ODH Roadmap (draft)

ODH Community Meeting, 2020-04-06

Václav Pavlín, vasek@redhat.com

Current state

- Most of the components converted
  - https://github.com/opendatahub-io/odh-manifests
- Kubeflow Operator basic functionality verified on OpenShift
- Open Data Hub CSV in development
- Already adding new feature for future versions:
  - Airflow - https://github.com/opendatahub-io/odh-manifests/pull/18
  - Redesign of JupyterHub & Spark (or anything else) integration
    - https://github.com/vpavlin/jupyterhub-singleuser-profiles/pull/26
- Discussing the relationship with KF community
  - https://github.com/kubeflow/kubeflow/issues/4916

Future

- KF 1.0 on OpenShift
- Disconnected deployment
- Open Data Hub CI/CD
- Kubeflow on OpenShift CI
- UBI based ODH & KF
- Multitenancy model
- Mixing KF & ODH
- ODH on Power 9

KF 1.0 on OpenShift

- Verify KF 1.0 works
- Provide patches upstream
- Document the verification process so that it can be automated as CI

Disconnected Deployment

- Disconnected/Offline deployment has been heavily requested feature for ODH pre-0.6 and is also important for Kubeflow
- Identify crucial components to start with
- Verify if it is possible to deploy from private registry
- Provide fixes
- Document the process

Open Data Hub CI/CD

- Implement continuous integration for ODH
  - Probably starting with tests for JupyterHub
- Implement continuous deployment for RH Internal Data Hub and Mass Open Cloud
- Prioritize other components for developing tests

Kubeflow on OpenShift CI

- Setup a CI pipeline for testing Kubeflow on OpenShift
- Might not be able to enforce gating on PRs, but run tests at least nightly and provide feedback to contributors

UBI based ODH & KF

- ODH images need to be verified and converted to UBI (Universal Base Image) to follow RH best practices
- There has been requests for UBI based Kubeflow
  - Analyze how many images need to be converted
  - Identify the critical ones to convert first
  - Provide documentation for conversion
  - A pipeline for verification of UBI based versions should be part of CI
  - Make sure Internal Data Hub and Mass Open Cloud uses UBI based versions

Multitenancy model (cluster vs. namespace)

- ODH is namespace scoped, KF is cluster scoped
- Each of the models have pros and cons
- Need to (re)investigate which one makes the most sense to focus on first
- Ideally we should be able to offer both eventually and be able to switch - i.e.
  - Extend ODH to cluster scope
  - Extend KF to namespace scope

Mixing ODH & KF

- We would like to be able to seamlessly mix ODH and KF components
- This is currently problematic - few random examples follow:
  - KF uses Istio, ODH uses OpenShift Routes
  - KF deploys to kubeflow namespace, ODH to opendatahub (KF has `kubeflow` namespace hardcoded for some parts)
  - Authentication/Authorization in OpenShift and KF is based on different premises
- Each divergence needs to be investigated and potential solutions discussed with KF community to see if we can get a solution upstream or need to provide it in ODH downstream

ODH on Power 9

- An initiative lead by Mass Open Cloud
- Be the operator and components running on Power 9 (ppc64le) architecture
- Priorities:
  - Operator
  - JupyterHub
  - Spark Operator
  - Other components