Theme 2: Deep Dive
HECC Project @ NAS

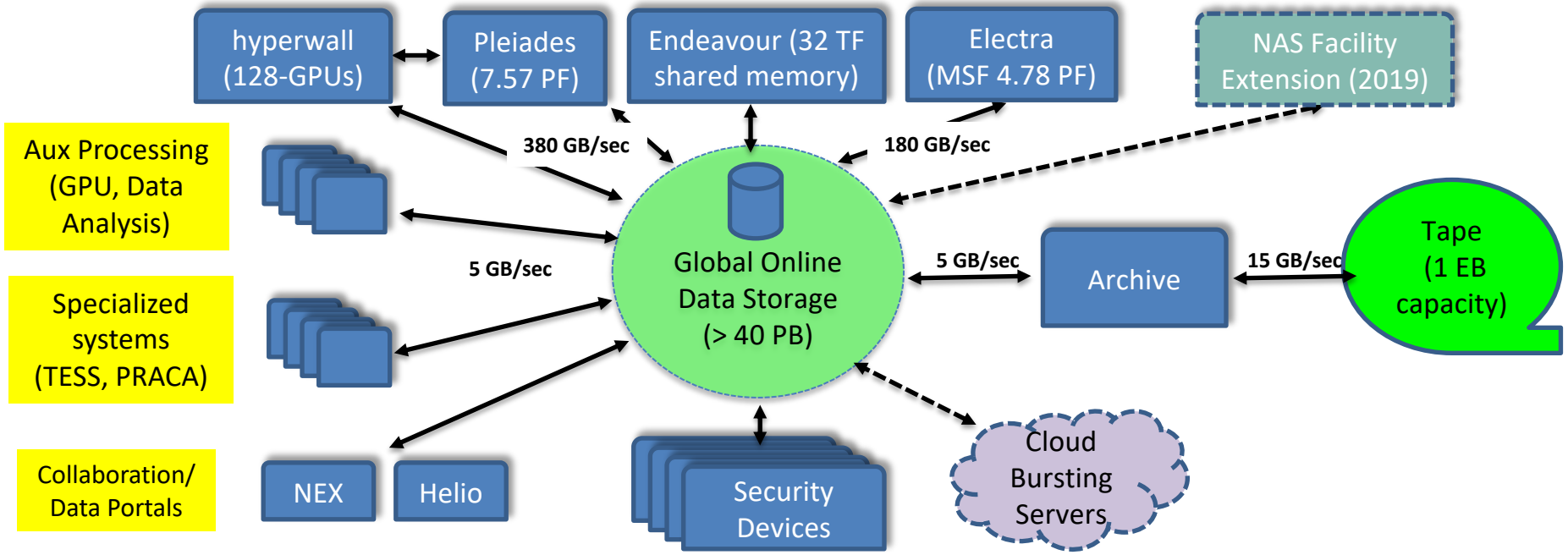
Piyush Mehrotra
Piyush.Mehrotra@nasa.gov

Chief, NASA Advanced Supercomputing (NAS) Division
Ames Research Center

HECC Environment

Mission Statement

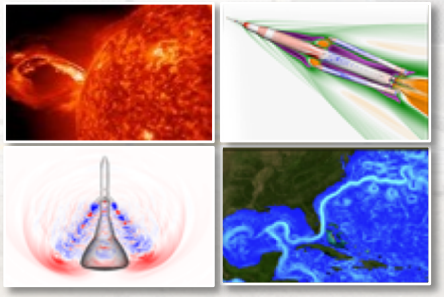
To accelerate and enhance NASA's mission of space exploration, scientific discovery, and aeronautics research by continually ensuring optimal use of a productive high-end computing environment .



HECC Services

HECC provides a suite of complimentary services to the user community to enhance the scientific and engineering results obtained from the hardware assets.

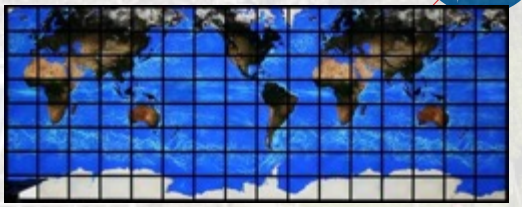
- Systems – Customized solutions including compute and storage solutions to meet specific project or mission requirements.
- Network – End-to-end network performance enhancements for user communities
- Application Performance and Productivity – Software solutions provided to research/engineering teams to better exploit installed systems.
- Visualization and Data Analysis – Custom visualization during traditional post-processing or concurrent during simulation to understand complex interactions of data.
- Data Analytics/Machine Learning – Exploitation of data sets through neural nets and emerging new techniques.
- Data Gateways – Custom data portals to support diverse projects
- Customized system support



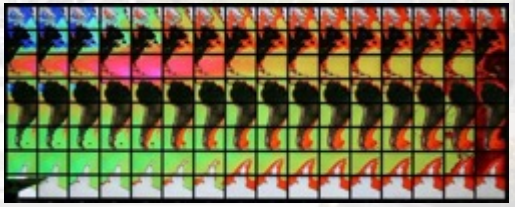
Interactive Analysis of Petascale Ocean Data

- MITgcm runs, up to 70,000 cores, have produced > 5PB dataset
- Customized visualization framework and software to interactively analyze/explore the data
- Supports 2600 different layouts to address different questions
- Utilizes 2TB NVMESSDs (1 per hyperwall node)
  - Filesystem access at near-memory speeds
  - No latency penalties for seeks, unlike spinning disks
  - In-house developed software layer exposes NVMe devices across hosts as unified block devices accessible over network
  - Yields scalable aggregate performance allowing browsing through entire dataset cutting across the storage grain at interactive speeds.

