



Monitoring DUNE Data Streaming Efficiency

Lisa Paton

Supervisor: Steven Timm

SIST 5 Minutes, 5 Slides

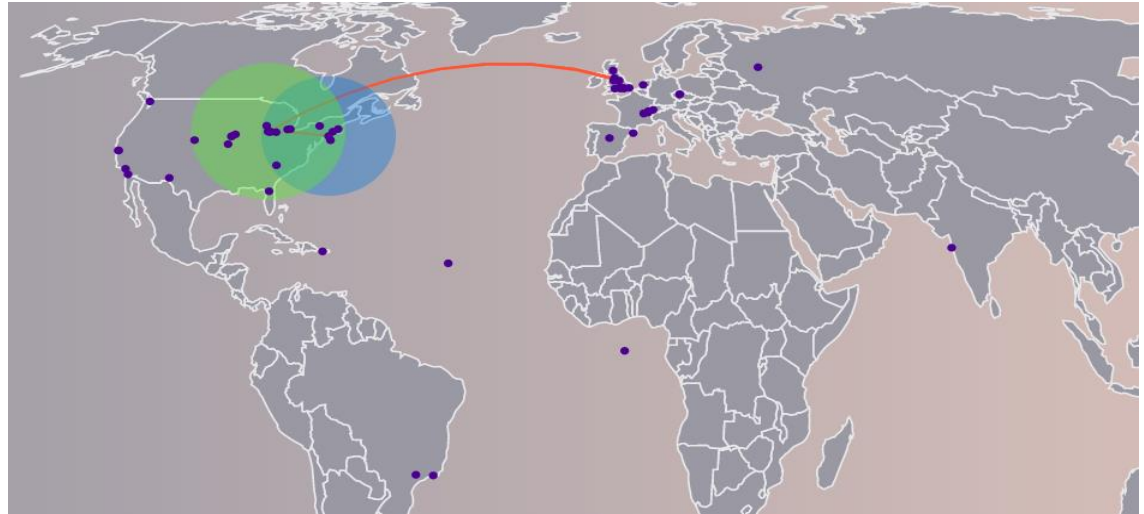
16 June 2021

DUNE Computing Sites

- The Deep Underground Neutrino Experiment (DUNE)
 - International Collaboration
 - 36 Computing Sites
 - 15 Storage sites (already 14 Petabytes of data)
- Any job running at the sites without data storage has to **stream** that data across the network (30,000 jobs at a time expected)
- Need to measure **who** is running, **where** are they reading data from, **how much** data, and **how fast**

Transfer Map

- React App
 - Online front end
 - Access transfer data based on user input date ranges
 - Data collected from Rucio
 - Visualize data transfer sizes, paths, etc.
 - Bigger circle \Rightarrow more data
 - Made by 3 Oregon State University students



Legend:



Data Sent



Data Received



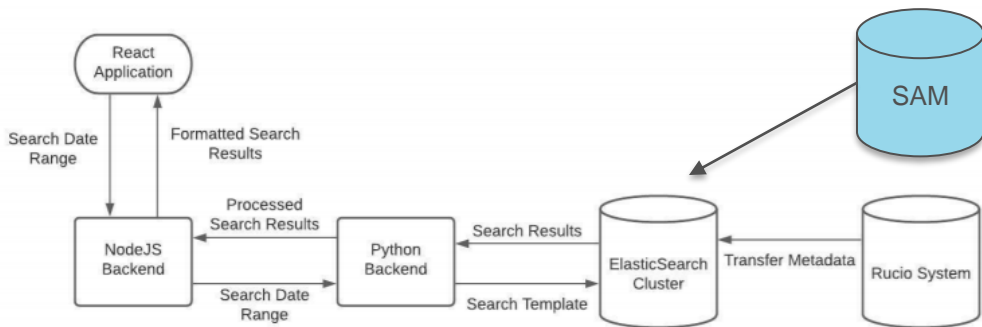
Dune Institution



Transfer Path

Objectives

- Collect streaming data
- Update Rucio transfer data visualization
- Rework Python back end to include Serial Access Metadata (SAM)
- SAM records start and end times of streaming, also file size
- Can calculate transfer rate (file size/duration)
- SAM records stored in elastic search
- Incorporate frequent updates to searchable date ranges



Objectives Continued

- Analyze data
 - Compare relative speed of transfers between pairs of sites
 - benchmarking, input/output operations per second (IOPS), network traffic, etc.
 - If the transfer rates are slow, why?
 - Slow source disk
 - Slow network
 - Slow local disk
 - Asking for data slowly (supposed to be slow)
- Model data
 - Predict network and storage needs
 - Storage capacity (measured using IOPS)
 - Network traffic (measured using Mb/s)