

# **Data Management Package for the novel data delivery system, ServiceX, and Applications to various physics analysis workflows**

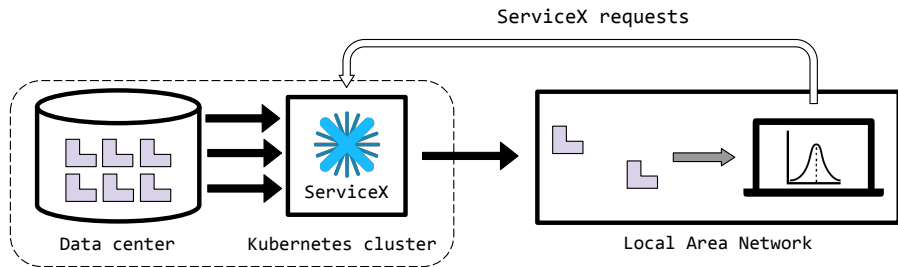
KyungEon Choi, Peter Onyisi

University of Texas at Austin  
Department of Physics

CHEP 2023 @ Norfolk, VA  
May 11, 2023

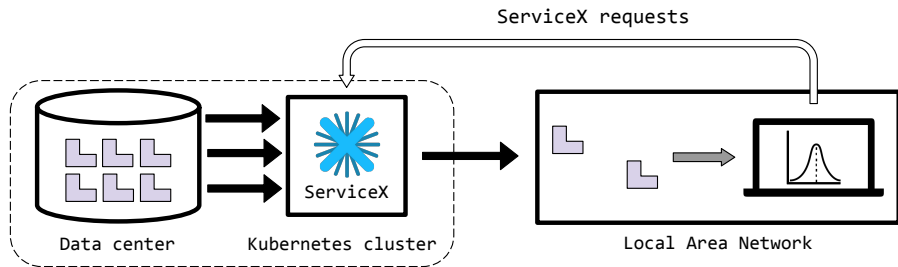


# What's ServiceX?



- A service to **easily** and **quickly** access large data at remote with **transformation**
- Deployed in a Kubernetes cluster and usually co-located with the data center to allow a wide network bandwidth.
- See more at [Ben Galewsky's talk on the ServiceX](#) ↗

# What's ServiceX?



- A service to **easily** and **quickly** access large data at remote with **transformation**
- Deployed in a Kubernetes cluster and usually co-located with the data center to allow a wide network bandwidth.
- See more at [Ben Galewsky's talk on the ServiceX](#) ↗

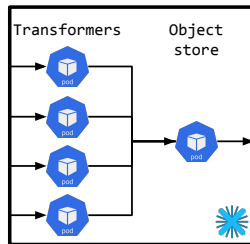
## Why ServiceX?

- ✓ A user wants few columns from datasets in the grid for ML study. Grid jobs?
- ✓ What about full-scale analysis?

# What's ServiceX?

## Under the hood

- 1 ServiceX spawns the so-called **Transformer** pods (Docker container) to read columns and filter events or do whatever it's designed for
- 2 Transformer pods are also scaled automatically using Kubernetes auto-scaling feature
- 3 Outputs from each transformer written to object store then delivered to a destination as soon as it becomes available or consumed later



## Transformers today

Transformer	Input data format
Uproot (uproot)	Flat ROOT ntuple
ATLAS R21 (atlasr21)	ATLAS R21 xAOD
CMS Run1 (cmssw-5-3-32)	CMS Run-1 AOD
Python (python)	*Runs user-provided python function



# What is ServiceX DataBinder?

Python library ( [GitHub](#) , [PyPI](#) )

for **easy** configurations of ServiceX delivery requests  
and handling of delivered data using a single configuration file  
and more

