Powering ODH With Ray

Erik Erlandson, Red Hat, Inc.
eje@redhat.com
@ManyAngled

Or...

Erik Erlandson, Red Hat, Inc.
eje@redhat.com
@ManyAngled

Jupyter & Ray In The Cloud

Erik Erlandson, Red Hat, Inc.
eje@redhat.com
@ManyAngled

Landscape

Ray At 10000 Meters

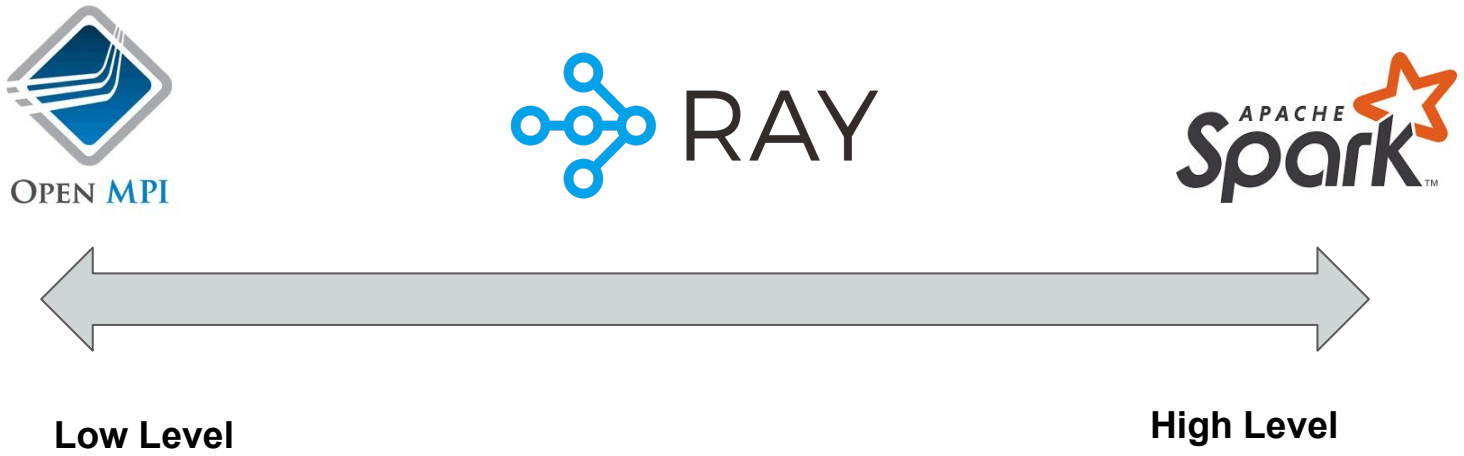
Jupyter and Open Data Hub in Context

Ray on ODH

Demo! 

Community Collaborations

Ray's Ecosystem Niche



Ray's Compute Model

Tasks

```
def f(x):
    # body
```

Actors

```
class foo(object):
    # body
```

Ray's Compute Model

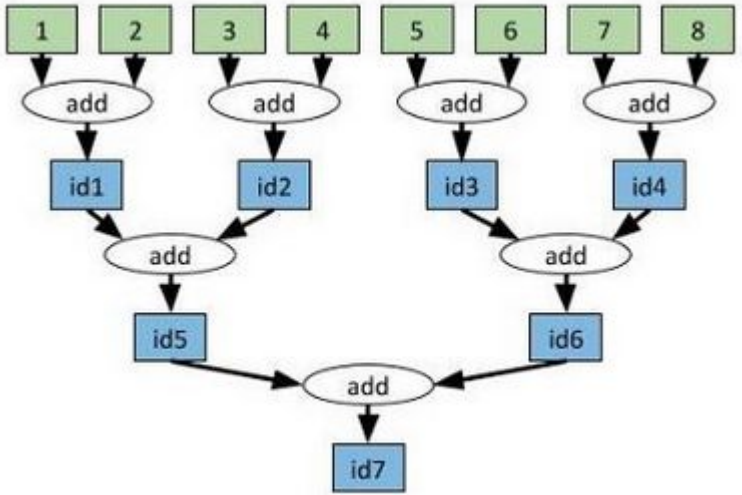
Tasks

```
@ray.remote
def f(x):
    # body
```

Actors

```
@ray.remote
class foo(object):
    # body
```

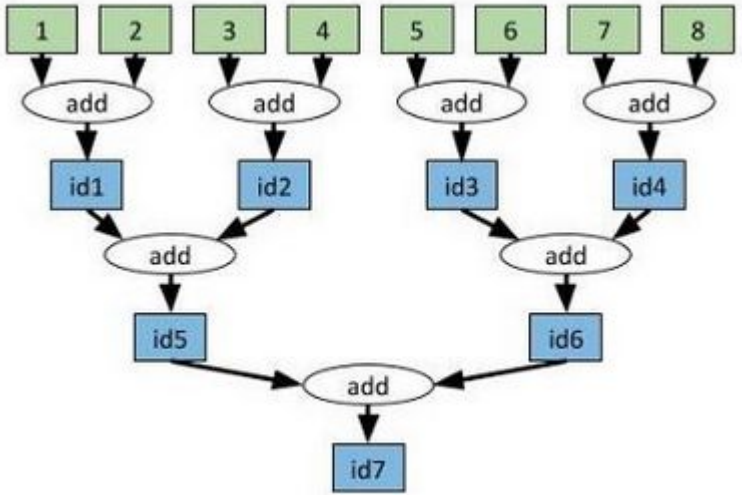
Dependency DAG




```
id1 = add.remote(1, 2)
id2 = add.remote(3, 4)
id3 = add.remote(5, 6)
id4 = add.remote(7, 8)
id5 = add.remote(id1, id2)
id6 = add.remote(id3, id4)
id7 = add.remote(id5, id6)
result = ray.get(id7)
```

Robert Nishihara
https://towardsdatascience.com/modern-parallel-and-distributed-python-a-quick-tutorial-on-ray-99f8d70369b8

Dependency DAG




```
id1 = add.remote(1, 2)
id2 = add.remote(3, 4)
id3 = add.remote(5, 6)
id4 = add.remote(7, 8)
id5 = add.remote(id1, id2)
id6 = add.remote(id3, id4)
id7 = add.remote(id5, id6)
result = ray.get(id7)
```

Robert Nishihara
https://towardsdatascience.com/modern-parallel-and-distributed-python-a-quick-tutorial-on-ray-99f8d70369b8

Ray's Data Model

- Plasma Object Store
- Typeless and Schemaless
- "Local First"
  - Pull remote data when needed
  - Read/Write is all local to worker node

Ray's Scheduling Model

- "Local First"
  - Prefers to run tasks on local scheduler
  - Submits to global scheduler if necessary

