anonymize_images.ipynb	a month ago	3.18 kB
keras-nn-x-ray-predict-pneumonia-86-54-TF2.ipynb	a month ago	211 kB
keras-nn-x-ray-predict-pneumonia-86-54.ipynb	a month ago	10.7 kB
make_demo_images.ipynb	a month ago	4.05 MB
Test Boto.ipynb	22 days ago	20.7 kB
Untitled.ipynb	Running a month ago	72 B
Untitled1.ipynb	a month ago	11.7 kB
upload_images.ipynb	2 months ago	3 kB
x-ray-predict-pneumonia-keras-training.ipynb	a month ago	211 kB
X-ray.ipynb	2 months ago	22.1 kB
first_try.h5	a month ago	4.88 MB
FreeMono.ttf	2 months ago	593 kB
notebooks.tar.gz	8 minutes ago	15.2 MB
pneumonia_model.h5	a month ago	9.76 MB
report.txt	a month ago	1.61 kB

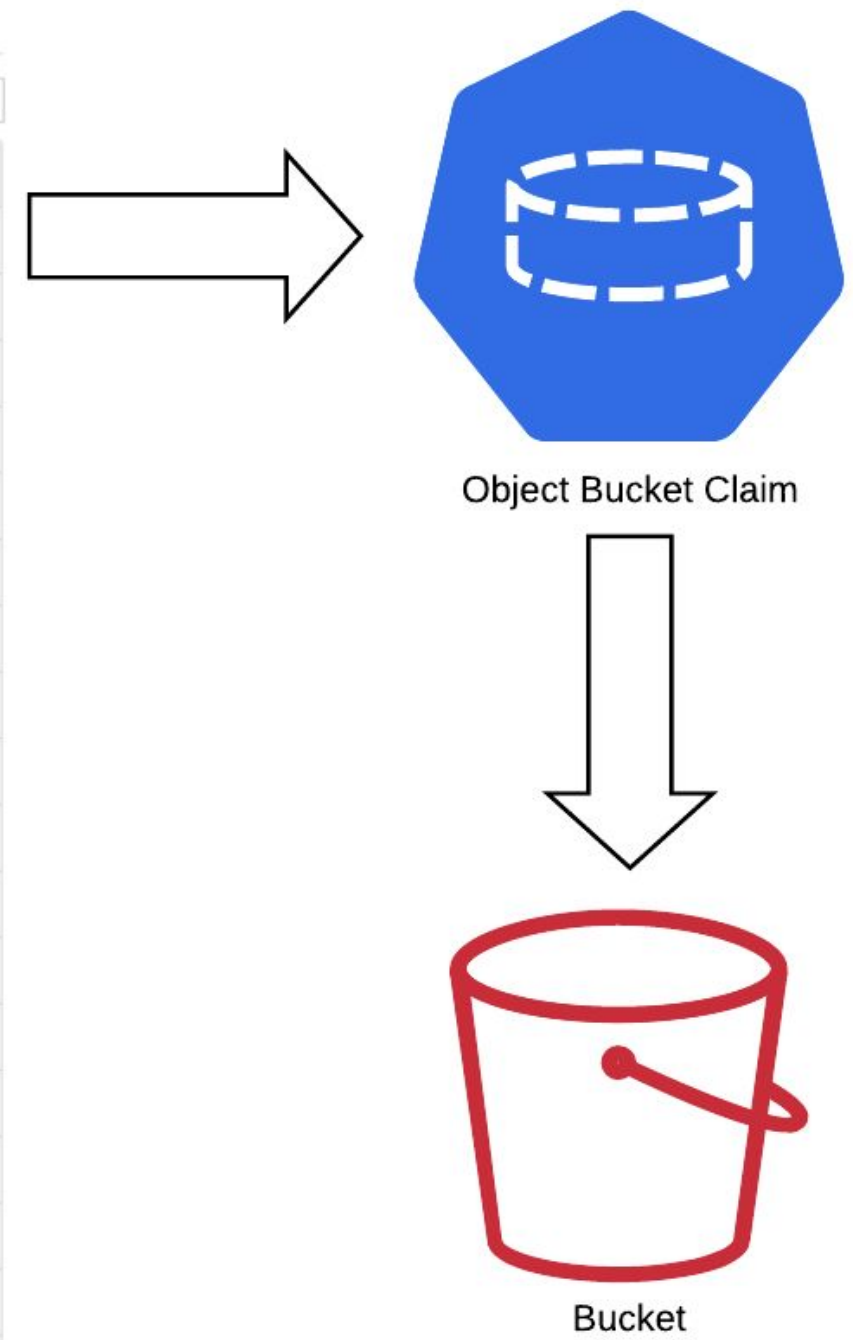


## Open Data Hub with OpenShift Container Storage



The screenshot shows the JupyterLab interface with the 'Files' tab selected. The file browser displays a list of files and folders. The folder 'datalake\_odh-6d70498b-dd8c-428a-bf21-37afcf2e722a' is highlighted with a green box. The interface includes a 'Logout' button and a 'Control Panel' button in the top right corner. Below the file list, there are buttons for 'Upload', 'New', and a refresh icon. The file list has columns for 'Name', 'Last Modified', and 'File size'.