# Use-case: Fast column extraction from remote data for ML study

Base Transformer in this config

File paths of delivered files written in a yaml

4 Samples are defined

Keep configuration tidy!  
→ Definition block

```
1 General:
2   ServiceXName: servicex-uc-af
3   Transformer: uproot
4   OutputFormat: root
5   OutputDirectory: /Users/kchoi/data_for_MLstudy
6   WriteOutputDict: fileset_ml_study
7
8 Sample:
9   - Name: Signal
10     RucioDID: user.kchoi:user.kchoi.fcnc_tHq_ML.ttH.v8,
11              user.kchoi:user.kchoi.fcnc_tHq_ML.ttW.v8,
12              user.kchoi:user.kchoi.fcnc_tHq_ML.ttZ.v8
13     Tree: nominal
14     FuncADL: DEF_ttH_nominal_query
15   - Name: Background1
16     XRootDFiles: DEF_ggH_input
17     Tree: mini
18     Filter: lep_n>2
19     Columns: lep_pt, lep_eta
20   - Name: Background2
21     Transformer: atlasr21
22     RucioDID: DEF_Zee_input
23     FuncADL: DEF_Zee_query
24   - Name: Background3
25     LocalPath: /Users/kchoi/Work/data/fcnc
26
27 Definition:
28   DEF_ttH_nominal_query: "Where(lambda e: e.met_met>150e3). \
29     Select(lambda event: {'el_pt': event.el_pt, 'jet_e': event.jet_e, \
30     'jet_pt': event.jet_pt, 'met_met': event.met_met})"
31   DEF_ggH_input: "root://eospublic.cern.ch/eos/opendata/atlas/OutreachDatasets\
32     /2020-01-22/4lep/MC/mc_345060.ggH125_ZZ4lep.4lep.root"
33   DEF_Zee_input: "mc15_13TeV:mc15_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.\
34     merge.DAOD_STDM3.e3601_s2576_s2132_r6630_r6264_p2363_tid05630052_00"
35   DEF_Zee_query: "SelectMany('lambda e: e.Jets('\AntiKt4EMTopoJets\')). \
36     Where('lambda j: (j.pt() / 1000) > 30'). \
37     Select('lambda j: j.pt() / 1000.0'). \
38     AsROOTTree('junk.root', 'my_tree', ['\JetPt\'])"
```

OutputFormat can be root or parquet

Path where data will be delivered

Remote data via Rucio (RSE)

Multiple inputs separated by comma

TCut syntax for Uproot  
(filter only for scalar type variables)

Background2 is only available as DAOD,  
thus extract using xAOD transformer

Files for Background3 is already  
at local storage → Just bind it together

# Use-case: Fast column extraction from remote data for ML study

Base Transformer in this config

File paths of delivered files written in a yaml

OutputFormat can be root or parquet

Path where data will be delivered

Remote data via Rucio (RSE)

Multiple inputs separated by comma

TCut syntax for Uproot (filter only for scalar type variables)

Background2 is only available as DAOD, thus extract using xAOD transformer

Files for Background3 is already at local storage → Just bind it together

Keep configuration tidy!  
→ Definition block

```
1 General:
2 ServiceXName: servicex-uc-af
3 Transformer: uproot
4 OutputFormat: root
5 OutputDirectory: /Users/kchoi/data_for_MLstudy
6 WriteOutputDict: fileset_ml_study
7
8 Sample:
9 - Name: Signal
10   RucioDID: user.kchoi:user.kchoi.fcnc_tHq_ML.ttH.v8,
11             user.kchoi:user.kchoi.fcnc_tHq_ML.ttW.v8,
12             user.kchoi:user.kchoi.fcnc_tHq_ML.ttZ.v8
13   Tree: nominal
14   FuncADL: DEF_ttH_nominal_query
15 - Name: Background1
16   XRootDFiles: DEF_ggH_input
17   Tree: mini
18   Filter: lep_n>2
19   Columns: lep_pt, lep_eta
20 - Name: Background2
21   Transformer: atlasr21
22   RucioDID: DEF_Zee_input
23   FuncADL: DEF_Zee_query
24 - Name: Background3
25   LocalPath: /Users/kchoi/Work/data/fcnc
26
27 Definition:
28 DEF_ttH_nominal_query: "Where(lambda e: e.met_met>150e3). \
29   Select(lambda event: {'el_pt': event.el_pt, 'jet_e': event.jet_e, \
30   'jet_pt': event.jet_pt, 'met_met': event.met_met})"
31 DEF_ggH_input: "root://eospublic.cern.ch/eos/opendata/atlas/OutreachDatasets\
32   /2020-01-22/4lep/MC/mc_345060.ggH125_ZZ4lep.4lep.root"
33 DEF_Zee_input: "mc15_13TeV:mc15_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.\
34   merge.DAOD_STDM3.e3601_s2576_s2132_r6630_r6264_p2363_tid05630052_00"
35 DEF_Zee_query: "SelectMany('lambda e: e.Jets('\AntiKt4EMTopoJets\')). \
36   Where('lambda j: (j.pt() / 1000) > 30'). \
37   Select('lambda j: j.pt() / 1000.0'). \
38   AsROOTTree('junk.root', 'my_tree', ['JetPt'])"
```

