ServiceX: Backend Tests

David Liu

Gordon Watts

21 September 2020

UNIVERSITY of WASHINGTON

iris hep

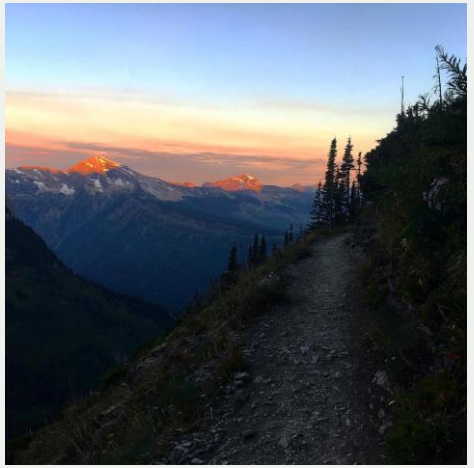
SX

Overview

1. Speaker introduction
2. Discussion of ServiceX
3. Project results and further work

About Me

- Incoming physics graduate student at University of Washington
- Graduated from the Ohio State University December 2019
- Hobbies: backpacking, rock climbing, everything currently inaccessible because of a pandemic



Project Goals

Project goal is to develop a testing suite for ServiceX and to check compatibility between its two backends. This was done successfully over the course of the summer.

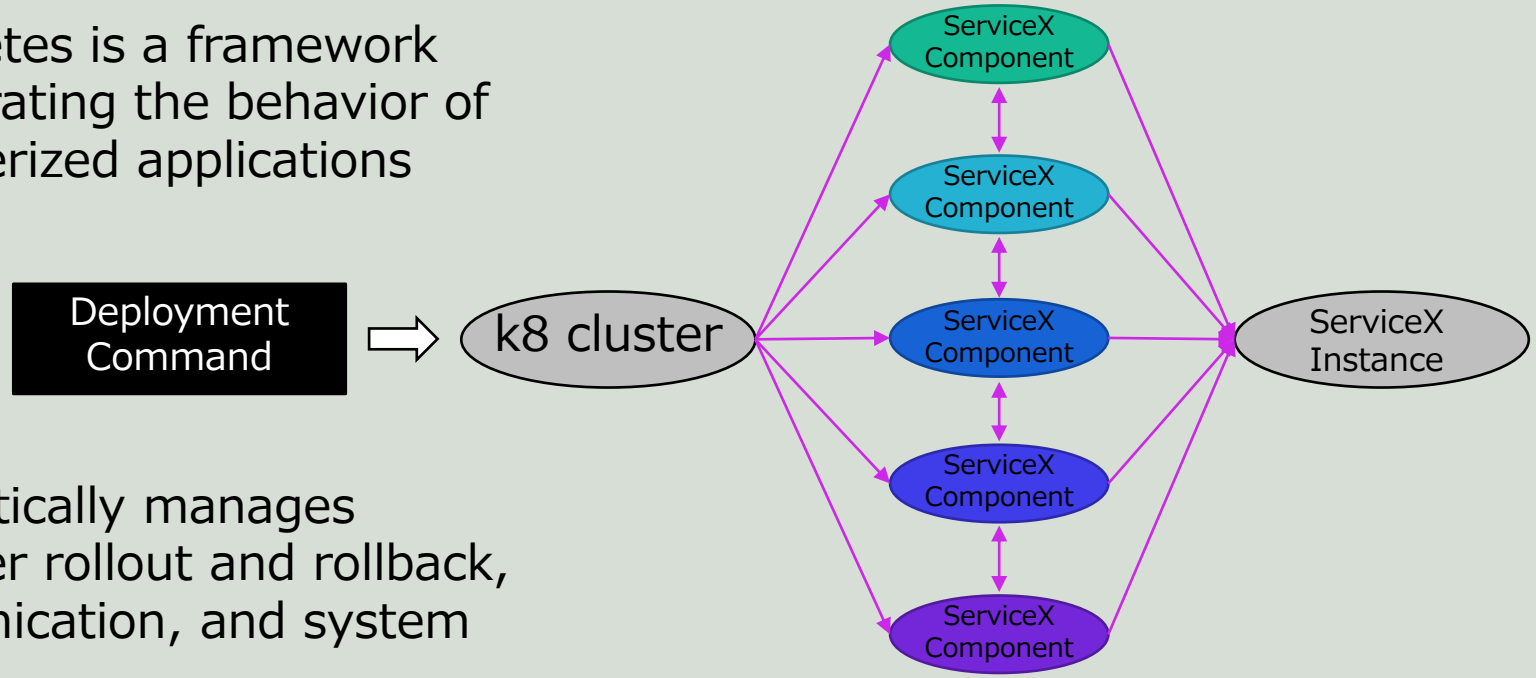
ServiceX: An Overview

- Currently in-development data delivery system using Kubernetes hosted at the University of Chicago
- Selects only desired portions of columnar datasets for user download.
- ServiceX is located close to the data and equipped with a high bandwidth connection to improve speeds
- Reduces user bandwidth, download time, and time spent sorting data, improving efficiency and allowing physicists to do the important thing – analysis!