Native Ray Libraries

- **Tune**: Scalable Hyperparameter Tuning
- **RLlib**: Scalable Reinforcement Learning
- **RaySGD**: Distributed Training Wrappers
- **Ray Serve**: Scalable and Programmable Serving

Ray Community Integrations

- XGBoost
- Dask
- Horovod
- sklearn
- Spacy
- huggingface

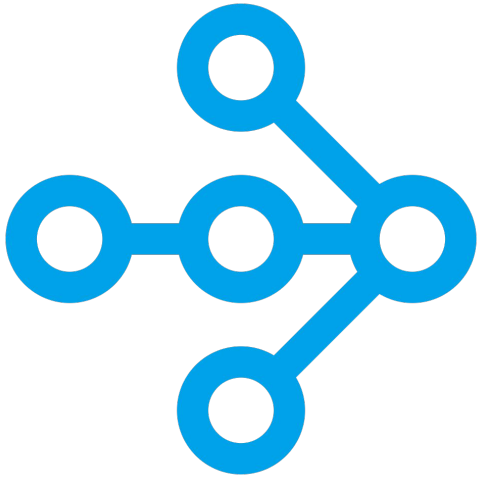
https://docs.ray.io/en/master/ray-libraries.html

Ray Community Integrations

- XGBoost
- Dask
- Horovod
- sklearn
- Spacy
- huggingface

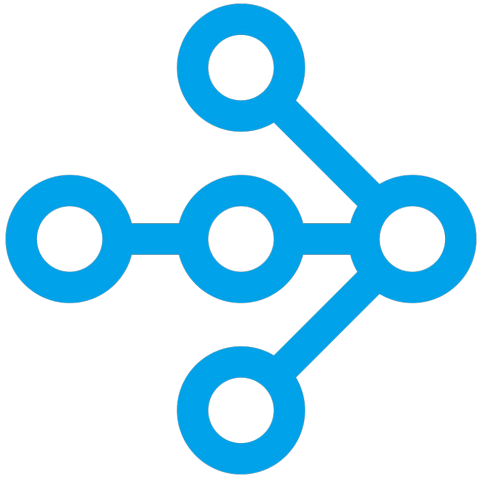
https://docs.ray.io/en/master/ray-libraries.html

Literate And Interactive Ray...



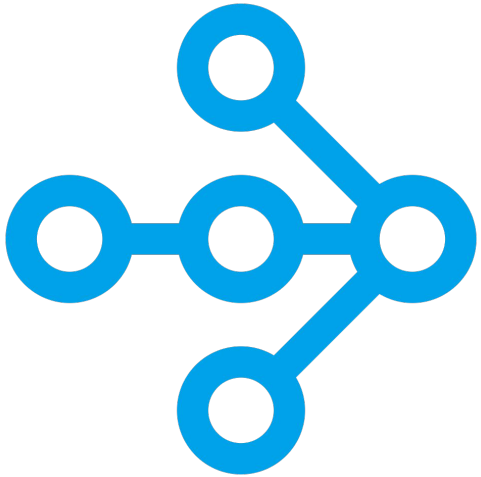
https://docs.ray.io/en/master/ray-libraries.html