# Use-case: Fast column extraction from remote data for ML study

Base Transformer in this config

File paths of delivered files written in a yaml

4 Samples are defined

Keep configuration tidy!  
→ Definition block

```
1 General:
2   ServiceXName: servicex-uc-af
3   Transformer: uproot
4   OutputFormat: root
5   OutputDirectory: /Users/kchoi/data_for_MLstudy
6   WriteOutputDict: fileset_ml_study
7
8 Sample:
9   - Name: Signal
10     RucioOID: user.kchoi:user.kchoi.fcnc_tHq_ML.ttH.v8,
11              user.kchoi:user.kchoi.fcnc_tHq_ML.ttW.v8,
12              user.kchoi:user.kchoi.fcnc_tHq_ML.ttZ.v8
13     Tree: nominal
14     FuncADL: DEF_ttH_nominal_query
15   - Name: Background1
16     XRootDFiles: DEF_ggH_input
17     Tree: mini
18     Filter: lep_n>2
19     Columns: lep_pt, lep_eta
20   - Name: Background2
21     Transformer: atlasr21
22     RucioOID: DEF_Zee_input
23     FuncADL: DEF_Zee_query
24   - Name: Background3
25     LocalPath: /Users/kchoi/Work/data/fcnc
26
27 Definition:
28   DEF_ttH_nominal_query: "Where(lambda e: e.met_met>150e3). \
29     Select(lambda event: {'el_pt': event.el_pt, 'jet_e': event.jet_e, \
30     'jet_pt': event.jet_pt, 'met_met': event.met_met})"
31   DEF_ggH_input: "root://eospublic.cern.ch/eos/opendata/atlas/OutreachDatasets\
32     /2020-01-22/4lep/MC/mc_345060.ggH125_ZZ4lep.4lep.root"
33   DEF_Zee_input: "mc15_13TeV:mc15_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.\
34     merge.DAOD_STDM3.e3601_s2576_s2132_r6630_r6264_p2363_tid05630052_00"
35   DEF_Zee_query: "SelectMany('lambda e: e.Jets('\AntiKt4EMTopoJets\')). \
36     Where('lambda j: (j.pt() / 1000) > 30'). \
37     Select('lambda j: j.pt() / 1000.0'). \
38     AsROOTTree('junk.root', 'my_tree', ['JetPt'])"
```

OutputFormat can be root or parquet

Path where data will be delivered

Remote data via Rucio (RSE)

Multiple inputs separated by comma

TCut syntax for Uproot  
(filter only for scalar type variables)

Background2 is only available as DAOD,  
thus extract using xAOD transformer

Files for Background3 is already  
at local storage → Just bind it together

# Use-case: Fast column extraction from remote data for ML study

Base Transformer in this config

File paths of delivered files written in a yaml

4 Samples are defined

Keep configuration tidy!  
→ Definition block

```
1 General:
2   ServiceXName: servicex-uc-af
3   Transformer: uproot
4   OutputFormat: root
5   OutputDirectory: /Users/kchoi/data_for_MLstudy
6   WriteOutputDict: fileset_ml_study
7
8 Sample:
9   - Name: Signal
10     RucioID: user.kchoi:user.kchoi.fcnc_tHq_ML.ttH.v8,
11             user.kchoi:user.kchoi.fcnc_tHq_ML.ttW.v8,
12             user.kchoi:user.kchoi.fcnc_tHq_ML.ttZ.v8
13     Tree: nominal
14     FuncADL: DEF_ttH_nominal_query
15   - Name: Background1
16     XRootDFiles: DEF_ggH_input
17     Tree: mini
18     Filter: lep_n>2
19     Columns: lep_pt, lep_eta
20   - Name: Background2
21     Transformer: atlasr21
22     RucioID: DEF_Zee_input
23     FuncADL: DEF_Zee_query
24   - Name: Background3
25     LocalPath: /Users/kchoi/Work/data/fcnc
26
27 Definition:
28   DEF_ttH_nominal_query: "Where(lambda e: e.met_met>150e3). \
29     Select(lambda event: {'el_pt': event.el_pt, 'jet_e': event.jet_e, \
30     'jet_pt': event.jet_pt, 'met_met': event.met_met})"
31   DEF_ggH_input: "root://eospublic.cern.ch/eos/opendata/atlas/OutreachDatasets\
32     /2020-01-22/4lep/MC/mc_345060.ggH125_ZZ4lep.4lep.root"
33   DEF_Zee_input: "mc15_13TeV:mc15_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.\
34     merge.DAOD_STDM3.e3601_s2576_s2132_r6630_r6264_p2363_tid05630052_00"
35   DEF_Zee_query: "SelectMany('lambda e: e.Jets('AntiKt4EMTopoJets')). \
36     Where('lambda j: (j.pt() / 1000) > 30'). \
37     Select('lambda j: j.pt() / 1000.0'). \
38     AsROOTTree('junk.root', 'my_tree', ['JetPt'])"
```

OutputFormat can be root or parquet

Path where data will be delivered

Remote data via Rucio (RSE)

Multiple inputs separated by comma

TCut syntax for Uproot  
(Filter only for scalar type variables)

