BigData Containers

Project Logistics:

Mentors: Dan McPherson email: dmcphers@redhat.com; email: ;
Min-max team size: 2-4
Expected project hours per week (per team member): 6-8
Will the project be open source: yes

Preferred Past Experience:

Containers/Docker/Kubernetes/OpenShift Nice to have
Git Valuable
Dataverse Nice to have

Project Overview:

Background:
The MOC has enabled researchers to use its 20 petabyte datalake in conjunction with jobs launched on OpenStack. However, many analytics type workloads have been moving to container platforms such as OpenShift (Ex: radanalytics.io) to take advantage of its performance and multi-cloud benefits.

In OpenShift 3.7, a service broker concept has been added to enable OpenShift to provision and connect to internal or external services. These service brokers are based on an open standard (Open Service Broker API) and are shared between Kubernetes and Cloud Foundry.

Project Specifics:
The goal of this project is to build a service broker for the Open Dataverse API on the MOC to enable analytics jobs on OpenShift to consume data from Dataverse. This will enable researchers to use the power of OpenShift (from the MOC or any other cloud) to provide a more optimized compute option that what's currently available.

Some Technologies you will learn/use:

Containers/Docker/Kubernetes/OpenShift
Dataverse
OpenStack/MOC
Software Engineering (Agile/Scrum/Git/etc.)
BigData Analytics (Spark, etc.)