Name	Last Modified	File size
<input type="checkbox"/> /		
<input type="checkbox"/> datalake_odh-6d70498b-dd8c-428a-bf21-37afcf2e722a	50 years ago	
<input type="checkbox"/> lost+found	21 hours ago	
<input type="checkbox"/> tools	21 days ago	
<input type="checkbox"/> anonymize_images.ipynb	a month ago	3.18 kB
<input type="checkbox"/> keras-nn-x-ray-predict-pneumonia-86-54-TF2.ipynb	a month ago	211 kB
<input type="checkbox"/> keras-nn-x-ray-predict-pneumonia-86-54.ipynb	a month ago	10.7 kB
<input type="checkbox"/> make_demo_images.ipynb	a month ago	4.05 MB
<input type="checkbox"/> Test Boto.ipynb	22 days ago	20.7 kB
<input type="checkbox"/> Untitled.ipynb	Running a month ago	72 B
<input type="checkbox"/> Untitled1.ipynb	a month ago	11.7 kB
<input type="checkbox"/> upload_images.ipynb	2 months ago	3 kB
<input type="checkbox"/> x-ray-predict-pneumonia-keras-training.ipynb	a month ago	211 kB
<input type="checkbox"/> X-ray.ipynb	2 months ago	22.1 kB
<input type="checkbox"/> first_try.h5	a month ago	4.88 MB
<input type="checkbox"/> FreeMono.ttf	2 months ago	593 kB
<input type="checkbox"/> notebooks.tar.gz	8 minutes ago	15.2 MB
<input type="checkbox"/> pneumonia_model.h5	a month ago	9.76 MB
<input type="checkbox"/> report.txt	a month ago	1.61 kB



## Prerequisites:

- OCS installed, s3 endpoint (route or service in the openshift-storage project)
- A project where Open Data Hub will be deployed (here "odh")

## Custom JH config (through special config map), which will, for each user:

- Create an Object Bucket Claim if not present,
- Retrieve configuration and access/secret keys,
- Inject everything as environment variables in the user's Pod.

```
oc create cm jupyterhub-cfg2 --from-file=./jupyterhub_config.py -n odh
```

## ODH Deployment:

- jupyterhub\_configmap\_name: jupyterhub-cfg2
- s3\_endpoint\_url: http://s3.openshift-storage/
- storage\_class: ocs-storagecluster-ceph-rbd

```
oc create -f odh-deploy.yaml
```



# Role and Rolebinding:

```
oc apply -f role_jupyterhub-hub-obc.yaml
```

```
---
apiVersion: rbac.authorization.k8s.io/v1
kind: Role
metadata:
  labels:
    app: jupyterhub
    name: jupyterhub-hub-obc
    namespace: odh
rules:
- apiGroups:
  - ""
  resources:
  - secrets
  verbs:
  - get
- apiGroups:
  - objectbucket.io
  resources:
  - objectbucketclaims
  verbs:
  - create
---
apiVersion: rbac.authorization.k8s.io/v1
kind: RoleBinding
metadata:
  name: jupyterhub-hub-obc
  namespace: odh
roleRef:
  apiGroup: rbac.authorization.k8s.io
  kind: Role
  name: jupyterhub-hub-obc
subjects:
- kind: ServiceAccount
  name: jupyterhub-hub
  namespace: odh
```

# Custom Notebooks:

```
#####  
# Directories mapping #  
#####  
import boto3  
from s3contents import S3ContentsManager  
from hybridcontents import HybridContentsManager  
from notebook.services.contents.filemanager import FileContentsManager  
  
# We use HybridContentsManager (https://github.com/viaduct-ai/hybridcontents),  
# FileContentsManager for accessing local volumes  
# and S3ContentsManager (https://github.com/danielfrg/s3contents) to connect to the datalake  
c.NotebookApp.contents_manager_class = HybridContentsManager  
  
# Intialize Hybrid Contents Manager with local filesystem  
c.HybridContentsManager.manager_classes = {  
    # Associate the root directory with a FileContentsManager.  
    # This manager will receive all requests that don't fall under any of the  
    # other managers.  
    '': FileContentsManager  
}  
  
# Get S3 credentials from environment variables  
aws_access_key_id = os.environ.get("AWS_ACCESS_KEY_ID")  
aws_secret_access_key = os.environ.get("AWS_SECRET_ACCESS_KEY")  
endpoint_url = os.environ.get("S3_ENDPOINT_URL")
```

# Result:

### Spawner Options

#### JupyterHub Server Image

Select desired notebook image

- s2i-minimal-s3-notebook:3.6
- s2i-minimal-notebook:3.6
- s2i-minimal-s3-notebook:3.6**
- s2i-scipy-notebook:3.6
- s2i-scipy-s3-notebook:3.6
- s2i-spark-minimal-notebook:3.6
- s2i-spark-scipy-notebook:3.6
- s2i-tensorflow-notebook:3.6

The screenshot shows the JupyterHub interface. At the top left is the Jupyter logo and the word "jupyter". At the top right are "Logout" and "Control Panel" buttons. Below the logo are tabs for "Files", "Running", and "Clusters". The "Files" tab is active. Below the tabs is a message: "Select items to perform actions on them." To the right of this message are "Upload", "New", and a refresh icon. Below this is a table listing files and folders. The first row is highlighted with a green box. The table has columns for "Name", "Last Modified", and "File size".

	Name	Last Modified	File size
<input type="checkbox"/>	0		
<input type="checkbox"/>	📁 datalake_odh-91f8f03e-60d8-4178-8502-dc61ec196e6e	50 years ago	
<input type="checkbox"/>	📁 lost+found	10 days ago	
<input type="checkbox"/>	📄 SimpleS3Connect.ipynb	5 days ago	2.4 kB


# Thank you!

<http://opendatahub.io/>


<http://openshift.com/ai-ml>

<https://github.com/quimou/dataprep-code/odh-ocs>

Red Hat is the world's leading provider of enterprise open source software solutions. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500.

 [linkedin.com/company/red-hat](https://www.linkedin.com/company/red-hat)

 [facebook.com/redhatinc](https://www.facebook.com/redhatinc)

 [youtube.com/user/RedHatVideos](https://www.youtube.com/user/RedHatVideos)

 [twitter.com/RedHat](https://twitter.com/RedHat)