Background2 is only available as DAOD,  
thus extract using xAOD transformer

Files for Background3 is already  
at local storage → Just bind it together

# Use-case: Fast column extraction from remote data for ML study

Base Transformer in this config

File paths of delivered files written in a yaml

4 Samples are defined

Keep configuration tidy!  
→ Definition block

```
1 General:
2   ServiceXName: servicex-uc-af
3   Transformer: uproot
4   OutputFormat: root
5   OutputDirectory: /Users/kchoi/data_for_MLstudy
6   WriteOutputDict: fileset_ml_study
7
8 Sample:
9   - Name: Signal
10     RucioID: user.kchoi:user.kchoi.fcnc_tHq_ML.ttH.v8,
11             user.kchoi:user.kchoi.fcnc_tHq_ML.ttW.v8,
12             user.kchoi:user.kchoi.fcnc_tHq_ML.ttZ.v8
13     Tree: nominal
14     FuncADL: DEF_ttH_nominal_query
15   - Name: Background1
16     XRootDFiles: DEF_ggH_input
17     Tree: mini
18     Filter: lep_n>2
19     Columns: lep_pt, lep_eta
20   - Name: Background2
21     Transformer: atlasr21
22     RucioID: DEF_Zee_input
23     FuncADL: DEF_Zee_query
24   - Name: Background3
25     LocalPath: /Users/kchoi/Work/data/fcnc
26
27 Definition:
28   DEF_ttH_nominal_query: "Where(lambda e: e.met_met>150e3). \
29     Select(lambda event: {'el_pt': event.el_pt, 'jet_e': event.jet_e, \
30     'jet_pt': event.jet_pt, 'met_met': event.met_met})"
31   DEF_ggH_input: "root://eospublic.cern.ch/eos/opendata/atlas/OutreachDatasets\
32     /2020-01-22/4lep/MC/mc_345060.ggH125_ZZ4lep.4lep.root"
33   DEF_Zee_input: "mc15_13TeV:mc15_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee.\
34     merge.DAOD_STDM3.e3601_s2576_s2132_r6630_r6264_p2363_tid05630052_00"
35   DEF_Zee_query: "SelectMany('lambda e: e.Jets('AntiKt4EMTopoJets')). \
36     Where('lambda j: (j.pt() / 1000) > 30'). \
37     Select('lambda j: j.pt() / 1000.0'). \
38     AsROOTTree('junk.root', 'my_tree', ['JetPt'])"
```

OutputFormat can be root or parquet

Path where data will be delivered

Remote data via Rucio (RSE)

Multiple inputs separated by comma

TCut syntax for Uproot  
(Filter only for scalar type variables)

Background2 is only available as DAOD,  
thus extract using xAOD transformer

Files for Background3 is already  
at local storage → Just bind it together

[FuncADL for Uproot](#)  
[FuncADL for xAOD](#)

# Use-case: Fast column extraction from remote data for ML study

```
In [1]: from servicex_databinder import DataBinder
```

```
In [2]: sxdb = DataBinder('config_MLstudy.yaml')
```

```
INFO - Loading DataBinder config file: config_MLstudy.yaml  
INFO - 4 Samples and 5 ServiceX requests
```

5 ServiceX requests running in parallel

```
In [*]: out = sxdb.deliver()
```

```
INFO - Deliver via ServiceX endpoint: https://servicex.af.uchicago.edu/
```

Signal - nominal: 96% 45/47 [00:28]

Signal - nominal Downloaded: 70% 33/47 [00:26]

Background2: 12% 2/17 [00:25]

Background2 Downloaded: 12% 2/17 [00:25]

Signal - nominal: 46% 37/80 [00:24]

Signal - nominal Downloaded: 46% 37/80 [00:24]

Signal - nominal: 58% 32/55 [00:18]

Signal - nominal Downloaded: 55% 30/55 [00:18]

```
INFO - Background1 | mini | ['root://eospublic.cern.ch//eos/opendata/atlas/OutreachDatasets/2020-01-22/4lep/MC/mc_345060.ggH125_'] is delivered
```

- Rucio datasets for **Signal** in ROOT Ntuple **~200 GB**
- Rucio dataset for **Background2** in ATLAS xAOD **~75 GB**
- Deliver **1-4 columns** to my laptop in Norfolk
- **Wall time: 1min 6s**

# How to run ServiceX DataBinder?

## ❶ Prepare ServiceX access file

- YAML file containing ServiceX endpoint information



```
1 api_endpoints:
2   - name: servicex-uc-af
3     endpoint: https://servicex.af.uchicago.edu/
4     token: <TOKEN>
5     type: uproot
```