Hosted In The Cloud



https://docs.ray.io/en/master/ray-libraries.html

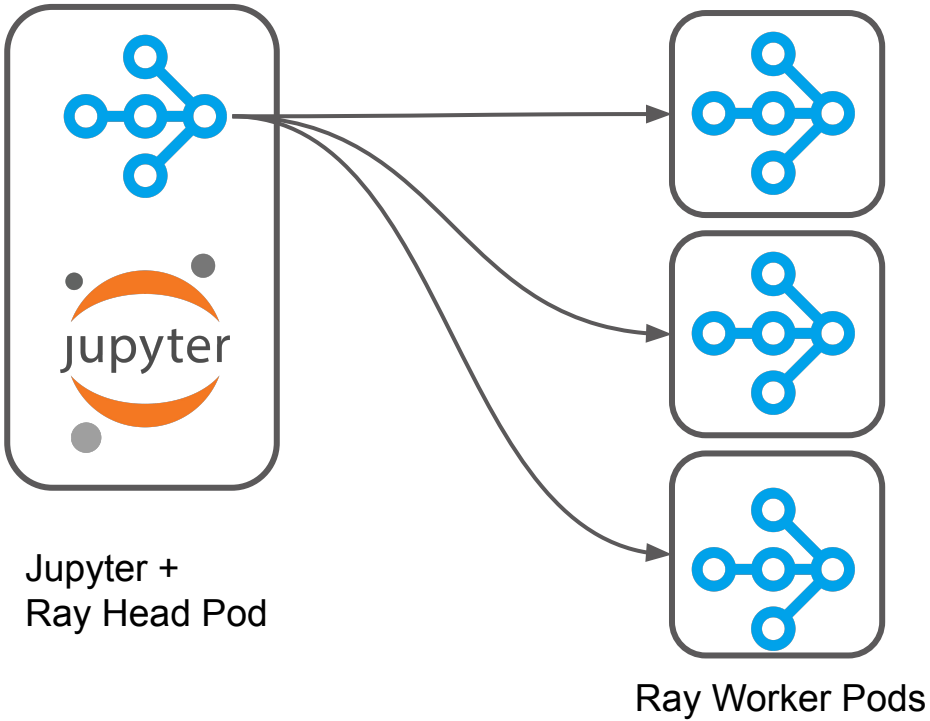
Hosted In The Cloud



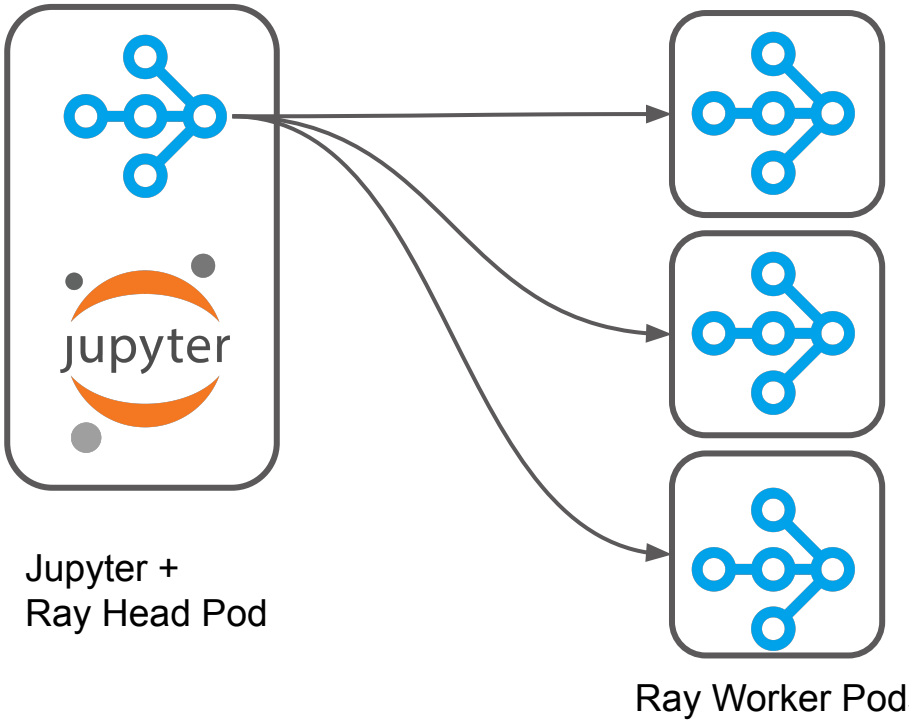
OPENSHIFT

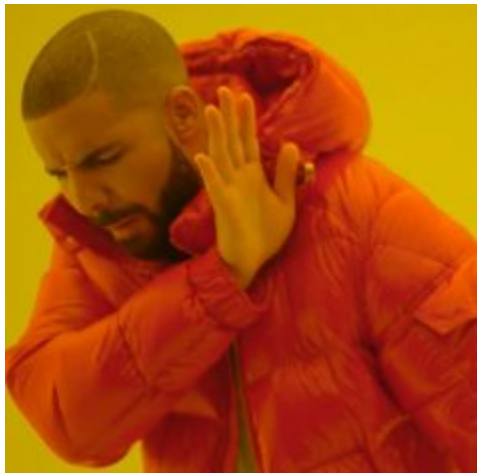
https://docs.ray.io/en/master/ray-libraries.html

Jupyter + ray.init()

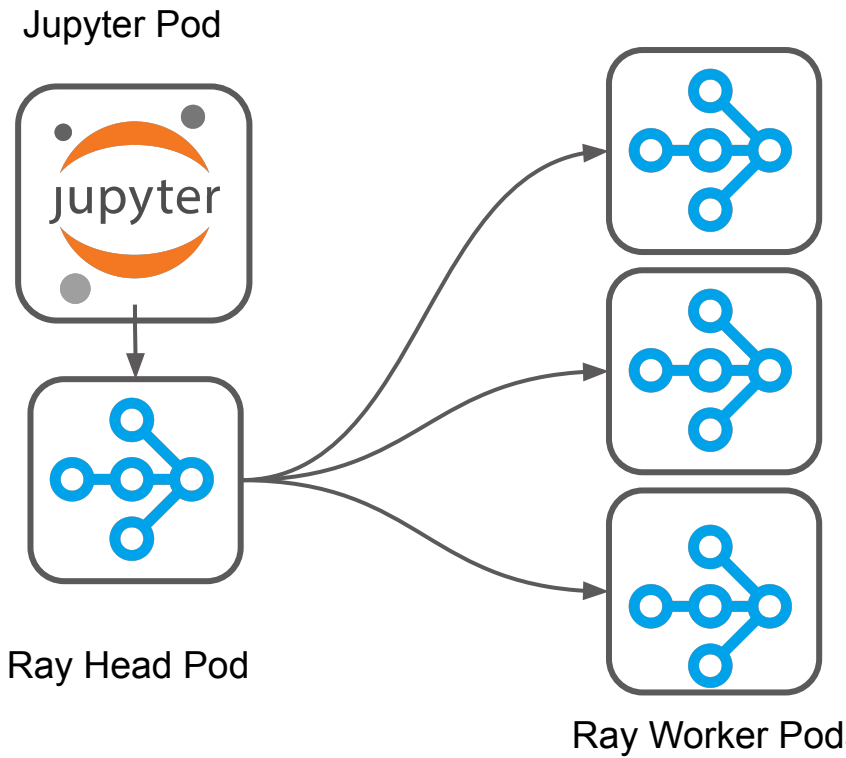


Jupyter + ray.init()



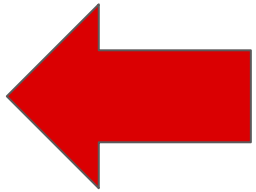


Jupyter + ray.util.connect()



Jupyter ...

Jupyter via Open Data Hub



OPEN DATA HUB

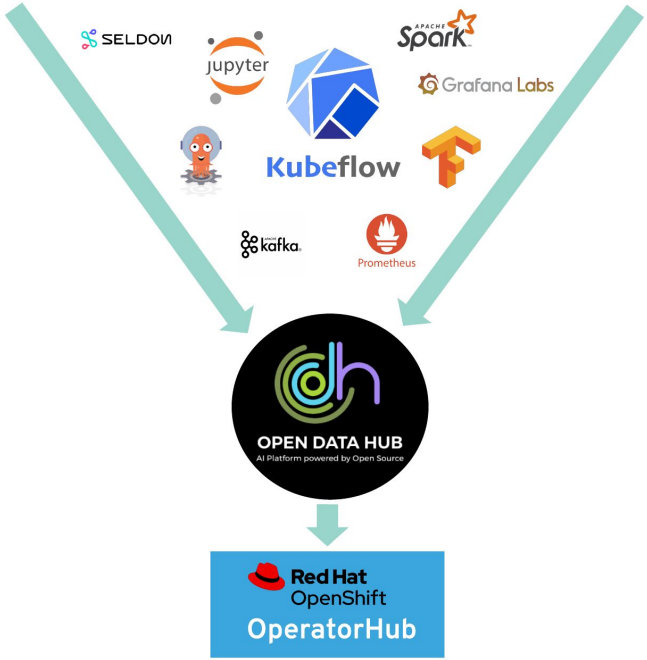
AI Platform powered by Open Source

Open Data Hub Is ...

Open Source Downstream

Reference Platform

Federated

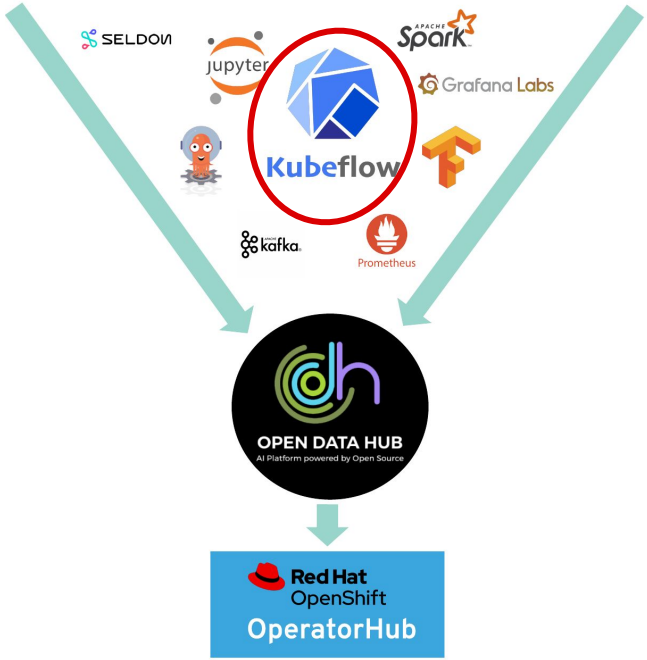


Open Data Hub Is ...

