



Using NetSage to Support ACCESS

Jason Zurawski - zurawski@es.net

Lawrence Berkeley National Laboratory / Energy Sciences Network
(ESnet)

2023 Internet2 Technology Exchange
September 20th, 2023



Monitoring using 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
- Originally a NSF funded collaboration between Indiana University, LBNL, and University Hawaii Manoa, now primary development is at the Texas Advanced Computing Center
- ACCESS-CI compute resources: <https://access.netsage.io>



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?

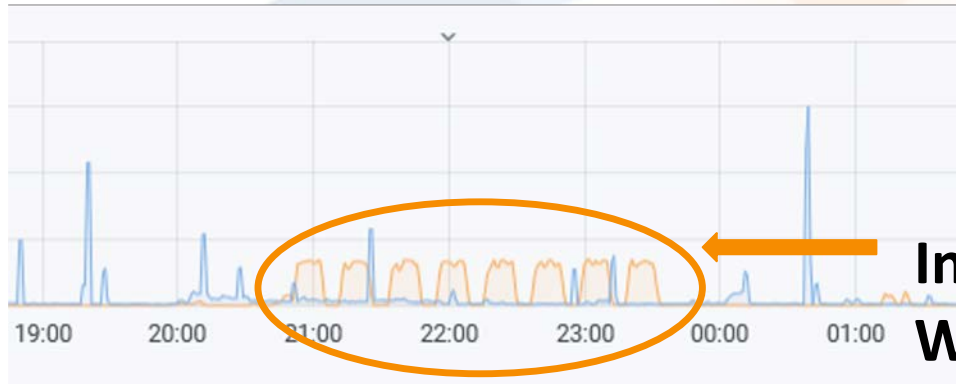
What are the current flow data summary statistics?

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



Built around answering questions



# Flows	Total Volume	Avg Rate
2,791	2.7 TB	4.6 Mb/s

Interesting pattern.
What is it?

Why so slow?

Common Information				
Timestamp	Sensor	5-Tuple Hash	Protocol	Duration
2021-07-15 03:20:53	SingAREN Los Angeles, United States	c9cef62cde44cadab1d152f5e3d37d197f3fa7bf034e74a44b105b1508a3a7f9	tcp	00:00:26

Source Information		Destination Information	
Field	Data	Field	Data
Organization	Nanyang Technological University	Organization	Education Bureau, Kaohsiung City Government, Taiwan
Country	Singapore	Country	Taiwan
ASN	9419	ASN	1659
Subnet	155.69.240.x	Subnet	140.127.254.x
Port	443	Port	55125

Singapore to Taiwan
via LA?