## ❷ Prepare DataBinder configuration file



```
1 General:
2   ServiceXName: servicex-uc-af
3   OutputFormat: root
4
5 Sample:
6   - Name: ggH125_ZZ4lep
7     XRootDFiles: "root://eospublic.cern.ch/eos/opendata/atlas/OutreachDatasets\
8                  /2020-01-22/4lep/MC/mc_345060.ggH125_ZZ4lep.4lep.root"
9     Tree: mini
10    FuncADL: "Select(lambda event: {'lep_pt': event.lep_pt, 'lep_eta': event.lep_eta})"
```

## ❸ Deliver!!



```
1 from servicex_databinder import DataBinder
2 sx_db = DataBinder('config_minimum.yaml')
3 out = sx_db.deliver()
```



# How to run ServiceX DataBinder?


In the Jupyter notebook,

```
In [1]: from servicex_databinder import DataBinder
```

```
In [2]: sxdb = DataBinder('config_minimum.yaml')
```

```
INFO - Loading DataBinder config file: config_minimum.yaml
```

```
INFO - 1 Samples and 1 ServiceX requests
```



# How to run ServiceX DataBinder?

In the Jupyter notebook,

```
In [1]: from servicex_databinder import DataBinder
```

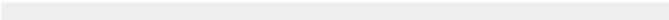
```
In [2]: sxdb = DataBinder('config_minimum.yaml')
```

```
INFO - Loading DataBinder config file: config_minimum.yaml  
INFO - 1 Samples and 1 ServiceX requests
```

```
In [*]: out = sxdb.deliver()
```

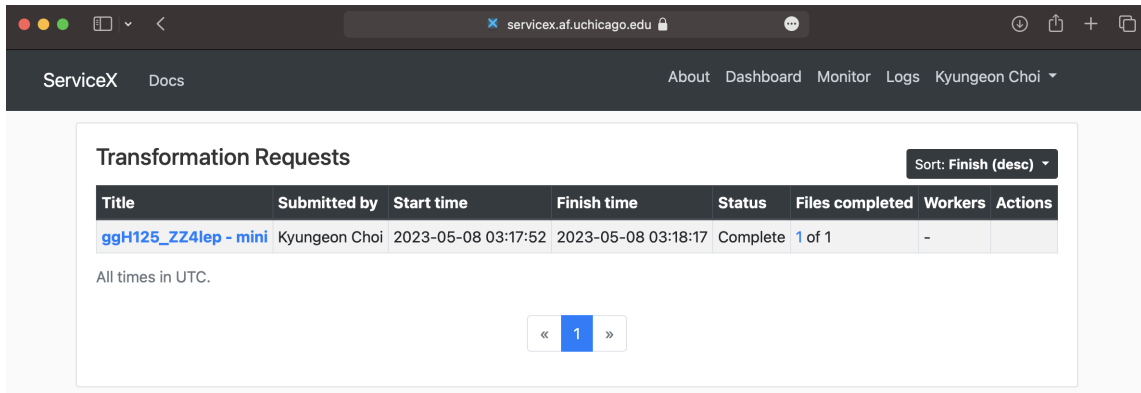
```
INFO - Deliver via ServiceX endpoint: https://servicex.af.uchicago.edu/
```

ggH125\_ZZ4lep - mini: 0%  0/1 [00:05]

ggH125\_ZZ4lep - mini Downloaded: 0%  0/1 [00:05]

# How to run ServiceX DataBinder?

Also from ServiceX Dashboard,



The screenshot shows a web browser window with the URL `servicex.af.uchicago.edu`. The page has a dark header with the ServiceX logo, a 'Docs' link, and navigation links for 'About', 'Dashboard', 'Monitor', 'Logs', and a user profile 'Kyungeon Choi'. The main content area is titled 'Transformation Requests' and includes a 'Sort: Finish (desc)' dropdown. Below this is a table with the following data:

Title	Submitted by	Start time	Finish time	Status	Files completed	Workers	Actions
<a href="#">ggH125_ZZ4lep - mini</a>	Kyungeon Choi	2023-05-08 03:17:52	2023-05-08 03:18:17	Complete	1 of 1	-	

Below the table, it states 'All times in UTC.' and there is a pagination control showing '1' of 1 page.