Open Source Downstream

Reference Platform

Federated



Open Data Hub Is ...

Open Source Downstream

Reference Platform

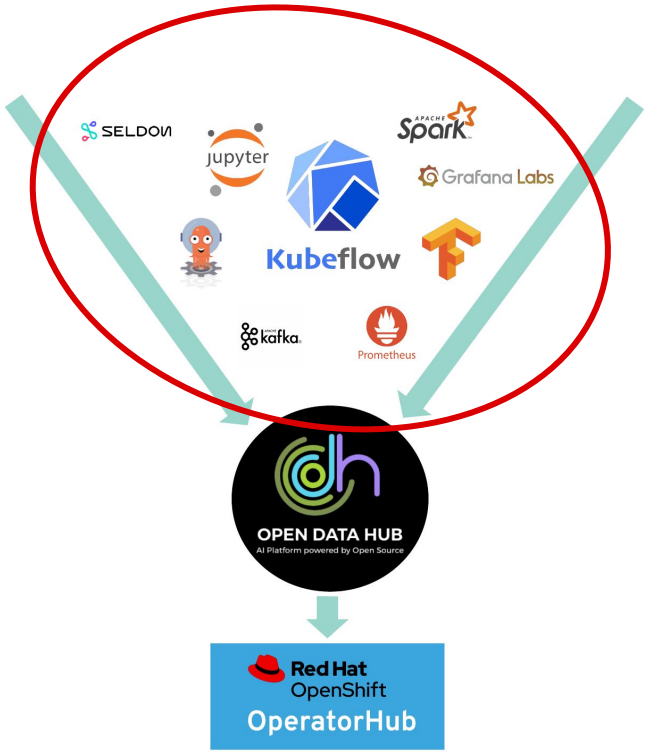
Federated

Open Data Hub Is ...

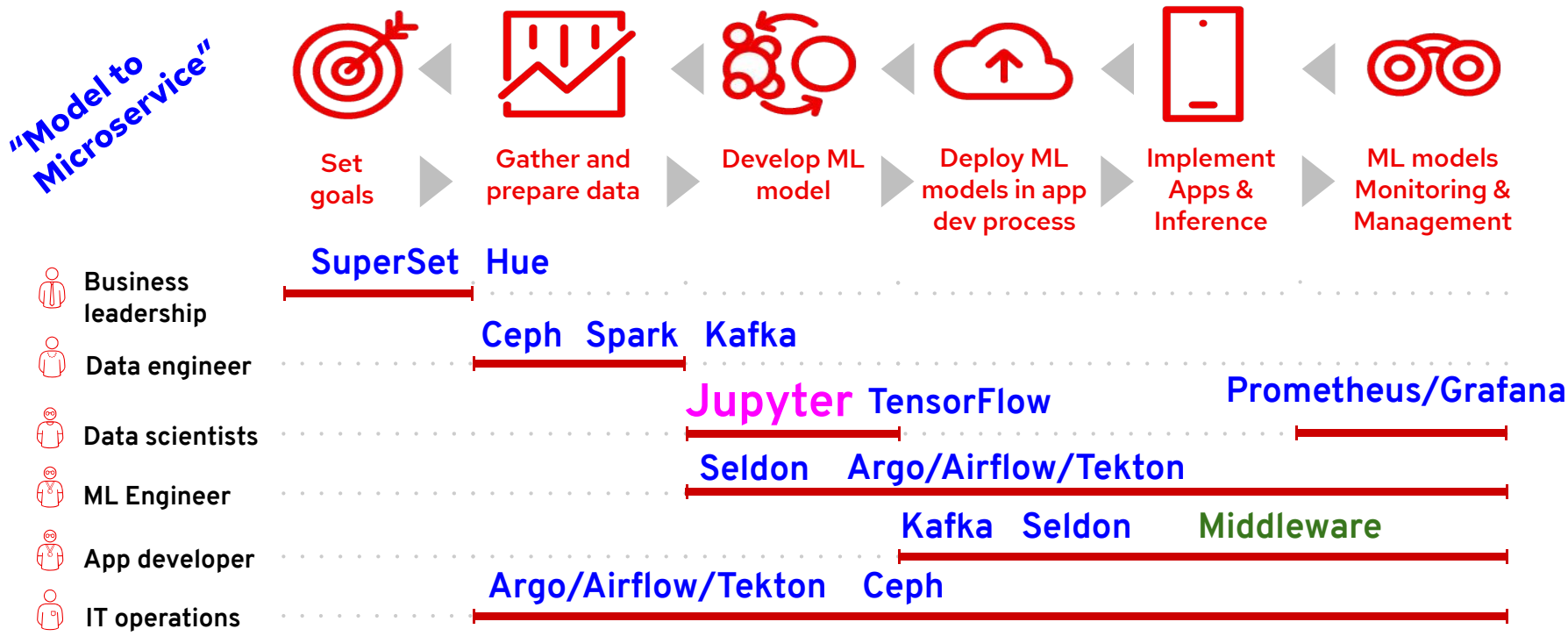
Open Source Downstream

Reference Platform

Federated



Data Science with ODH



Dog-Fooding ODH at Red Hat

Application Logs

Applications in the product release pipeline store their **runtime logs** in our system. These groups are also engaged for **anomaly detection**

Cluster Metrics

Operational metrics from OpenShift clusters. AIOps is engaged here.

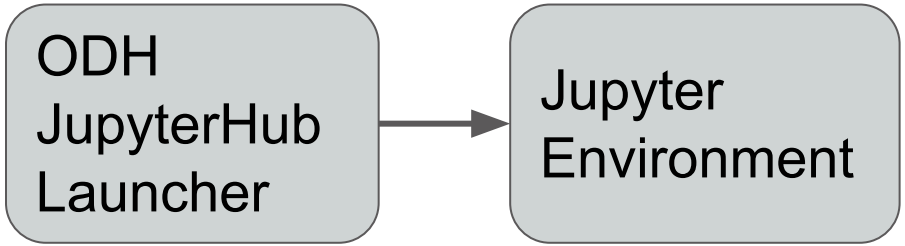
Customer Support Data

Storage of customer data like **SOSReports**, **customer feedback**, etc.