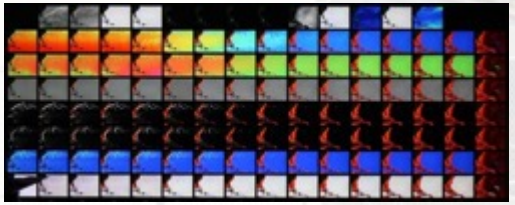
Global View



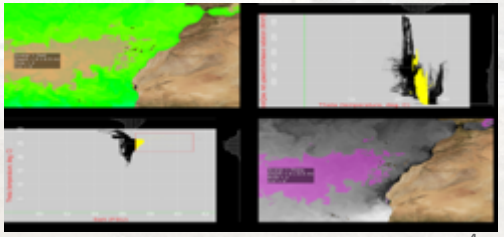
Different Depths



All Scalars



Liked Scatterplot and map



NASA Earth Exchange (NEX)

A Collaborative environment to engage and enable Earth scientists

Collaboration Portal and Knowledgebase
- Web server
- Database server
- 503 registered members (both NASA and non-Nasa)

Compute Resources

Sandbox: for prototyping – available to all

HPC: 720-core dedicated queue + access to rest of HECC systems

Data Repository
NFS storage - 2.3 PBs
- Cached Datasets: LANDSAT, Modis, TRMM, …
- Available internally and externally

Lustre storage – 3.2 PBs available internally

OpenNEX (Cloud Infrastructure)
- 50+ TBs data
- Images
- Tutorials
- Workshops

Heliophysics Portal
Querying Integrated Database of Solar Flares

- Heliophysics Portal provides highlights of the latest solar events
- Multi-instrument database provides an integrated view of reported flares and ground-based observations
- Web interface to search for unique flare events based on their physical characteristics and other pre-defined criteria, in order to investigate their radiation properties, including extreme ultraviolet radiation and X-ray radiation.
- Data from three primary flare lists (NOAA, NASA, and Lockheed Martin) and event catalogs from spacecraft and ground-based observations are integrated into the database for data starting from 2002
- http://helioportal.nas.nasa.gov.

Commercial Cloud Trade Study

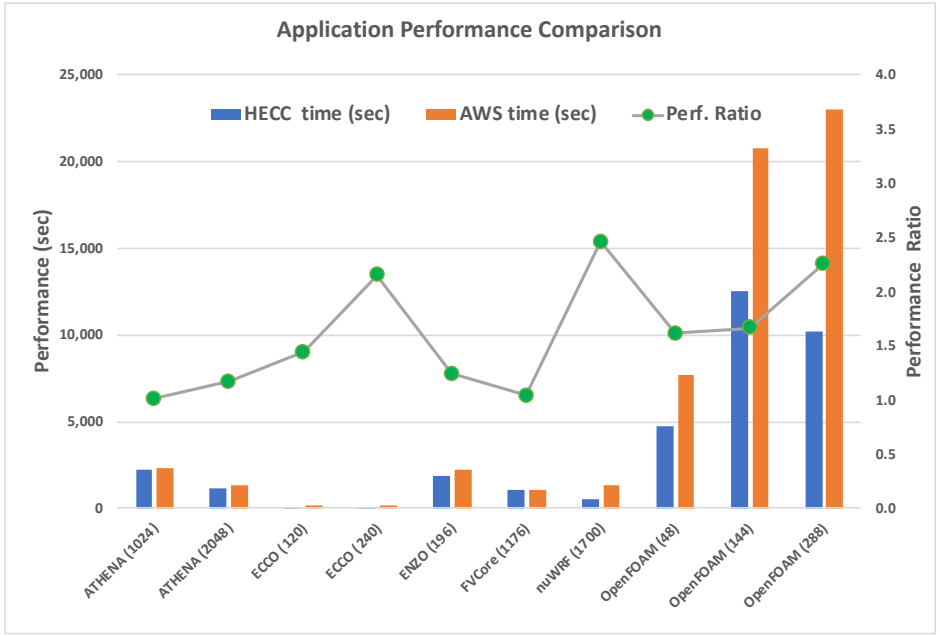
Goal: *Evaluate the suitability of commercial clouds for HPC Applications*

