U.S. DEPARTMENT of ENERGY

BERKELEY LAB

ESnet
ENERGY SCIENCES NETWORK

Networks For AI

Eli Dart
Network Engineer, Science Engagement
dart@es.net

SIG-NGN Meeting
Karlsruhe, Germany
9 October 2025

Context

- Huge focus on AI currently
  - Lots of funding
  - Meetings like this one!
- AI is data-driven
  - Science community has ground truth data
  - Lots of interest in feeding this data to AI
- “Move the compute to the data!”
  - Of course we do this when we can
  - In many cases, we can’t
    - AI training resources and data not colocated
    - Computing can’t be moved
    - **Need to move the data to the computing**
- So how do we do this? With networks.

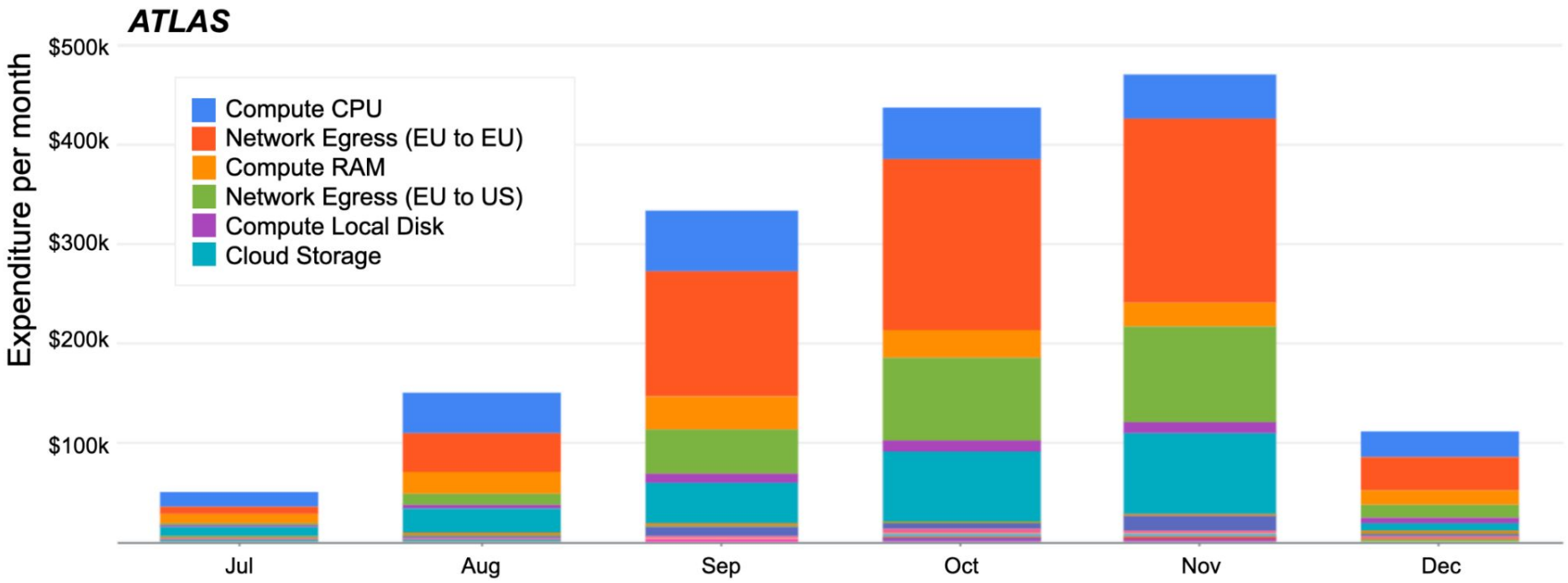
Lessons From LHC: Networks and Storage

- Original MONARC model assumed unreliable networks
  - Hierarchical data replication
  - Storage-intensive
- Networks are reliable, storage is expensive
  - Change in computing model
  - Use remote I/O instead of copying data
  - Burn network to save on storage
- This model has persisted because it works
  - Networking is now a core component of LHC computing model
  - This is unlikely to change in the near term
- Storage is the hard part
  - This will be true until something changes
  - Waiting for disruptive technology…

