



Monitoring DUNE Data Streaming Efficiency

Lisa Paton

Supervisor: Steven Timm

11 August 2021

SIST Final Presentation

Background

- DUNE (Deep Underground Neutrino Experiment) has
 - International Collaboration
 - 36 Computing Sites
 - 15 of 36 have Storage Capacity
 - Data streaming using xroot
 - Xroot – TCP based protocol for file transfers
 - Specifically designed to stream physics analysis files (ROOT Branches)
 - Sends only what you need, as you need it (like Netflix)

Project Goals

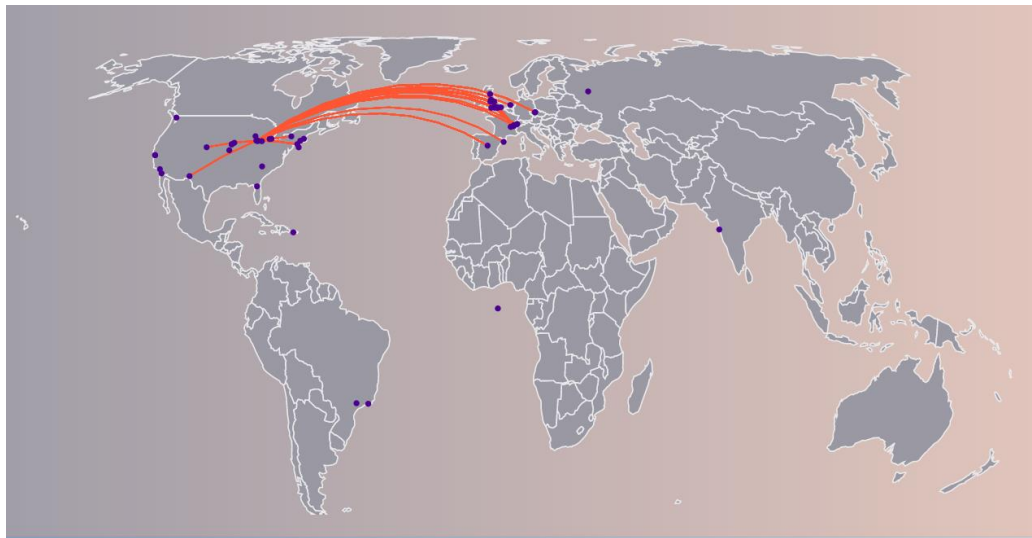
Monitor/Analyze transfer data

- Relevant info
 - Who is running jobs?
 - Where is their data stored?
 - How much data are they streaming?
 - How fast is the data streaming?
- Identify bottlenecks