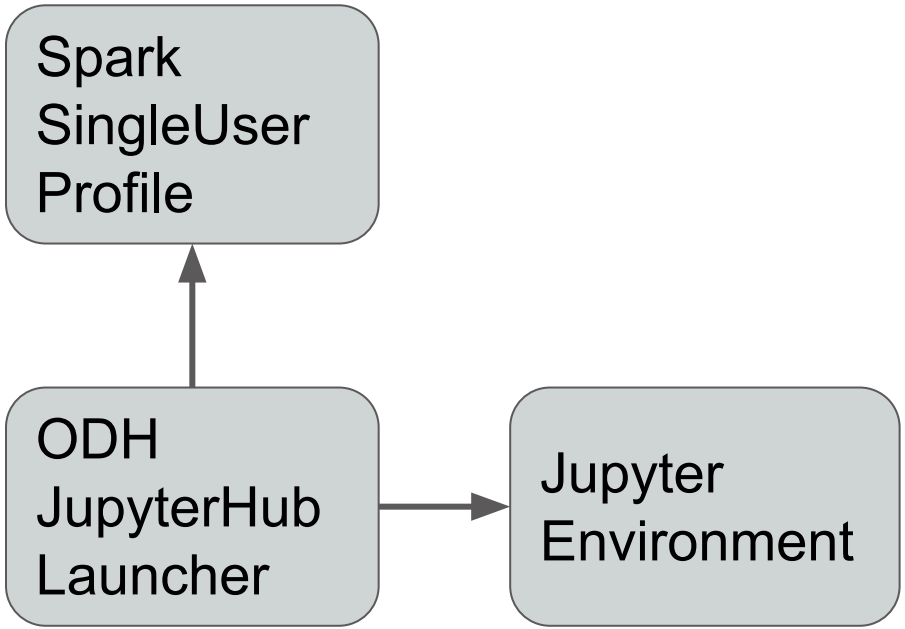
Analogy: Spark on ODH

ODH
JupyterHub
Launcher

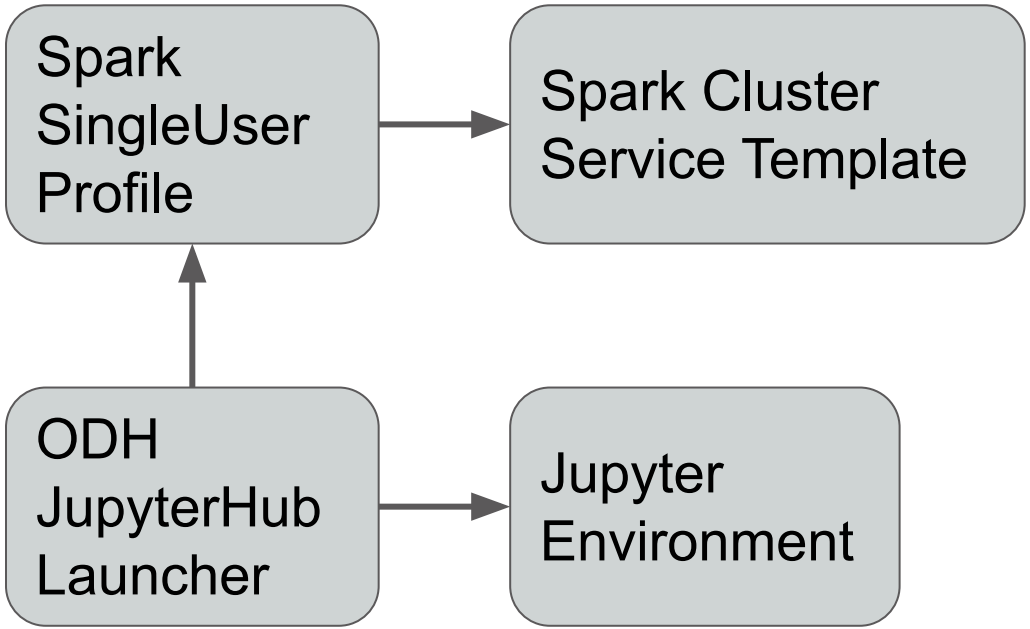
Analogy: Spark on ODH



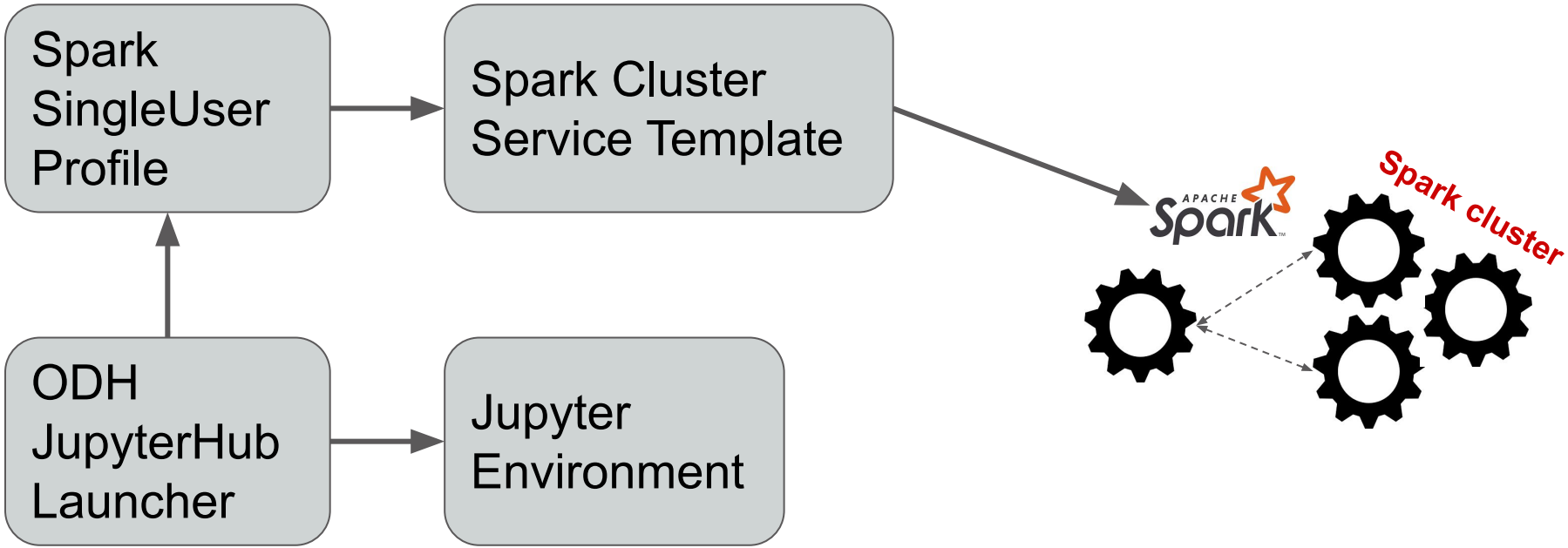
Analogy: Spark on ODH



