CICompass

Funded by the National Science Foundation, Grant #2127548

Migrating two Major Facilities from on-premise data centers to a common cloud platform mid-merger

By Chad Trabant

The March Toward the Clouds: MF Perspectives

March 2, 2022
CI Compass Cyberinfrastructure for NSF Major Facilities Workshop

USC Viterbi School of Engineering Information Sciences Institute
INDIANA UNIVERSITY
UNIVERSITY OF NOTRE DAME
THE UNIVERSITY OF UTAH
TEXAS TECH UNIVERSITY
renci

Funded by the National Science Foundation awards
EAR-1851048, EAR-1851037, EAR-1851159, EAR-1851163

Migrating two Major Facilities from on-premise data centers to a common cloud platform mid-merger

Chad Trabant, Incorporated Research Institutions for Seismology

GAGE Geodetic Facility for the Advancement of Geoscience

UNAVCO

SAGE Seismological Facility for the Advancement of Geoscience

IRIS

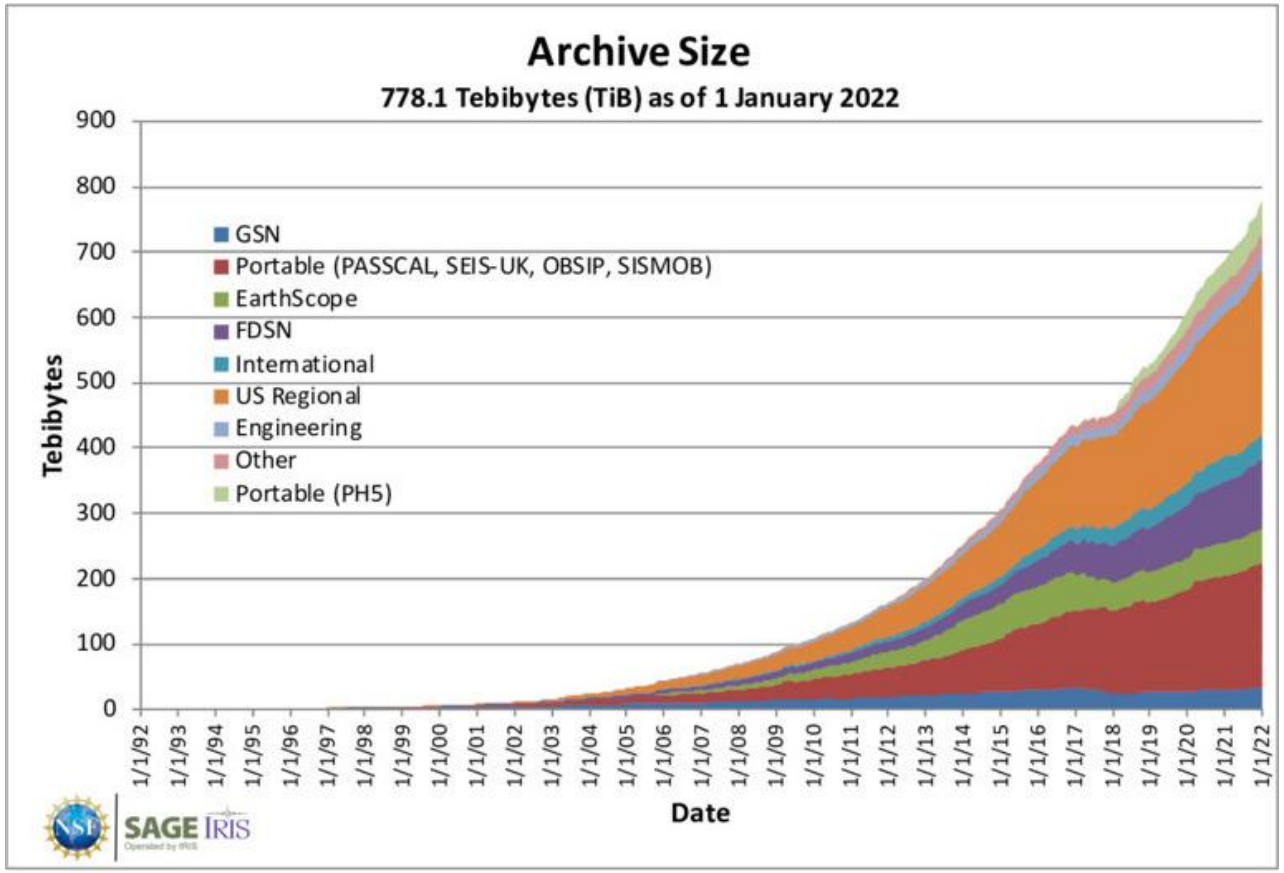
NSF awards EAR-1851159 (GAGE-EAR), EAR-1851163 (GAGE-OPP)

NSF awards EAR-1851048 (SAGE-EAR), EAR-1851037 (SAGE-OPP)

Funded by the National Science Foundation awards
EAR-1851048, EAR-1851037, EAR-1851159, EAR-1851163

Who: the current facilities

- Primary data facilities supporting **seismological** and **geodetic** research in the US as well as many international researchers
- Dependent on on-premise systems
- Operating for nearly 3 decades
- Petabytes of data, 10s of thousands of stations, 100s of thousands of data streams



Growth of the primary archive of the SAGE facility operated by the IRIS Data Management Center. **All repositories across both facilities are expected to total 2-3 petabytes.**

