Understanding LHC data movement with NetSage

Doug Southworth (dojosout@iu.edu)

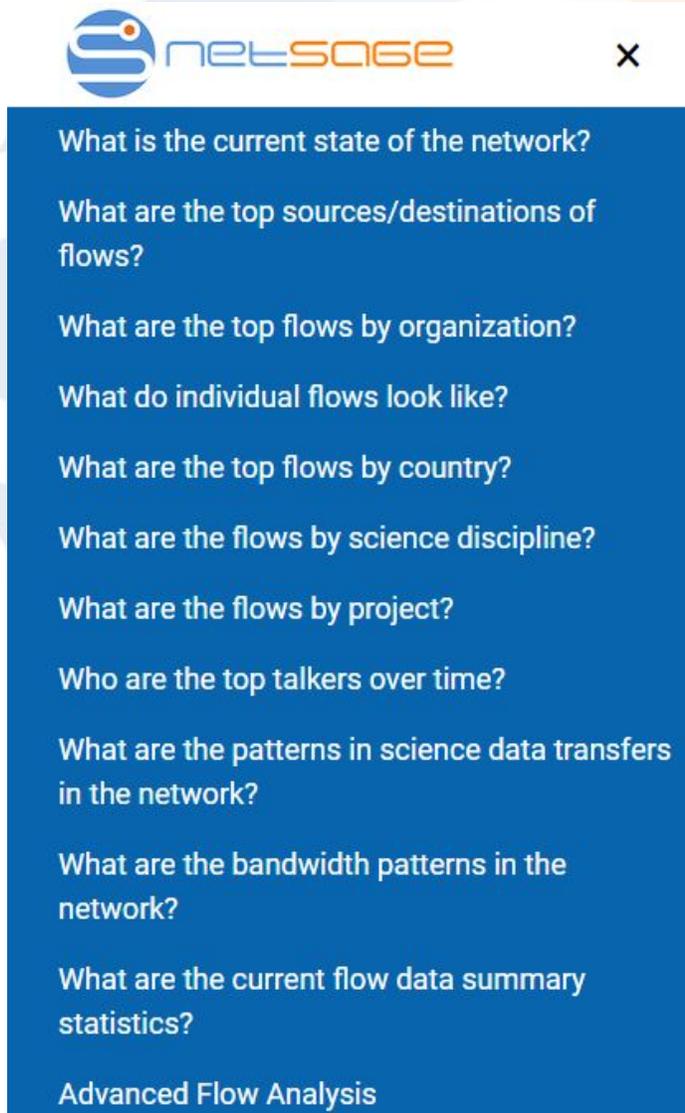Indiana University

October 12, 2021

Quick overview of NetSage

- NetSage advanced measurement services for R&E data traffic
  - Better understanding of current traffic patterns across instrumented circuits
  - Better understanding of large flow sources/sinks
  - Performance information for data transfers
- Collaboration between Indiana University, LBNL, and University of Hawaii Manoa
- Originally funded by the NSF for IRNC circuits, NetSage is now deployed in a variety of configurations across the globe

Built around answering questions

What is the current state of the network?

What are the top sources/destinations of flows?

What are the top flows by organization?

What do individual flows look like?

What are the top flows by country?

What are the flows by science discipline?

What are the flows by project?

Who are the top talkers over time?

What are the patterns in science data transfers in the network?

What are the bandwidth patterns in the network?

What are the current flow data summary statistics?

Advanced Flow Analysis

- Answers questions asked by network engineers, network owners, and end-users
- Human-readable summaries and patterns
- Big picture overview helps highlight trends and events that can make in-depth analysis of local data more fruitful