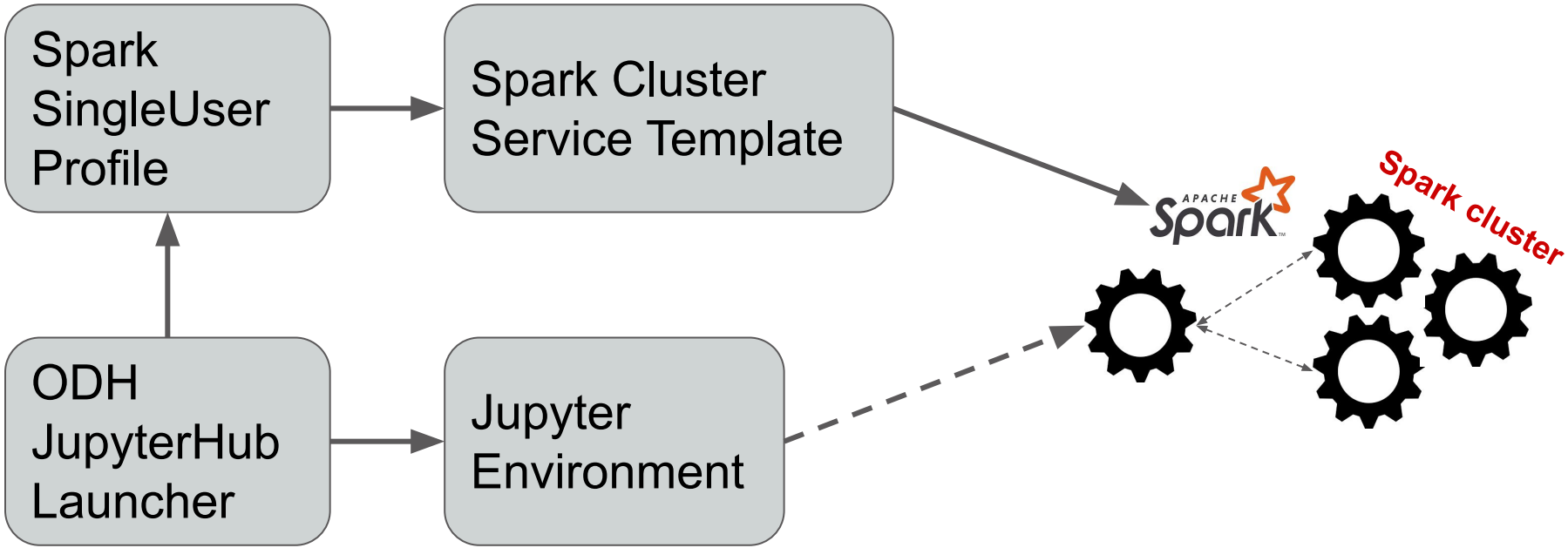
Analogy: Spark on ODH



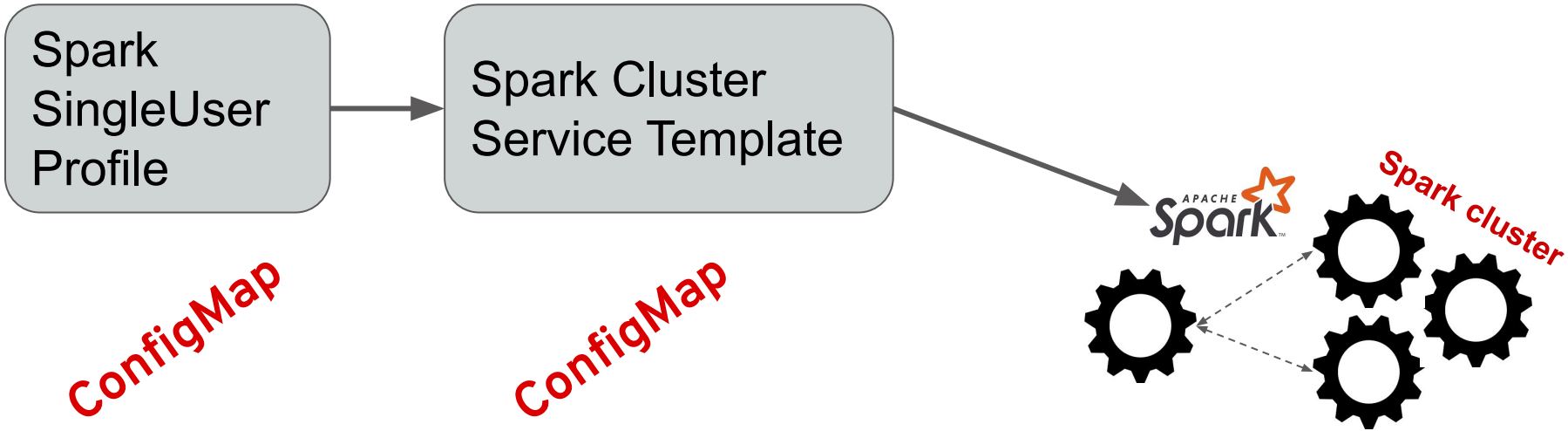
Analogy: Spark on ODH



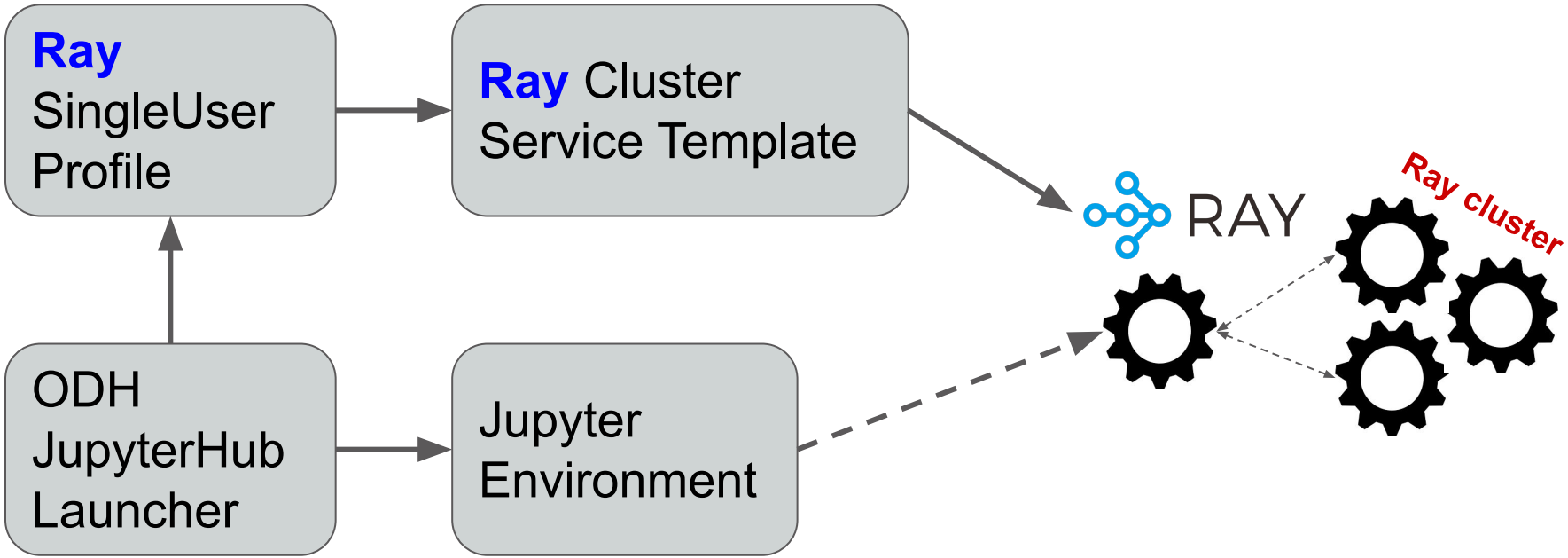
Analogy: Spark on ODH



Analogy: Spark on ODH



Ray on ODH?