# How to run ServiceX DataBinder?


In the Jupyter notebook,

```
In [1]: from servicex_databinder import DataBinder
```

```
In [2]: sxdb = DataBinder('config_minimum.yaml')
```

```
INFO - Loading DataBinder config file: config_minimum.yaml  
INFO - 1 Samples and 1 ServiceX requests
```

```
In [3]: out = sxdb.deliver()
```



```
INFO - Deliver via ServiceX endpoint: https://servicex.af.uchicago.edu/
```

```
INFO - ggH125_ZZ4lep | mini | ['root://eospublic.cern.ch//eos/opendata/atlas/OutreachDatasets/2020-01-22/4lep/MC/  
mc_345060.ggH125_'] is delivered
```



```
INFO - Delivered at /Users/kchoi/Work/UTAustin/Computing/ServiceX/ServiceXDataBinder/ServiceXData
```

# How to run ServiceX DataBinder?

In the Jupyter notebook,

```
In [1]: from servicex_databinder import DataBinder
```


```
In [2]: sxdb = DataBinder('config_minimum.yaml')
```

```
INFO - Loading DataBinder config file: config_minimum.yaml  
INFO - 1 Samples and 1 ServiceX requests
```

```
In [3]: out = sxdb.deliver()
```

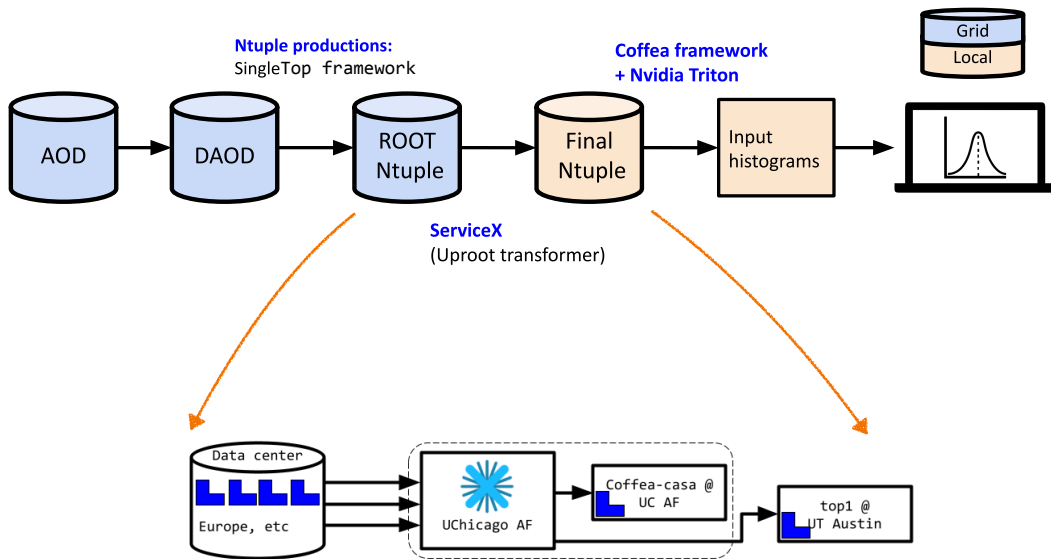
```
INFO - Deliver via ServiceX endpoint: https://servicex.af.uchicago.edu/  
INFO - ggH125_ZZ4lep | mini | ['root://eospublic.cern.ch//eos/opendata/atlas/OutreachDatasets/2020-01-22/4lep/MC/  
mc_345060.ggH125_ is delivered  
INFO - Delivered at /Users/kchoi/Work/UTAustin/Computing/ServiceX/ServiceXDataBinder/ServiceXData
```

```
In [4]: out
```

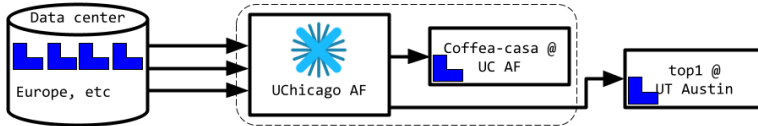


```
Out[4]: {'ggH125_ZZ4lep': {'mini': ['/Users/kchoi/Work/UTAustin/Computing/ServiceX/ServiceXDataBinder/ServiceXData/ggH125_Z  
Z4lep/mini/root__192.170.240.18_1094__root__eospublic.cern.ch__eos_opendata_atlas_OutreachDatasets_2020-01-22_4le  
p_MC_mc_345060.ggH125_ZZ4lep.4lep.root.root']}}}
```

# Use-case: ATLAS Run-2 Physics Analysis

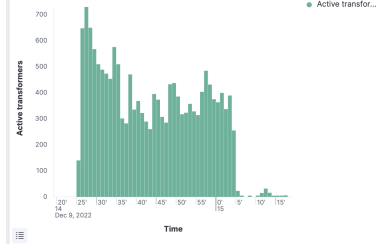


# Use-case: ATLAS Run-2 Physics Analysis



- ServiceX spawns 140 Uproot transformer pods and scaled up/down (up to  $\sim 800$  pods) to extract 30-70 columns from 130 trees out of  $\sim 1.1$  TB ROOT ntuples ( $> 600$  Rucio datasets)
- Wall time:  $\sim 53$  mins to UC Analysis Facility (AF)
- DataBinder configuration file only about 400 lines
- Returned fileset from DataBinder is directly passed to coffea as input
- Accommodate various needs
  - ✓ Delivered data can be also processed with ROOT
  - ✓ Data can be delivered to AF, or University cluster