Lessons From LHC: Storage and Cloud

- ATLAS Google trial
  - https://arxiv.org/abs/2405.13695
  - https://doi.org/10.48550/arXiv.2405.13695
- Established as part of an ongoing evaluation of the use of commercial clouds by the ATLAS Collaboration
  - Anticipating potential future adoption of such resources by WLCG
  - Fulfil or complement computing pledges
- Integration of computing was successful
  - Ran for several months
  - Real-world results
- Total Cost Of Ownership calculation is instructive for this discussion

Lessons From LHC: Storage and Cloud



ESnet

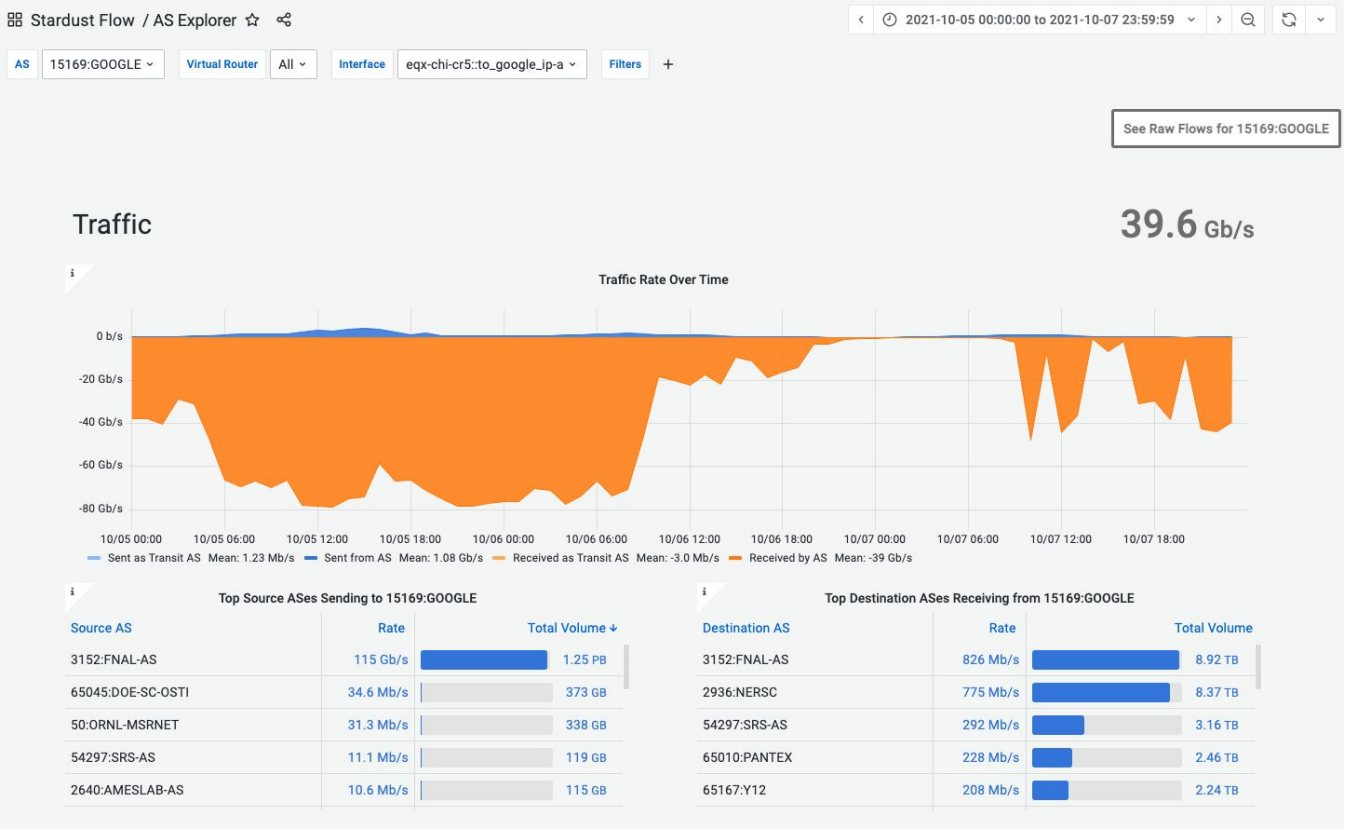
Lessons From LHC: Storage and Cloud

- Primary cloud expenses:
  - Network egress
  - Storage
- This selects heavily against some Cloud workflows
  - Data storage in commercial cloud
  - Serving data from commercial cloud
  - These are not viable unless cost models change
- See previous LHC lesson: burn network to save storage