Photo of the RIVER cluster at UChicago. Photo taken from the public RIVER website.

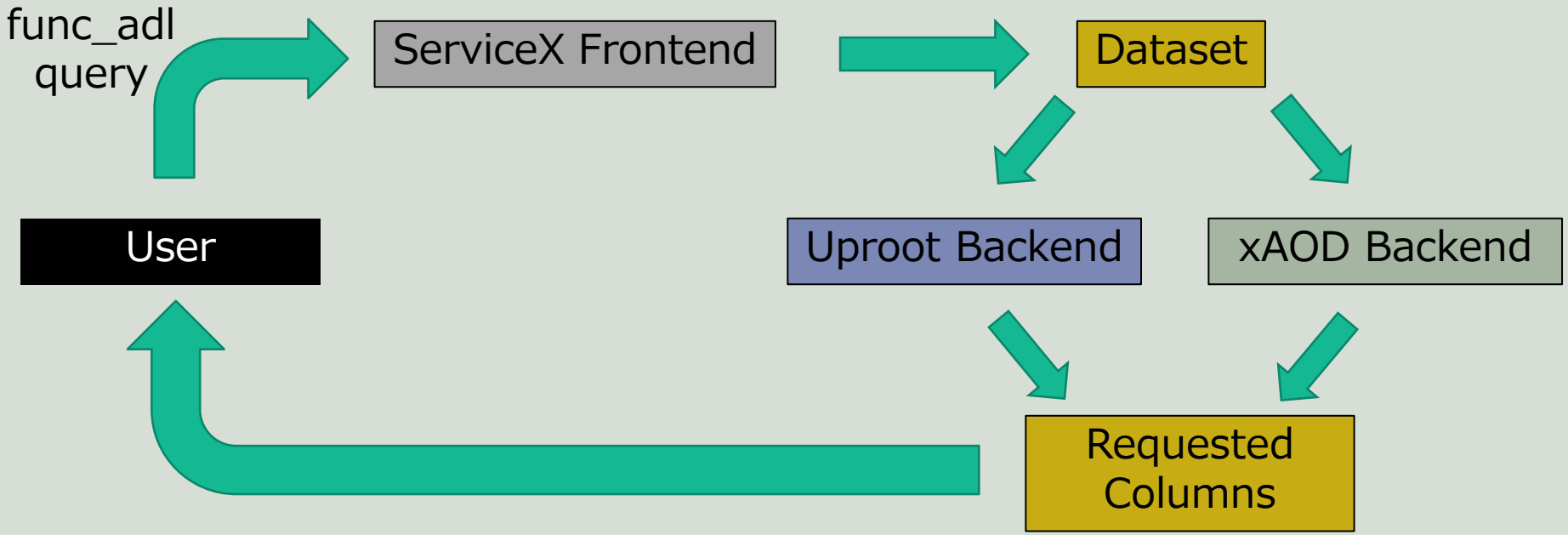
ServiceX and Kubernetes

- Kubernetes is a framework orchestrating the behavior of containerized applications