Active transformers over time

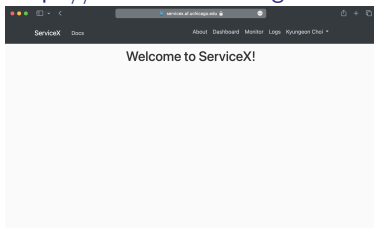




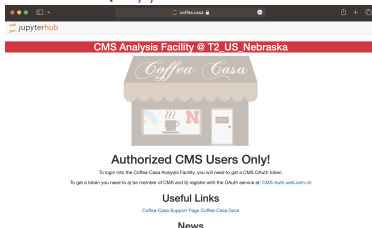


# Where is ServiceX?

<https://servicex.af.uchicago.edu>



<https://coffea.casa>



<https://coffea-opendata.casa>

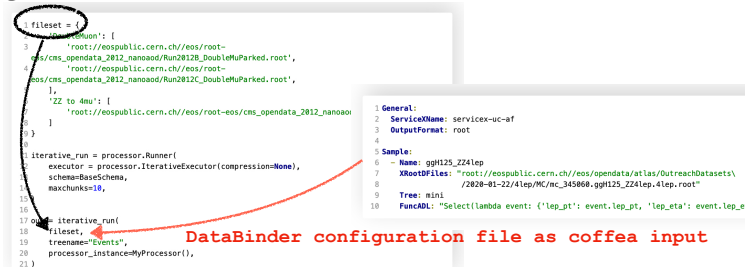


- <https://servicex.af.uchicago.edu> at University of Chicago Analysis Facility for ATLAS users
  - <https://coffea.casa> at University of Nebraska-Lincoln (UNL) for CMS users
  - <https://coffea-opendata.casa> at UNL for OpenData access
- ⇒ ServiceX endpoint accessible only inside UNL coffea-casa facilities but ServiceX access file is pre-generated for users

More about coffea-casa at [Oksana's Plenary talk](#)

# Future plans

- ✓ **Support other ServiceX data delivery options**
  - Streaming from ServiceX object-store or from local path
- ✓ **Better integration into coffea framework**



- ✓ **Support (upcoming) ServiceX updates**
  - multi-tree Uproot transformer, Python transformer, ATLAS R22 transformer, ...
- ✓ **Towards more generic data management package**

# Questions?

Special thanks to the ServiceX team!



Norfolk, Virginia, USA • May 8-12, 2023

**CHEP**  
2023

Computing in High Energy & Nuclear Physics

# Backup

# Delivery of full datasets - a snippet of DataBinder config

## General:

```
ServiceXBackendName: uproot_uc_af
OutputDirectory: /data_ceph/kyungeon/fcnc_tHq_ML/ServiceXData_v7
OutputFormat: root
WriteOutputDict: out_paths_v7
```

## Sample:

```
- Name: ttH
  RucioID: user.kchoi:user.kchoi.fcnc_tHq_ML.ttH.v7
  TransformerImage: kyungeonchoi/servicex_func_adl_uproot_transformer:fcnc_nominal
  Tree: nominal
  FuncADL: DEF_funcadl_prompt_NOMINAL

- Name: ttH
  RucioID: user.kchoi:user.kchoi.fcnc_tHq_ML.ttH.v7
  TransformerImage: kyungeonchoi/servicex_func_adl_uproot_transformer:fcnc_fullsim
  Tree: sys_fullsim
  FuncADL: DEF_funcadl_prompt_SYS

...
```

## Definition:

```
DEF_funcadl_prompt_NOMINAL: "Where(lambda e: e.el_truthIFFClass.Where(lambda i: i==2).Count() == e.el_truthIFFClass.Count())
Where(lambda e: e.mu_truthIFFClass.Where(lambda i: i==4).Count() == e.mu_truthIFFClass.Count()).
Select(lambda e: {'mu_pt': e.mu_pt, 'mu_eta': e.mu_eta, 'mu_phi': e.mu_phi, 'mu_e': e.mu_e, 'mu_charge': e.mu_ch
                  'el_pt': e.el_pt, 'el_eta': e.el_eta, 'el_phi': e.el_phi, 'el_e': e.el_e, 'el_charge': e.el_ch
                  'met_met': e.met_met, 'met_phi': e.met_phi,
                  'jet_pt': e.jet_pt, 'jet_eta': e.jet_eta, 'jet_phi': e.jet_phi, 'jet_e': e.jet_e, 'jet_tagWeig
                  'weights': e.weight_mc*e.weight_pileup*e.weight_leptonSF*e.weight_bTagSF_DL1r_Continuous*e.wei
                  'mcChannelNumber': e.mcChannelNumber, 'runNumber': e.runNumber, 'eventNumber': e.eventNumber,
                  'mc_generator_weights': e.mc_generator_weights,
```