Ray
SingleUser
Profile
Ray Cluster
Service Template
ODH
JupyterHub
Launcher
Jupyter
Environment
RAY
Ray cluster

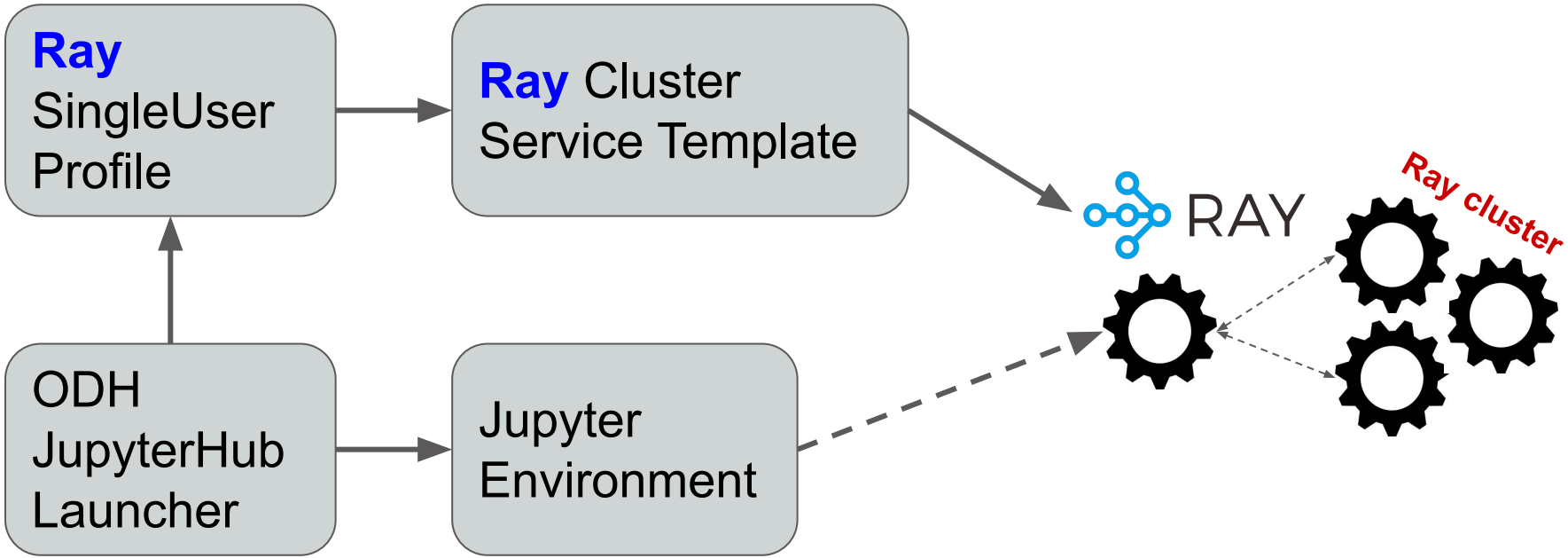
Ray Single User Profile

```
data:
  jupyterhub-singleuser-profiles.yaml: |
    profiles:
      - name: Ray ML Notebook
        images:
          - 'ray-ml-notebook:experimental'
        services:
          ray-cluster:
            resources:
              - name: ray-cluster-template
                path: rayClusterResourceTemplate
            configuration:
              ray_image: 'ray-ml-node:experimental'
              memory_request: '1024Mi'
              cpu_request: '1'
            return:
              RAY_CLUSTER: 'metadata.labels[odh-ray-cluster-service]'
```

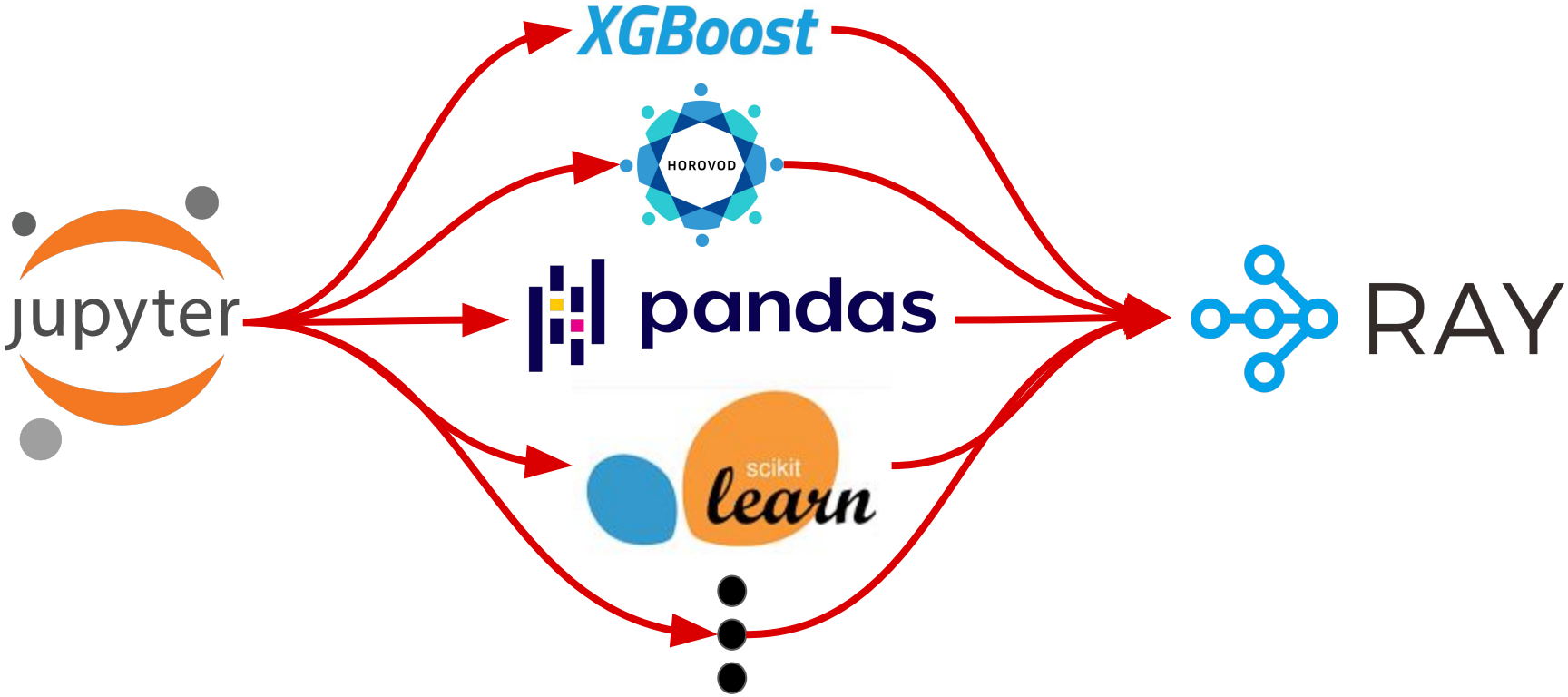
Ray Cluster Service Template

```
data:
  rayClusterResourceTemplate: |
    kind: RayCluster
    apiVersion: cluster.ray.io/v1
    metadata:
      name: 'ray-cluster-{{ user }}'
      labels:
        odh-ray-cluster-service: 'ray-cluster-{{ user }}-ray-head'
    spec:
      maxWorkers: 5
      upscalingSpeed: 1.0
      idleTimeoutMinutes: 5
      # much much more ...
```