SAM Data Transfers Visualization

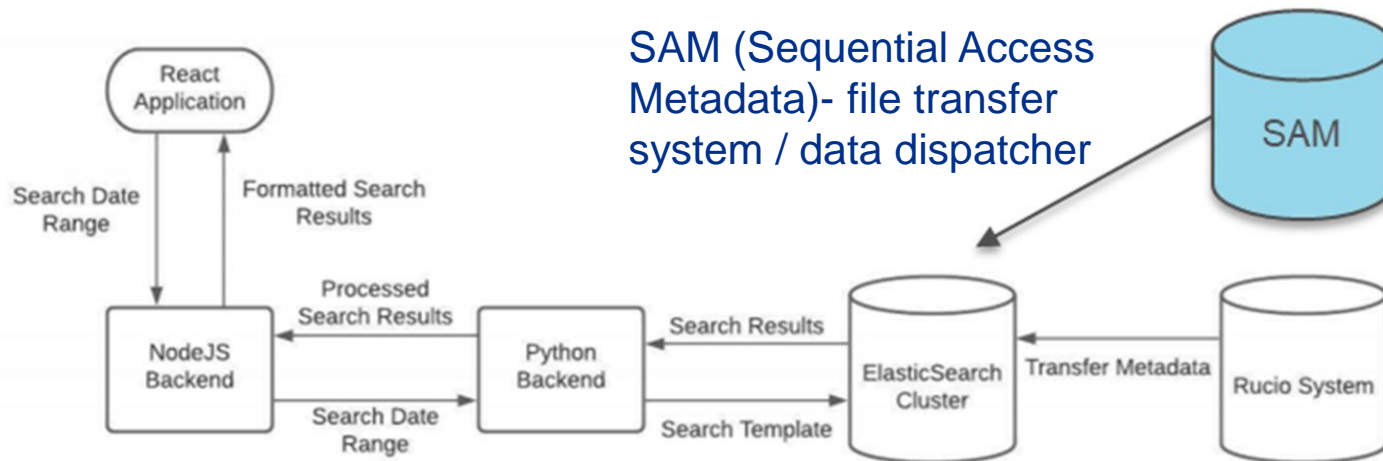
Mapping SAM transfers

- React App
 - Free, open-source library to build UI's using JavaScript
 - Tailored by Oregon State University Students to display transfers
 - Zac Lee, Lydia Brynmoor, Luke Penner, working with Dr. Schellman
 - Fast and intuitive
 - Better suited for network monitoring than Kibana
- Branched off
 - Altered Python backend
 - Query xroot instead of Rucio
 - Implemented caching system for faster loading



React App display of SAM transfers over 5 days. Blue dots show DUNE Computing Sites. Red lines show data streaming paths.

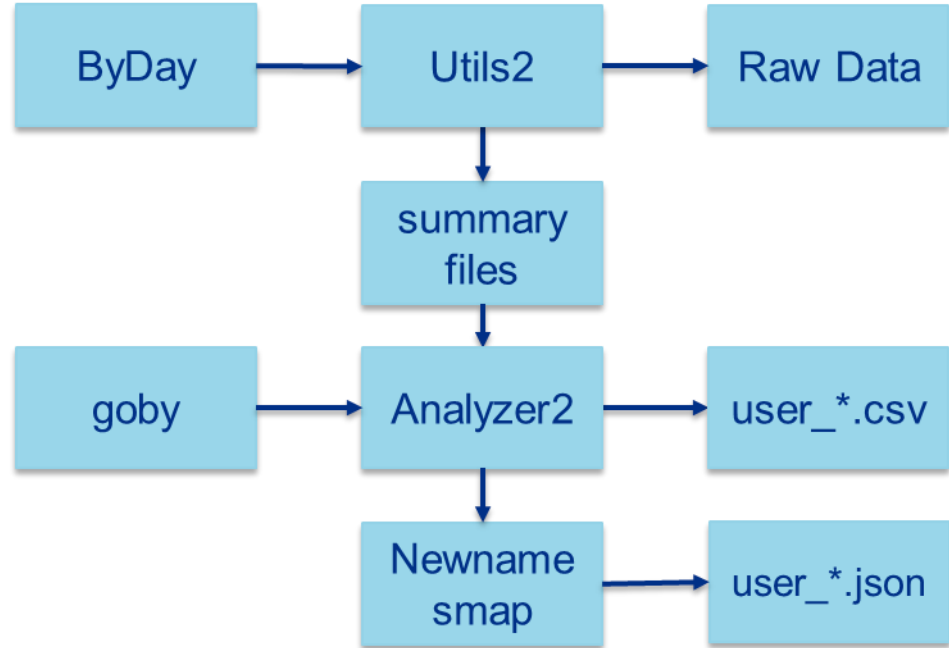
Architecture Diagram



Architecture Diagram (reads from left to right) showing how data gets to the React App.

Xroot data fetching

- ByDay → Utils2
 - Queries projects for each day, separated by project ID
 - Makes summaries for each day
 - Uses SAM file states and time stamps to determine speed/duration of streaming
- Goby → Analyzer2
 - Makes graphs and user csv files per day
- Newnamesmap/es_client
 - Reformats csv to json, for compatibility with NodeJS backend
 - Concatenates daily user json files