# Delivery of full datasets - Run DataBinder in Jupyter Notebook

```
[2]: from servicex_databinder import DataBinder
```

```
[9]: sx_db = DataBinder('config_fcnc_v7.yml')
```

```
INFO - Loading DataBinder config file: config_fcnc_v7.yml
INFO - 21 Samples and 61 ServiceX requests
```

```
[4]: %time
out = sx_db.deliver()
```

```
INFO - Deliver via ServiceX endpoint: https://uproot-atlas.servicex.af.uchicago.edu/
```

ttW - sys: 82%	<div><div></div></div>	41/50 [11:51]
ttW - sys Downloaded: 82%	<div><div></div></div>	41/50 [11:51]
others_non_prompt - ...: 100%	<div><div></div></div>	1101/1102 [11:41]
others_non_prompt - ... Downloaded: 100%	<div><div></div></div>	1101/1102 [11:41]
others_non_prompt - ...: 97%	<div><div></div></div>	28/29 [11:35]
others_non_prompt - ... Downloaded: 97%	<div><div></div></div>	28/29 [11:35]
ttZ - sys: 63%	<div><div></div></div>	37/59 [11:18]
ttZ - sys Downloaded: 63%	<div><div></div></div>	37/59 [11:18]
others_non_prompt - ...: 42%	<div><div></div></div>	25/59 [11:17]
others_non_prompt - ... Downloaded: 42%	<div><div></div></div>	25/59 [11:17]
others_non_prompt - ...: 8%	<div><div></div></div>	56/692 [11:19]
others_non_prompt - ... Downloaded: 8%	<div><div></div></div>	56/692 [11:19]
others_non_prompt - ...: 94%	<div><div></div></div>	44/47 [11:21]
others_non_prompt - ... Downloaded: 94%	<div><div></div></div>	44/47 [11:21]
others_prompt - sys: 59%	<div><div></div></div>	216/365 [11:23]
others_prompt - sys Downloaded: 59%	<div><div></div></div>	216/365 [11:23]
others_prompt - sys: 15%	<div><div></div></div>	8/54 [11:14]
others_prompt - sys Downloaded: 15%	<div><div></div></div>	8/54 [11:14]
others_prompt - sys: 40%	<div><div></div></div>	57/142 [11:03]
others_prompt - sys Downloaded: 40%	<div><div></div></div>	57/142 [11:03]
others_non_prompt - ...: 30%	<div><div></div></div>	111/365 [11:10]
others_non_prompt - ... Downloaded: 30%	<div><div></div></div>	111/365 [11:10]

# Delivery of full datasets - Dashboard

Transformation Requests							Sort: Finish (desc) ▾
Title	Submitted by	Start time	Finish time	Status	Files completed	Workers	Actions
<a href="#">others_non_prompt - sys</a>	Kyungeon Choi	2022-12-09 05:25:00	-	Running <div><div></div></div>	30 of 365	62	<button>Cancel</button>
<a href="#">others_non_prompt - sys</a>	Kyungeon Choi	2022-12-09 05:24:59	-	Running <div><div></div></div>	5 of 59	30	<button>Cancel</button>
<a href="#">others_prompt - sys</a>	Kyungeon Choi	2022-12-09 05:25:00	-	Running <div><div></div></div>	6 of 142	34	<button>Cancel</button>
<a href="#">fcnc_tHc_prod - sys</a>	Kyungeon Choi	2022-12-09 05:25:01	-	Running <div><div></div></div>	1 of 34	19	<button>Cancel</button>
<a href="#">ttZ - sys</a>	Kyungeon Choi	2022-12-09 05:24:59	-	Running <div><div></div></div>	6 of 59	38	<button>Cancel</button>
<a href="#">others_non_prompt - sys</a>	Kyungeon Choi	2022-12-09 05:25:01	-	Running <div><div></div></div>	4 of 142	22	<button>Cancel</button>
<a href="#">others_prompt - sys</a>	Kyungeon Choi	2022-12-09 05:24:59	-	Running <div><div></div></div>	111 of 1102	158	<button>Cancel</button>
<a href="#">others_non_prompt - sys</a>	Kyungeon Choi	2022-12-09 05:24:59	-	Running <div><div></div></div>	5 of 692	44	<button>Cancel</button>
<a href="#">others_prompt - sys</a>	Kyungeon Choi	2022-12-09 05:24:59	-	Running <div><div></div></div>	43 of 365	68	<button>Cancel</button>