K8s Selection Extension for SWAN

Sahil Jajodia
GSoc Student Developer - CERN HSF (SWAN), 2019
Mentors - Prasanth Kothuri, Piotr Mrowczynski, Enric Tejedor Saavedra, Diogo Castro
Service for Web based Analysis (SWAN)
Features of SWAN

- Analysis only with a web browser
  - No local installation needed
  - Based on Jupyter Notebooks
    - Calculations, input data and results “in the Cloud”
- Support for multiple analysis ecosystems and languages
  - Python, ROOT C++, R and Octave
- Easy sharing of scientific results: plots, data, code
- Integration with CERN resources
  - Software, storage, mass processing power
User Interface
What is spark and how do we use it at SWAN?
Integration with Spark

- Connection to CERN Spark Clusters
  - Spark: general purpose distributed computing framework
- Same environment across platforms (local/remote)
  - User data - EOS
  - Software - CVMFS
- Graphical Jupyter extensions developed
  - Spark Connector
  - Spark Monitor
- Not only used for Physics Analysis at CERN
• Currently used with general purpose Spark K8s Cluster
• No option to use any other cluster
Why do we need K8s Selection?
- Currently SWAN does not support user managed K8s clusters
- Users cannot use their own resources
- Some bottleneck in the cluster can prevent you from analysing data
- Cannot use clusters for personal work
What is K8s Selection?
The main window of the extension.
Add cluster to the config file

- Simple UI
- Different modes
- Extendable
Sharing access to users

Grant access

Username

Email

CreateUser
Why should you use it?
The common use cases are as follows

- Use different clusters for different purposes
- Quickly create cluster, use it and then dispose it
- Create a common cluster for your community
- If you are admin, then share the cluster with other users
- Spring up a cluster if some other cluster is not responding
Future work
- Adding GCloud mode to use GKE clusters for spark
- Creating abstractions (functions) so it is easy to extend
- Integrating it with current method used at SWAN
- Making it useful for services other than spark. E.g Distributed Tensorflow
Demo
Conclusion

- Easy to use extension for SWAN
- Have various use cases
- Provides flexibility during analysis
- Allows user to take advantage of their resources