



HPCMP Requirements for Metadata and Archiving at Scale



Amanda Tumminello, Navy DSRC

April 2019

DoD High Performance Computing Modernization Program

Distribution A: Approved for Public Release; distribution is unlimited

HPCMP Ecosystem

Users



DOD HPC

A technology-led, innovation-focused program providing the computational environments to solve the Department's critical mission challenges.

DoD Supercomputing Resource Centers (DSRCs)



Networking and Security

Defense Research & Engineering Network (DREN)



Computer Network Defense, Security R&D, and Security Integration

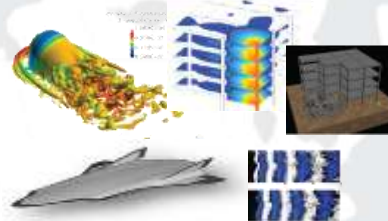


Software Applications



Results

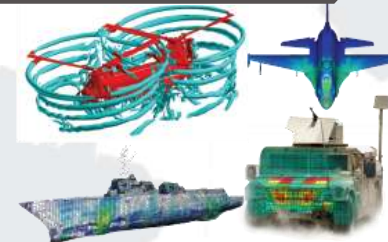
Science and Technology



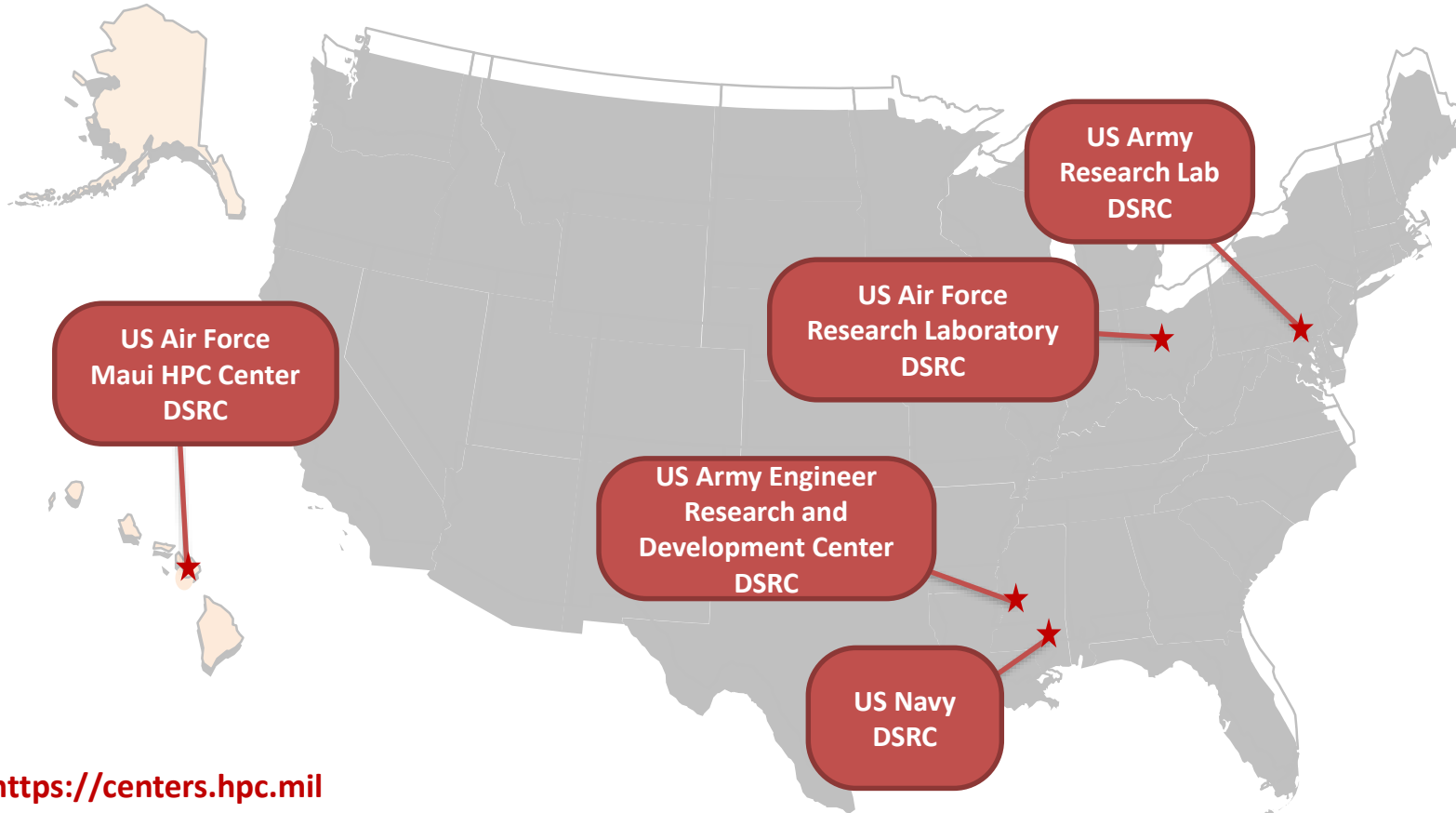
Test and Evaluation



Acquisition Engineering



DoD Supercomputing Resource Centers



<https://centers.hpc.mil>

Users



By the numbers

DoD Supercomputing Resource Centers (DSRCs)