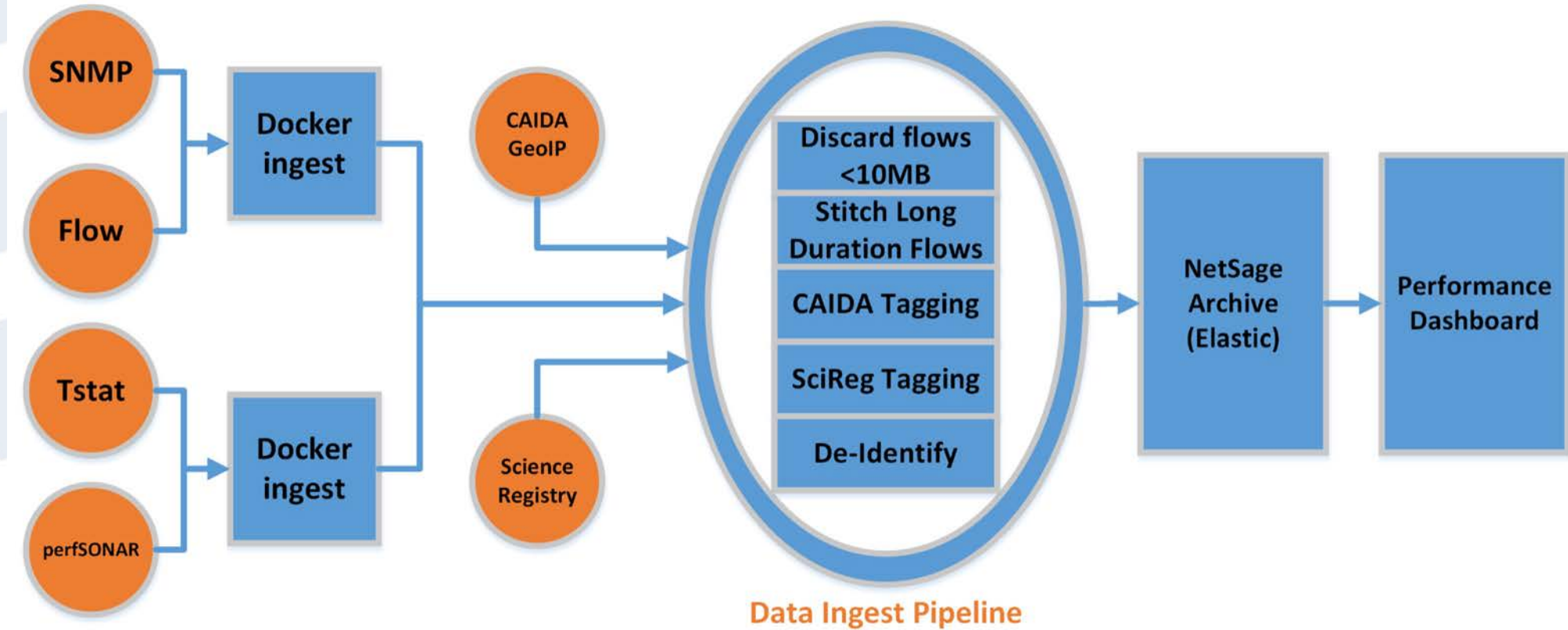
NetSage Focus on Use Cases

- Start with Flow Data Dashboards
 - What are the top sites using my circuits?
 - What are the top sources/destinations for an organization?
 - Who's using my archive?
- Dig deeper with debugging dashboards
 - What are the flows like between these two orgs?
 - There was a performance spike on my circuit – what was it?
 - Who's transferring a lot of data really slowly?

Flow Data collection

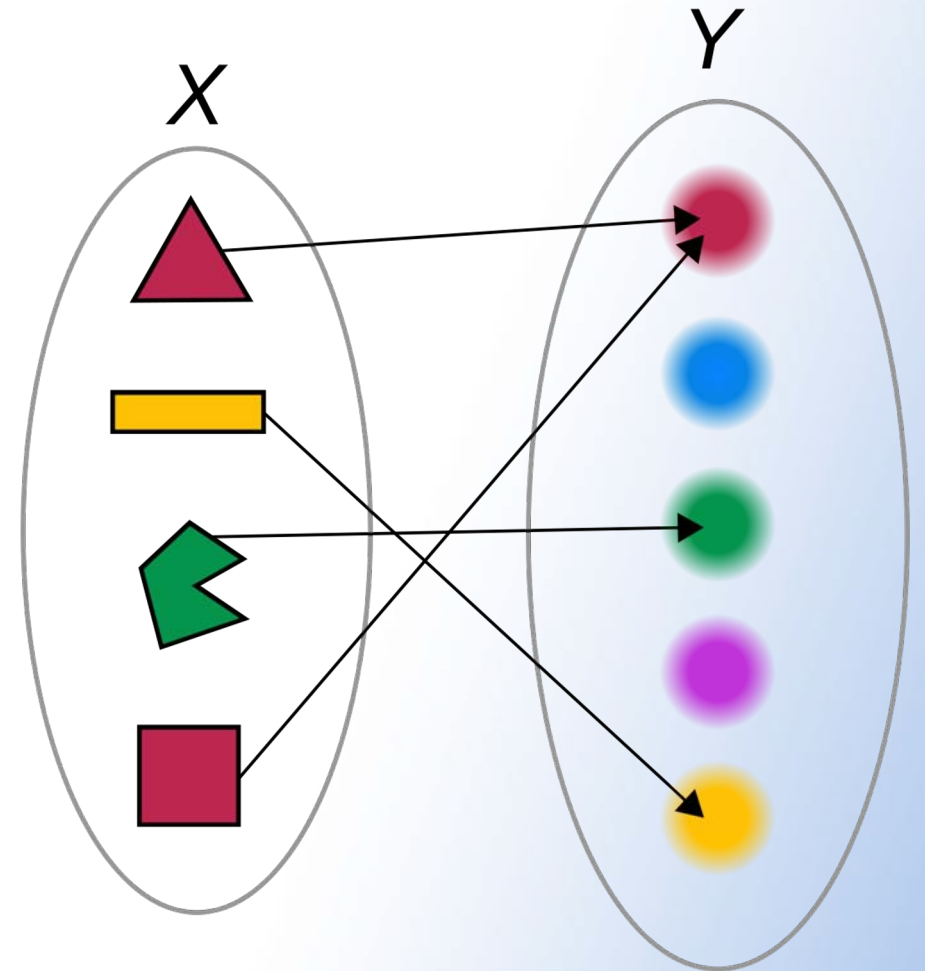
- Flow data is sent to a NetSage Ingest Pipeline, de-identified, and then sent to NetSage archive
- Runs as a set of Docker containers on existing hardware/VM infrastructure
 - Linux distro of your choice underneath
 - Can be anywhere on your network as long as your router can send raw flow data to it
 - Support from TACC if you run into any issues, but you handle general updates