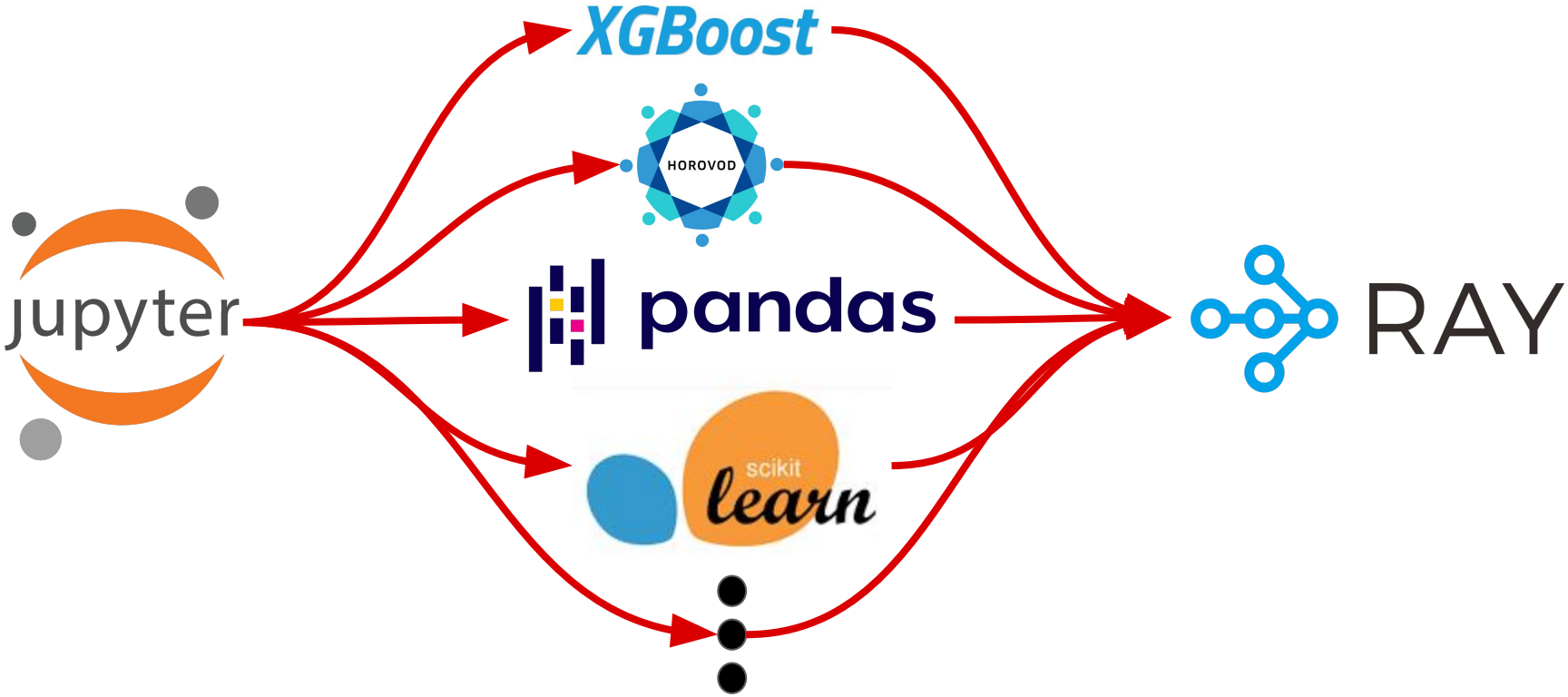
Demo: Ray on ODH!



Unified Scaling for Data Science



~~Unified~~ Simplified Scaling



Ray on ODH at the Mass-Open Cloud

Led by Boston University, the **MOC** is a collaborative effort among BU, Harvard, UMass Amherst, MIT, and Northeastern University, as well as the Massachusetts Green High-Performance Computing Center (MGHPCC) and Oak Ridge National Laboratory (ORNL).

It is supported by a broad alliance of industry partners, including Red Hat.

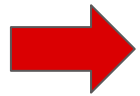
Ray on MOC

- Maximum 5 workers + 1 head
- 1 CPU, 3 GB memory
- Pre-installed:

```
tensorflow==2.4.1
horovod[ray]==0.21.3
keras==2.4.3
scikit-learn==0.24.1
dask==2021.2.0
pandas==1.1.5
scipy==1.5.4
```

Operate First

Developing Software In The Open



Operating Software and Services In the Open

https://www.operate-first.cloud/

Operate First PRs for Ray

The components being installed are:

- [x] the ray notebook imagestream (operate-first/apps#302)
- [x] the ray operator CRD (operate-first/apps#321)
- [x] the ray operator pod itself (operate-first/apps#336)
- [x] a singleuser profile for ray ML notebook (operate-first/apps#361)

Other PRs that were part of this integration:

- operate-first/apps#346
- operate-first/apps#350

https://github.com/operate-first/support/issues/102

Roadmap

- Community Ray Operator in Catalog
- Maintain Ray Images via Project Thoth
- Community Use Cases With Jupyter
- Formal Integration With KF and ODH
- KF Pipeline Nodes Backed by Ray

Call To Action

- Play with Ray on Jupyter up on MOC
- File issues and PRs with op-1st
- Report Back! eje@redhat.com

https://www.operate-first.cloud/users/moc-ray-demo/README.md

https://odh.operate-first.cloud/