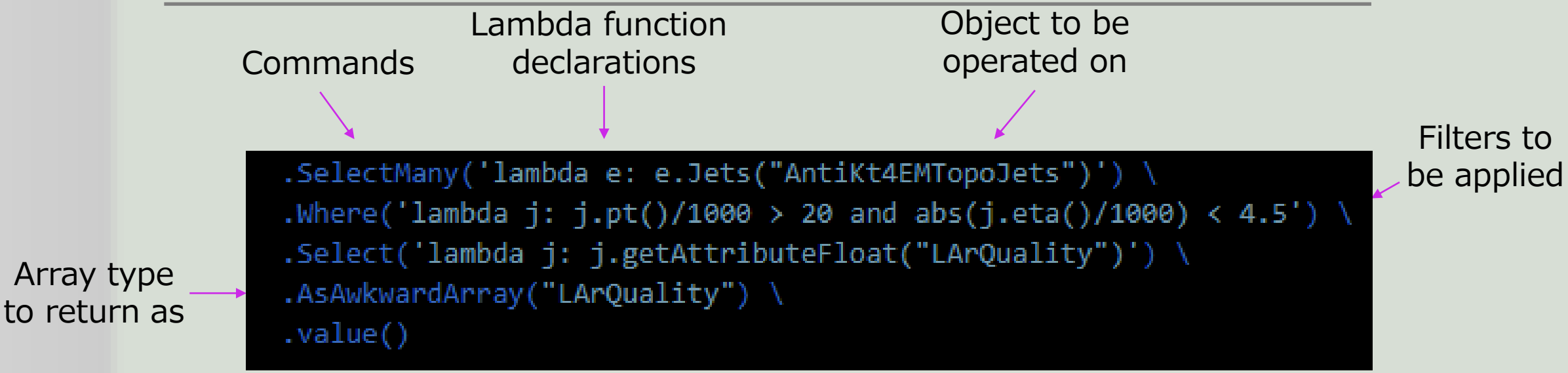


- Automatically manages container rollout and rollback, communication, and system health

ServiceX: Under the Hood



func_adl example



ServiceX: Issues

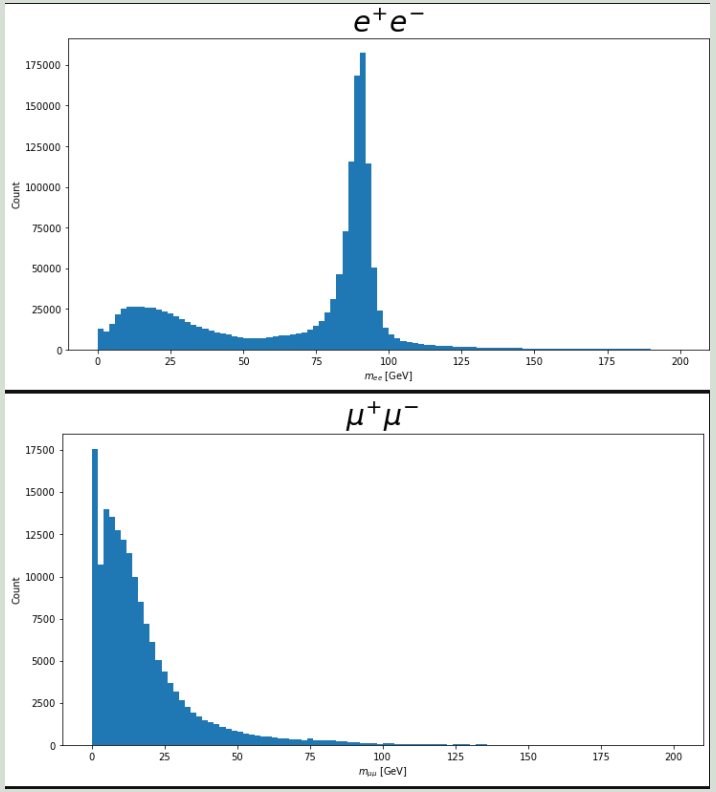
- Uproot and xAOD backends were designed separately, leading to potential incompatibilities
- Project is still in beta and experiences very rapid updates
- As a result, we need a testing kit to make sure system is still functional after updates, to check for backend compatibilities, and for integration tests

Project Results

- Slew of patches for bugs discovered during testing
- Updates to ServiceX functionality and capabilities
  - Support for additional datatypes from xAOD files
  - Implemented LazyArrays
- Improvements to documentation
  - Created documentation for func_adl language
  - Wrote new demos and examples for ServiceX

Project Results

- Developed a test suite using pytest for both the xAOD and uproot backends to rapidly test each functionality:
  - https://github.com/d-w-liu/servicex-backend-tests/tree/master/tests
- Tests ability to retrieve data accurately, ability to retrieve large amounts of data simultaneously, and ability to perform mathematical operations on data during retrieval
- Able to retrieve 100 columns of data from a 10 TB dataset in under 30 minutes



MC Z ➛ ee plots

Further Development

- Ultimately, bundle the testing kit with the official release
- Roll testing kit repo into the official ssl-hep GitHub
- Schedule the testing kit to run automatically
- Bundle testing kit in Docker for easy use

Acknowledgements

- Gordon Watts
- Marc Weinberg
- Ben Galewsky
- Mason Proffitt
- Everyone I interacted with in IRIS-HEP

Questions?

Why func_adl?

- Language-agnostic
- Functional SQL-like language
- Simple and clearly structured