Funded by the National Science Foundation awards
EAR-1851048, EAR-1851037, EAR-1851159, EAR-1851163

Why 1/2: motivations for facility operation

- **Cost effective operation**
- **Highly scalable operation**
- **Generalized systems - ready for other data types**
- **Opportunity to reduce legacy technical debt**
- **NSF requested a cloud-based prototype**

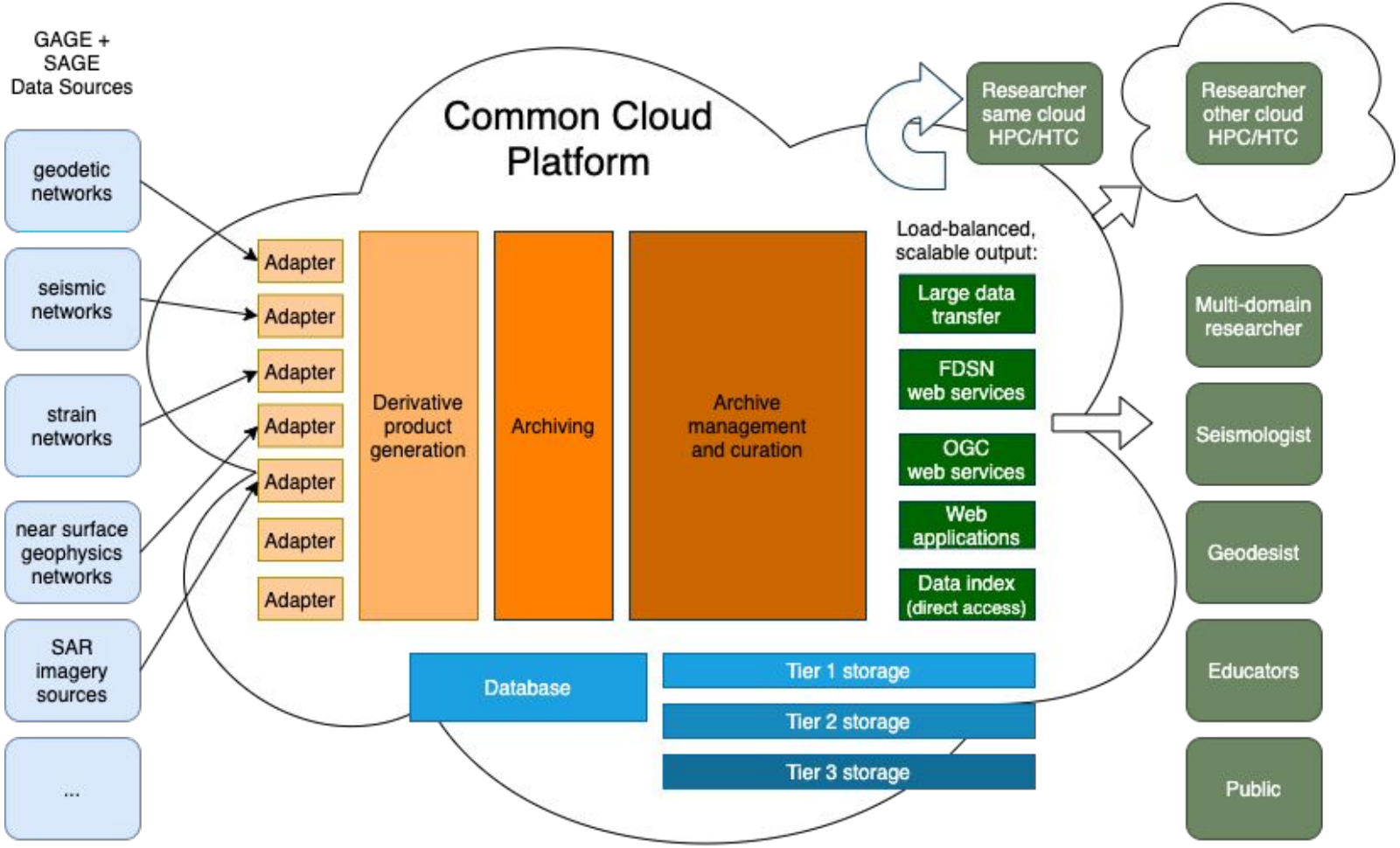
Funded by the National Science Foundation awards
EAR-1851048, EAR-1851037, EAR-1851159, EAR-1851163

Why 2/2: anticipated impacts on researchers

- **Data processing in the same or data proximate cloud environment**
- **Integrated geophysical data discovery**
- **Significantly enhanced ability to transfer large volumes of data**
- **More capacity for current access mechanisms, easy transition to enhanced services**

Funded by the National Science Foundation awards
EAR-1851048, EAR-1851037, EAR-1851159, EAR-1851163

What: a common geophysical data platform



Funded by the National Science Foundation awards
EAR-1851048, EAR-1851037, EAR-1851159, EAR-1851163

Primary current challenges

- **Adapting data management systems for the cloud**
- **Adapting to generalized components, de-siloing**
- **Understanding and mitigating new hotspots, e.g.:**
  - **Data egress costs and options, projected egress: 1-2 petabytes annually**
  - **Continuous and rapid cost management (FinOps)**
- **Workforce development, re-alignment for the cloud**
- **Research community training and support for new capabilities**

Concurrent with a corporate merger and while operating the on-premise systems!

We are grateful for the inclusion of CICompass expertise in our working groups!