NetSage Ingest



Global Science Registry

- Maps flows to specific resources
 - DTN
 - Instrument
 - Compute
- Contributions come from resource owners
- Currently matched by IP, however packet marking is in the roadmap for development



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 10MB for circuits
 - 1MB for archives
- Data Privacy Policy
 - <https://tinyurl.com/netsage-privacy>
- Prototypes are behind a password until we're told to make it public



NetSage Domestic:



- FRGP
- **GPN**
- PACWAVE
- SoX
- Sun Corridor
- TACC
- LEARN
- ...and now ACCESS-CI!



Netsage for ACCESS-CI

- One set of dashboards for all ACCESS-CI compute resources
 - Currently, TACC and PSC are contributing flow data
 - SDSC, NCSA, and others coming soon
 - Portal available at <https://access.netsage.io>
- Currently just flow data, however there are plans for future development, including:
 - In-band telemetry
 - Tie in flow data with job scheduling/allocation to see the big picture of flow data in, computation, and flow data out



Pick a sensor

Pick a timeframe

Flow Data

This dashboard provides flow data for the top ten sources and destinations by volume and rate. Click on an organization name in one of the tables to see the flow data dashboard for that organization.

Please note that rate will be zero if only one sampled flow was detected.

The slope graph below shows the top pairs of organizations by volume. Darker lines correspond with larger volume. All times are displayed in browser local time.

Links in tables for more specific information

Top Sources

By Volume ▾

Source	Total Vol. ▾	Largest Flow	# Flows
Vanderbilt University	140.9 TB	316.7 GB	153.6 K
University of Alabama at Birmingham	89.4 TB	1.5 TB	21.3 K
Vanderbilt University Medical Center	68.0 TB	182.4 GB	90.4 K
Georgia Institute of Technology	56.0 TB	186.4 GB	59.1 K
University of Georgia	36.8 TB	163.0 GB	39.1 K
Kennesaw State University	33.6 TB	91.1 GB	57.4 K
National Library of Medicine	32.3 TB	4.7 TB	23.3 K



What NetSage Does Best

- Answers questions asked by network engineers and network owners
- Human-readable summaries and patterns
- Gives people the higher level pattern so they can narrow down a time frame and then use local tools that have more detail
- Questions? Contact Jen Schopf & Doug Southworth (jms@tacc.utexas.edu & dsouthworth@tacc.utexas.edu)





Using NetSage to Support ACCESS

Jason Zurawski - zurawski@es.net

Lawrence Berkeley National Laboratory / Energy Sciences Network
(ESnet)

2023 Internet2 Technology Exchange
September 20th, 2023