Training and Inference in Commercial Cloud

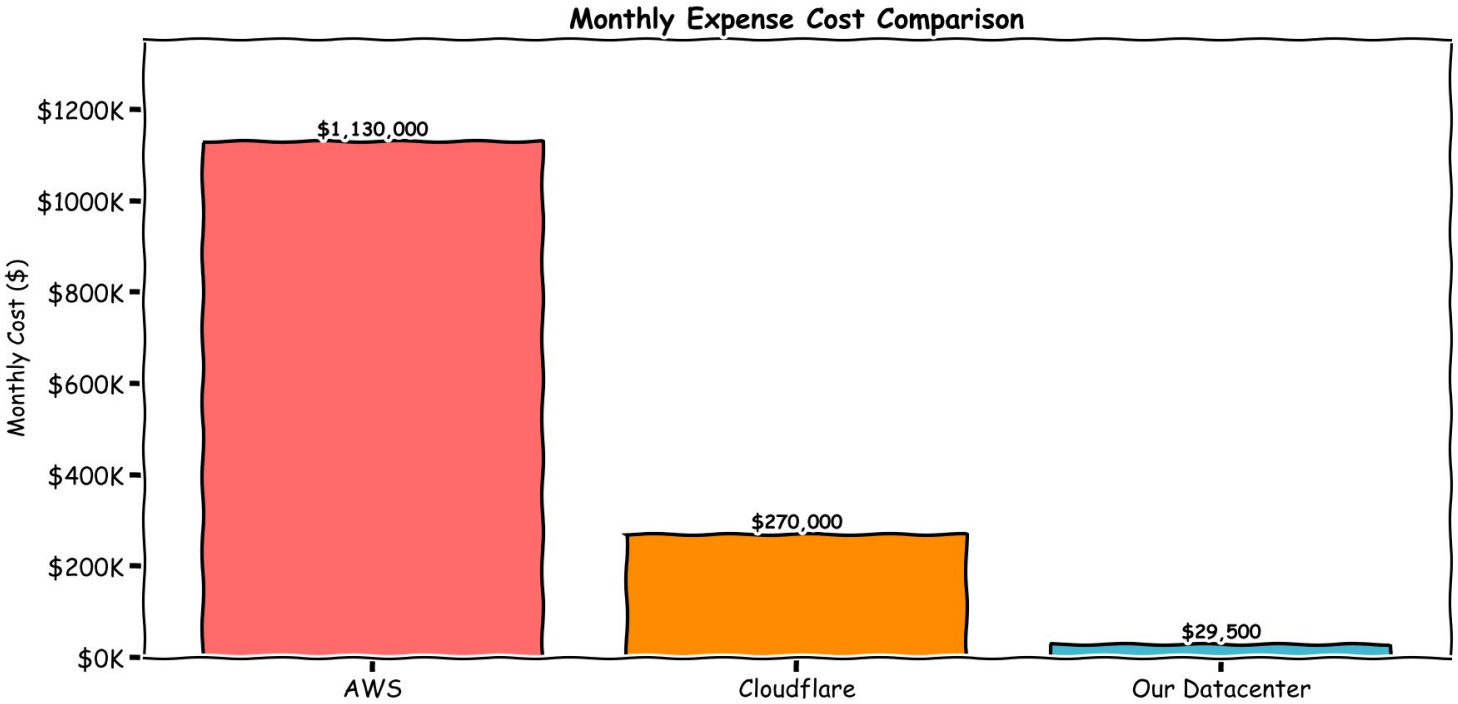
- DUNE collaboration
- ML based data reduction
- Near-real-time inferencing classifiers supporting neutrino physics workflows.
- Hybrid workflow
  - Data and CPU computing at FNAL
  - GPUs in Google
- Impossible without high performance networks
- Works great if you do have high performance networks