JSON example

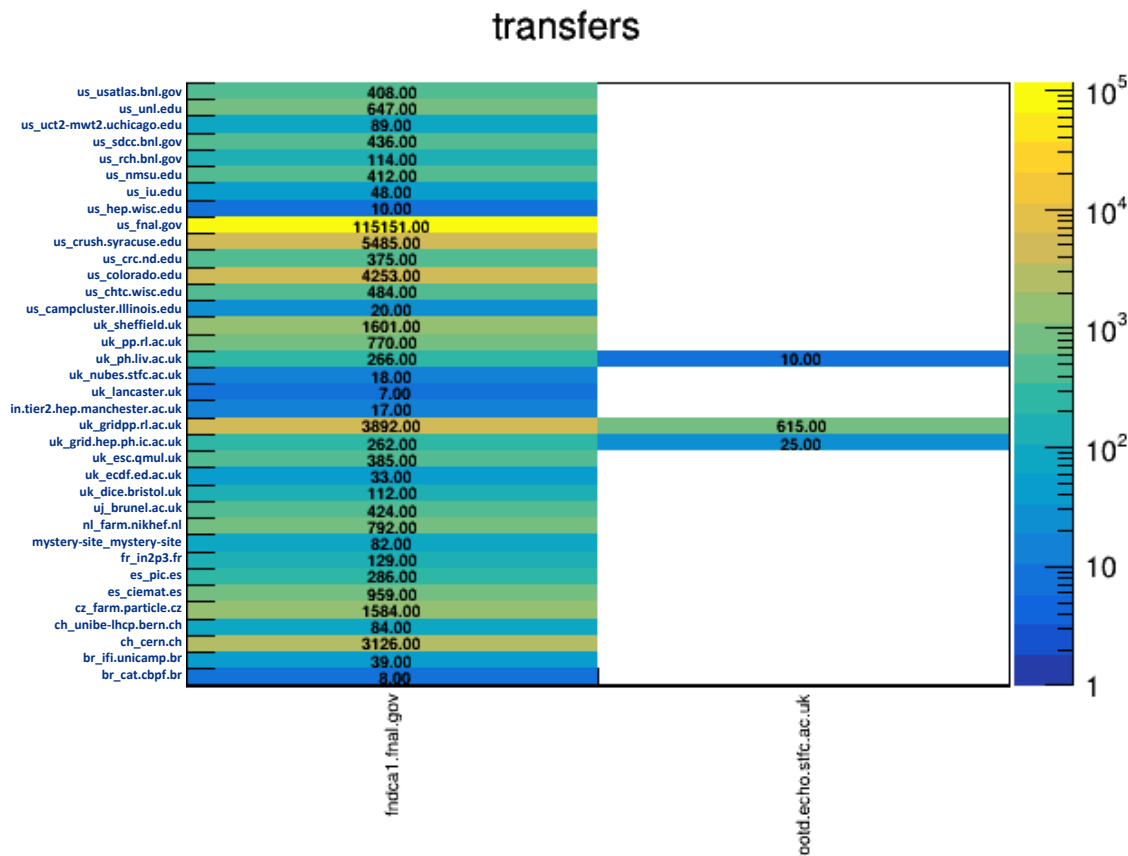
```
{
  "data": [
    {
      "Name": "np04_raw_run005786_0008_dl9.root",
      "Source": "fndcal.fnal.gov",
      "Destination": "us_fnal.gov",
      "File_size": "8243280609",
      "Start_time": "2021-07-02T06:47:09.453Z",
      "File_transfer_time": "63823.24799990654",
      "transfer_speed(MB/s)": "0.1291579615160305",
      "transfer_speed(B/s)": 129157
    },
    {
      "Name": "np04_raw_run005786_0031_dl7.root",
      "Source": "fndcal.fnal.gov",
      "Destination": "us_fnal.gov",
      "File_size": "8228122221",
      "Start_time": "2021-07-02T06:47:20.895Z",
      "File_transfer_time": "134162.20600008965",
      "transfer_speed(MB/s)": "0.06132965807818114",
      "transfer_speed(B/s)": 61329
    }
  ]
}
```

Output from search
Fed to React App

Transfers

Spanning April, number of transferred files between sites (coming from column and going to row)