NetSage Ingest

Data Ingest Pipeline
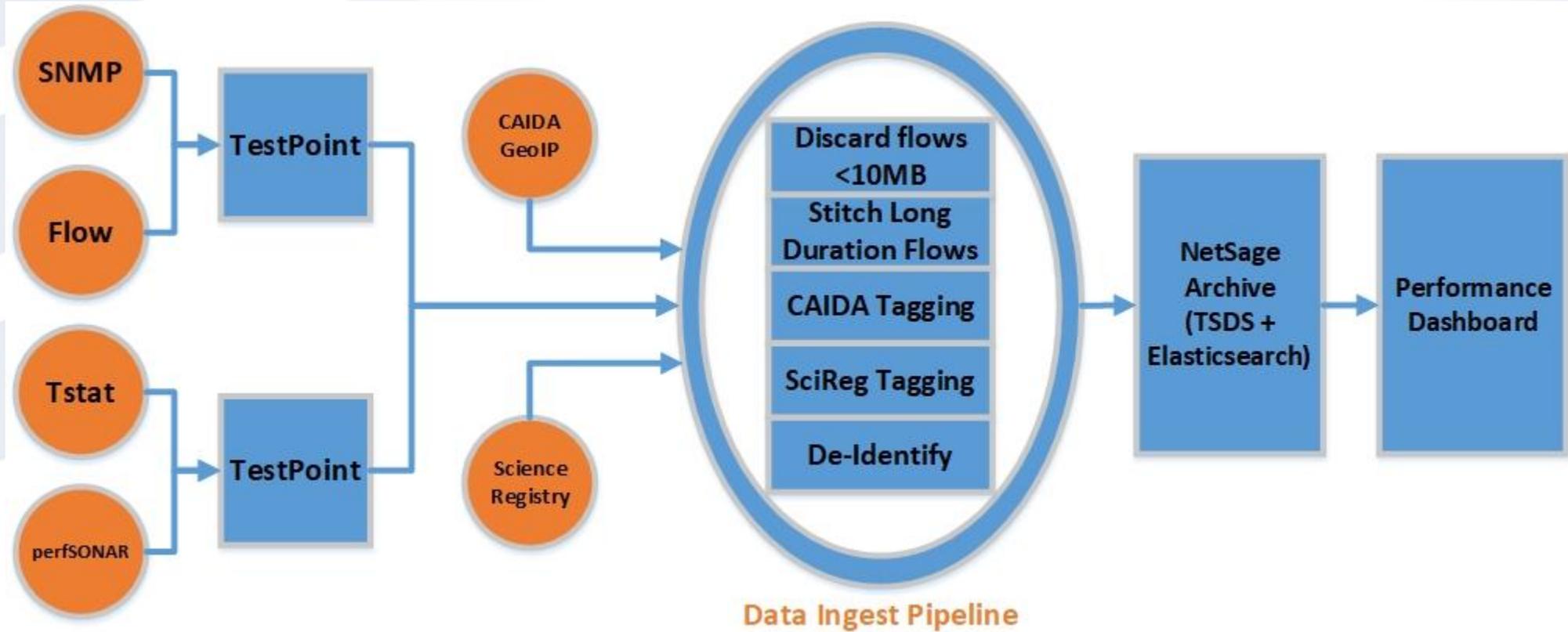
NetSage Privacy

- NetSage is committed to privacy, and preemptively addressing any security or data sharing concerns
  - No PII collected
  - Remove the last octet from IP address
  - Only keep data on flows over 10M for circuits
    - 1M for archives
- Fully GDPR Compliant
- Data Privacy Policy
  - http://www.netsage.global/home/netsage-privacy-policy
- Data Flow Data Retention (De-Identification) Policy
  - https://tinyurl.com/netsage-deid

What NetSage has access to currently:

- sflow:
  - NEA3R (NY-London, NY-Amsterdam)
  - GEANT (NY-Paris)
  - TransPAC (Seattle-Tokyo, Guam-Singapore)
  - SingAREN
  - Hawaii-Guam, Hawaii-LA
  - PIREN and PNWGP
- SNMP
  - Mostly the same as above, plus ANA links

What NetSage needs moving forward:

- Feedback
  - What questions do you want answers to?
  - What's the most useful way of displaying data?
  - What features are lacking? Which ones aren't needed?

- Data
  - Flow, SNMP, perfSONAR, data archives - NetSage can ingest whatever you have

What NetSage needs now:

- IP mappings
  - Instruments
  - DTNs
  - Compute resources

This data is used by the Science Registry for flow identification.

How can I get data into NetSage?

- Send flow/SNMP data directly to NetSage
  - GlobalNOC handles data ingest
  - Records stored securely by OmniSOC
- Run your own ingest pipeline
  - Full Docker deployment of ingest pipeline on-site in about one hour

Regardless of method, NetSage never has access to the full IP address and remains privacy compliant!

Upcoming Development

- Support for flow tagging
  - IPv6 header tagging
  - IPv4 UDP firefly packets
- LHC traffic toggle to view R&E links with LHC traffic removed
- Proposal for use in XSEDE supercomputing facilities

Existing NetSage Examples

- https://netsage.global
  - International circuits
  - Regional R&E networks
  - Datacenter archives
- Look and feel is similar, but Dashboards are added/removed depending on use case

LHC Dashboard Prototype

https://lhc.netsage.global

- username: netsage

- password: netsage-lhc

Feel free to email with comments/questions/suggestions. This is a springboard for future development, so your input is valuable and appreciated.

Doug Southworth - dojosout@iu.edu