



Reading CMS Run 1/2 miniAOD files with ServiceX and func_adl

Fellow: Baidyanath Kundu (Manipal
Institute of Technology)

Mentor: Gordon Watts (University of
Washington)

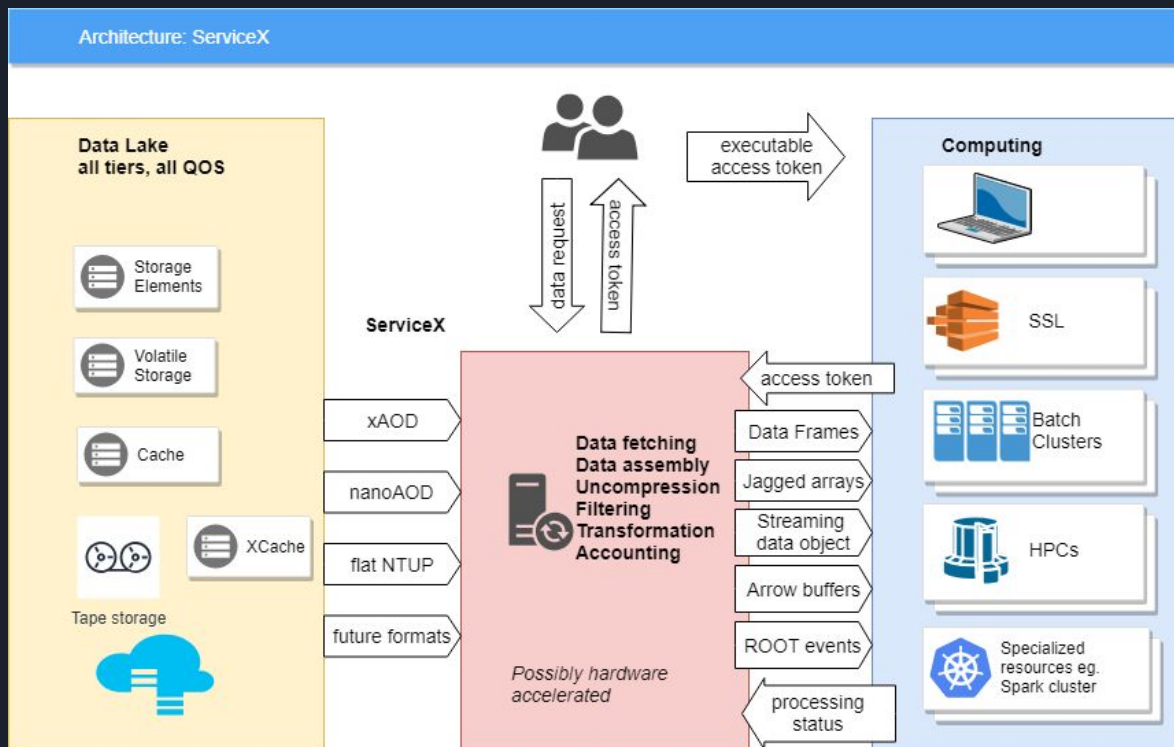
Introduction

- `func_adl_xAOD` was made for ATLAS xAOD format
- The project goal was to add CMS AOD support to the package
- The Higgs Discovery [demo](#) from CMS was used as a guide