Data From Lab to Cloud For Inference



- ~80Gbps for hours at a time
- Over 1PB moved in ~30 hours
- No need to replicate the data first (no file transfer workflow)

Random Startup Story



- Petabyte scale storage is cost prohibitive in commercial cloud
- File-based replication of data will be constrained by storage costs

ESnet

Random Startup Story

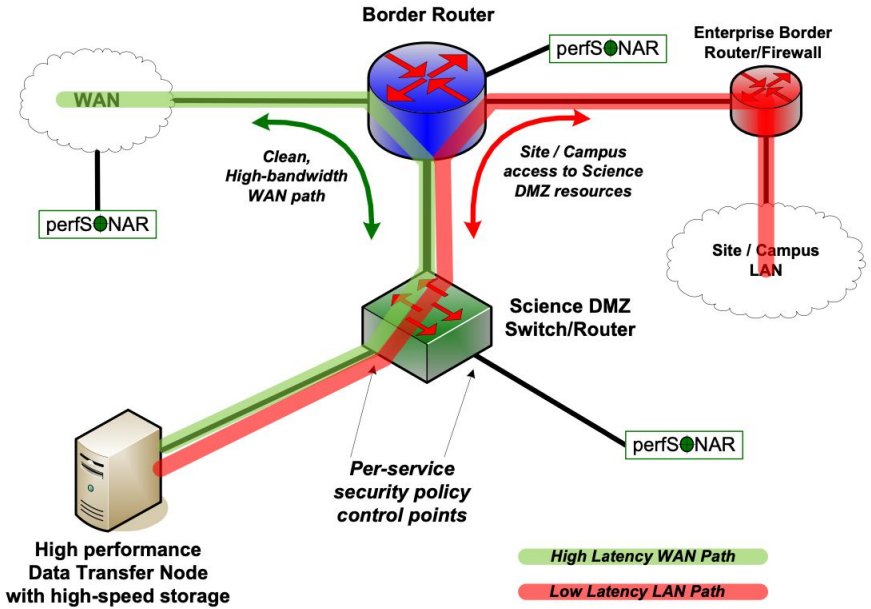
- Startup company trying to do AI/ML with video data
- https://si.inc/posts/the-heap/
- Orders of magnitude cost difference between cloud and self-hosted storage
  - Cloud is enough more expensive that self-hosted is a no-brainer
  - But storage is still difficult and expensive
  - If storage were easy we would all be doing more of it
- Not just the science community
  - Everyone is affected by the high cost of cloud storage
  - It's not just us
- As a community, how do we want to deal with this?

Putting A Picture Together

- Networks are essential for AI workflows
  - Move data to compute
  - Integrate multiple resources
- But data is the point
  - Data server interface to the network is key
  - Make sure this is included in the AI strategy
- We have a set of best practices

#include <science_dmz.h>

- I’m not going to give a Science DMZ talk here
  - You’ve probably all heard it anyway
- The key point is to include end systems in network design
  - DTNs are the WAN-facing interface to storage
  - Data APIs run on DTNs



ESnet

Path Forward: AI Needs Data and Networks

- Lots of money being directed toward AI
- If these efforts are to be successful, need good interfaces to networks
  - Data repositories
  - Data portals
  - Computing systems
- It is on us to ensure that the science community can use networks effectively
  - File access
  - Remote I/O
  - Streaming access
  - Remember that remote file read → stream → analysis is a real workflow
- Data challenge activities will be valuable
- Science DMZ model will be valuable

U.S. DEPARTMENT of ENERGY

BERKELEY LAB

ESnet
ENERGY SCIENCES NETWORK

Thanks!

Eli Dart

dart@es.net

https://my.es.net/

https://www.es.net/

https://fasterdata.es.net/