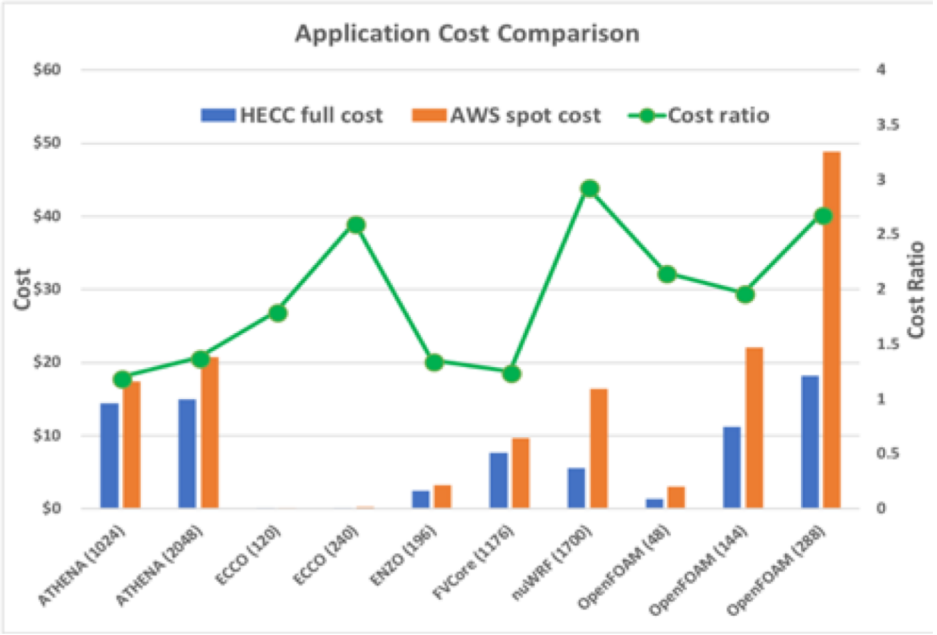
Approach

- **Workload:** NPBs, six full-sized applications (ATHENA++, ECCO, ENZO, FVCore, WRF, OpenFOAM)
- **Systems:** HECC systems Pleiades and Electra, Amazon Web Services (AWS), Penguin-on-Demand (POD)
- **Cost Basis:** HECC – NAS full cost of running (HW/SW, power, maintenance, staff, and facility costs); AWS and POD – only the compute costs from published rates and any publicly-known discounts (spot pricing, lease price, etc.)
- **Key Findings:** *Commercial clouds currently do not offer a viable, cost-effective approach for replacing in-house HPC resources for NASA HPC applications. However, there may be use cases where a commercial cloud is a viable alternatives, e.g.,, specialized hardware*
- **Actions:**
  - Continually evaluate the suitability of commercial clouds
  - Develop an environment to support bursting to commercial clouds (on a full-cost recovery basis) for S&E projects – Phase 1 pilot project available September end.

Evaluating the Suitability of Commercial Clouds for NASA's High Performance Computing Applications: A Trade Study: Chang et al, NAS Technical Report NAS-2018-01*, May 2018*
https://www.nas.nasa.gov/assets/pdf/papers/NAS_Technical_Report_NAS-2018-01.pdf

Perf. & Cost Comparison: HECC - AWS





Deep Dive: NAS HECC Project (1)

How does each division structure their archiving and what is its architecture?

- Archive Systems sit on high performance networks (56 gigabit) as peers to the supercomputing environment.
- Archive Systems have some direct analysis capabilities (analyze in place)
- System level automation moves data from archive disks to two tape copies in a reliable and transparent ways streaming copies in 10's of gigabytes/second

Do archives provide analysis tools? What are they?

- Some amount of limited commercial licensed software provided by HECC: MATLAB, ID, …
- Can install any software stack as needed by PI/user: Commercial, proprietary, open source, ..

How do user expectations drive any of their processes?

- HEC C supports general purpose environments at no cost to users and special purpose (custom) environments based on actual cost

Deep Dive: NAS HECC Project (2)

What requirements do they put on archiving that ensures its viability for how long?

- HECC continually tracks different storage technologies and opportunistically upgrades: when significant reductions in cost or improvements in performance become available
  - Largely driven by LTO tape costs
  - Current practice is to migrate existing data on tape from one generation to the next

How can computing centers help in data processing and data analytics?

- Best-case scenario for NASA scientists is to site copies of datasets next to the large-scale compute so as to increase productivity
- Continually enhance online and archival storage at very low incremental costs

Takeaway

NASA in-house HPC centers provide cost-effective environments for accessing, analyzing and archiving SMD observational and model datasets

- Large-scale computational resources
- Enhanced I/O capabilities
- Large-scale online storage for quick access and archival systems for long-term storage
- Support for visualization and data analytics
- (Will) work with cloud environments to provide hybrid resources
- Capability to provide specialized hardware/software systems for custom data analysis requirements

Comments/Questions?

piyush.mehrotra@nasa.gov

https://www.nas.nasa.gov/hecc