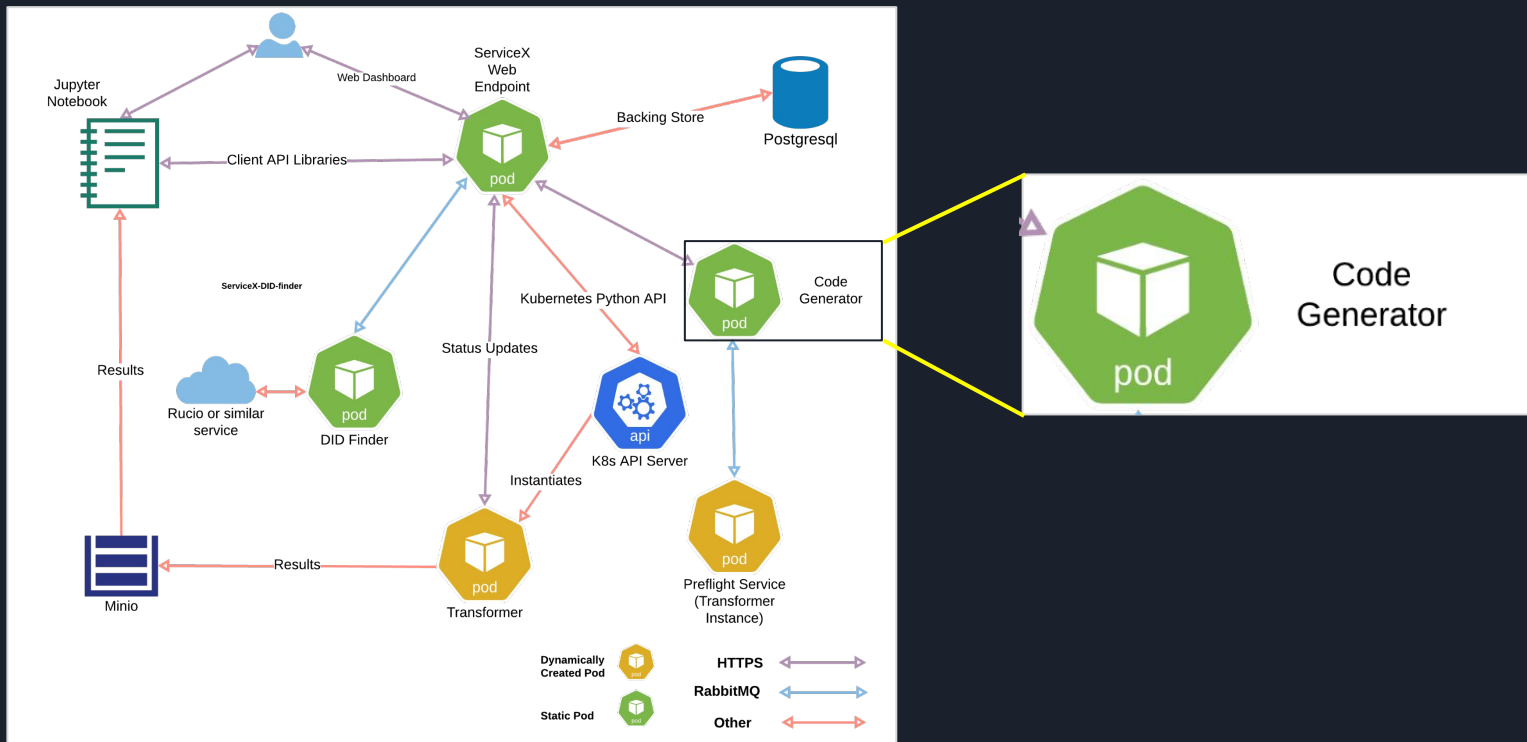
ServiceX: Inspiration



ServiceX: Overview



ServiceX: Architecture



Code Generation

Query Code (Python)

```
result = ServiceXSourceCMSRun1AOD("cernopendata://1507")
    .SelectMany(lambda e: e.TrackMuons("globalMuons"))
    .Select(lambda m: m.pt())
```



Generated Code (C++)

```
...
{
    edm::Handle<reco::TrackCollection> trackmuons0;
    {
        edm::Handle<reco::TrackCollection> result;
        iEvent.getByLabel("globalMuons", result);
        trackmuons0 = result;
    }

    for (auto &i_obj1 : *trackmuons0)
    {
        _col12 = i_obj1.pt();
        myTree->Fill();
    }
}
...
```


Result: ServiceX running on CMS AOD data

Simple CMS Open Data Plot

Demonstrates a simple plot of Muon Track p_T from the [CMS OpenData Higgs to 4 lepton dataset](#). The dataset number (picked from the URL) is 1507.

```
In [1]: from func_adl_servicex import ServiceXSourceCMSRun1AOD
from hist import Hist
import mplhep as mpl

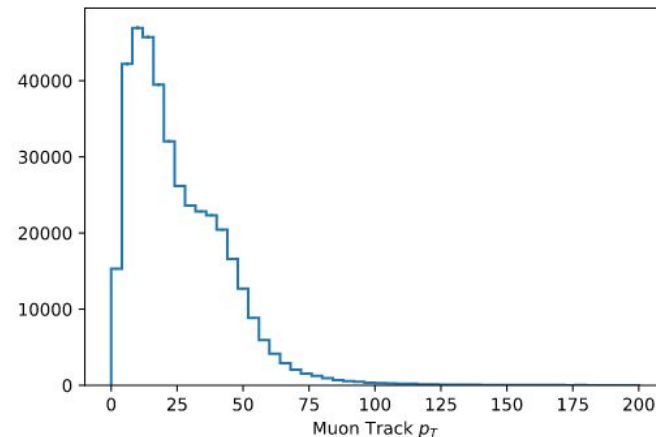
# This is CMS data...
# mpl.style.use(mpl.style.CMS)
```

The following fetches the data - for reference, when testing, on my home machine running a developer version of k8s, and 5 transformers, this took 8 minutes. My network connection was not saturated. **Replace this when we can run on something big like the river cluster.

```
In [2]: data = ServiceXSourceCMSRun1AOD("cernopendata://1507") \
        .SelectMany(Lambda e: e.TrackMuons("globalMuons*)) \
        .Select(Lambda m: m.pt()) \
        .AsAwkwardArray(['mu_pt']) \
        .value()
```

```
cernopendata://1507: 0% | 0/9000000000.0 [00:00]
cernopendata://1507: 0% | 0/9000000000.0 [00:00]
cernopendata://1507: 0% | 0/22 [00:00]
Downloaded: 0% | 0/22 [00:00]
```

```
In [3]: h = (Hist.new
            .Reg(50, 0, 200, name='mu_pt', label='Muon Track $p_T$')
            .Int64()
            )
h.fill(data['mu_pt'])
_ = h.plot()
```



The Notebook can be accessed [here](#)

Approach

- Identify similarities between the two formats (shockingly similar)
- Refactor the existing code to separate the common code from the code specific to each format
- Add the code required to support CMS AOD format
- Add code for CMS AOD type information

Challenges Faced

- Unfamiliarity with CMS software stack (CMSSW)
 - Thanks to Freya Blekman, Jim Pivarski & Brian Cruz for their help to overcome this
- Timezone and technical challenges
- COVID

Future work

- Using validation files for CMS data
- Add type information of more CMS AOD functions
- Add support for more formats. Eg. ATLAS xAOD Run 3

Thank you

Links to the repos contributed to for this project:

[func_adl](#), [func_adl_xAOD](#), [ServiceX Code Generator FuncADL_xAOD](#), [ServiceX_xAOD_CPP_transformer](#), [HiggsDemoTemplated](#)