Most transfers are from Fermilab to Fermilab



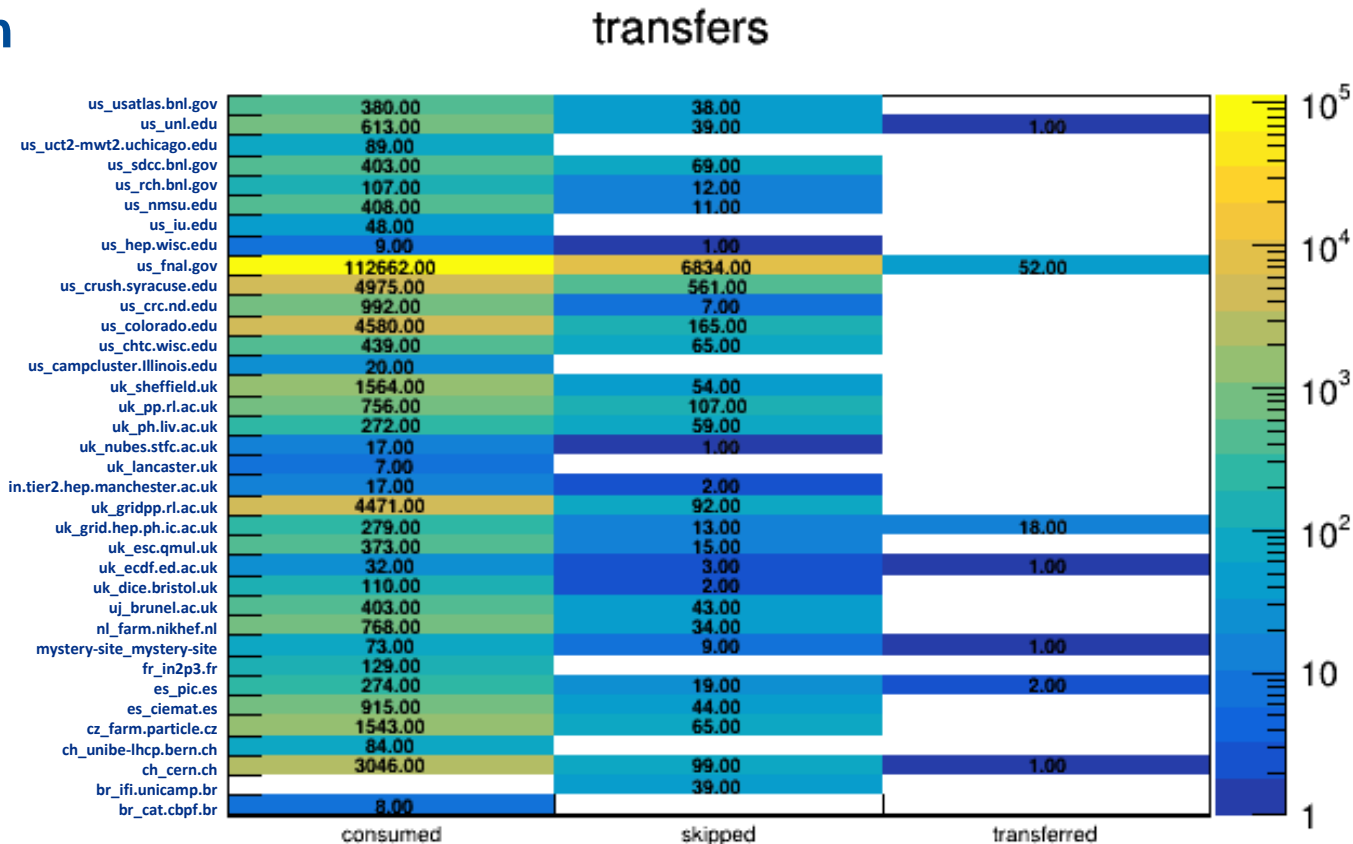
Transfers Graph

Number of transfers, separated by file status

Consumed-closed with successful state

Skipped- closed with unsuccessful state

Transferred-copied to local disc, but not closed

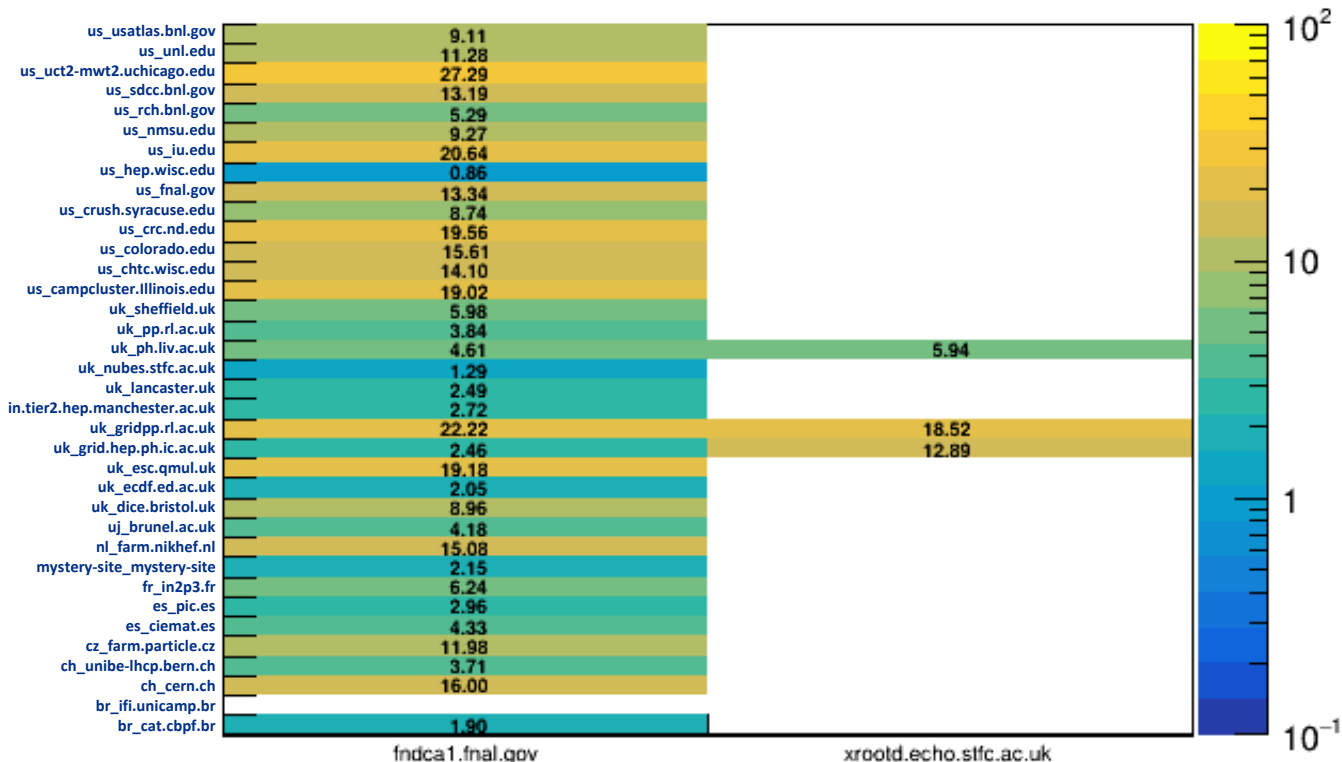


Rates for consumed

rate for consumed, MB/s user_2021-04-01_2021-04-30

Spanning April,
streaming rate (MB/s)
for consumed files
between sites
(coming from column
and going to row)

Most transfers are
from Fermilab



Summary and Conclusions

- Data from SAM and ElasticSearch give us useful information about DUNE Computing
 - I/O streaming rate within each application
- Streaming from the nearest data copy makes Analysis tasks faster
 - Less important for reconstruction since the jobs are CPU limited

Future work

- Analysis and Modeling
 - Find bottlenecks
 - Predict network traffic, expected bandwidth
 - Predict required storage capacity
 - Compare IOPS (input output per second) between pairs of sites
 - Best job distribution given N sites with a given bandwidth
- XrootParser
 - Automate goby/Analyzer2
 - Merge XrootParser Branches

Acknowledgments

Many thanks to

- Steven Timm
- Heidi Schellman,
- Zac Lee, Sean Gilligan, Lydia Brynmoor, and Luke Penner
- Aisha Ibrahim, Andres Quintero Parra, and Brian Vaughn
- SIST Committee