- Five DoD Supercomputing Resource Centers (DSRCs) in four states
- 350 staff
- ~2000 users from 3 DoD Services and additional DoD agencies
- 22 HPC systems from four manufacturers
- 995,896 cores
 - Over 700 GPUs
 - Over 700 accelerators (Phi and KNL)
- 45.6 Petaflops aggregate compute capability
- Over seven billion compute hours delivered annually
- 120 Petabytes of data stored
- 40 Gb interconnect between DSRCs

DoD Supercomputing Resource Centers



- **Four Allocated DSRCs** provide allocated resources to all HPCMP users

- **Air Force Research Laboratory (AFRL) DSRC** *Wright-Patterson Air Force Base, Dayton, OH*
- **Army Research Laboratory (ARL) DSRC** *Grounds, Aberdeen, MD*
- **Engineer Research and Development Center (ERDC) DSRC** *Information Technology Laboratory, Vicksburg, MS*
- **Navy DSRC** *Stennis Space Center, MS*

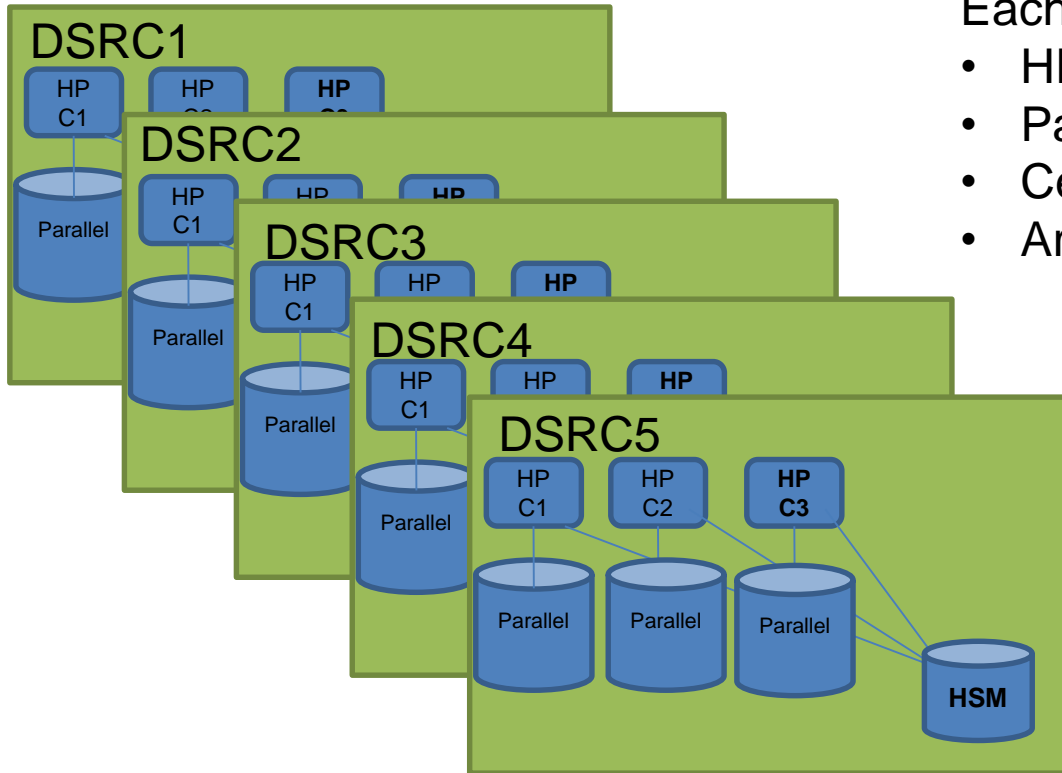
- **One Vanguard Center** provides exploratory technology evaluation to HPCMP and s

- **Maui High Performance Computing Center**

**Today, HPC Centers supports
22 HPC systems
with 995,896 cores and
45.6 PetaFLOPS capability**

HI

HPCMP Current Data Environment



Each DSRC is made up of the following:

- HPC systems
- Parallel File System (Lustre/GPFS)
- Center-Wide File system (GPFS)
- Archival Storage (SAM-FS/Tape)

Most of our applications currently require data access via POSIX interface

Current Customer Workflow

We have two completely different types of workflows

- Research
- Time Sensitive

Research Workflow

Users ingest data to desired DSRC HPC system (or systems)

- Users run applications on HPC systems, store data on regularly scrubbed parallel file system
- Three options for saving data semi-permanently or permanently:
 - Copy data to remote are outside of the program
 - Copy data to the Center-Wide File System
 - Copy data to the HSM at the site of their choice.

Time Sensitive

- **I/O variability hampers processing**
- **I/O tends to be a mix of large block sequential and small block random**
- **Ensured performance is a desired characteristic**

Current Limitations

- **Duplication of data across sites**
- **Staging data to the worksite**
- **Migrating data from old technology to new technology**
- **Quotas**
- **Ability to query data holding based on user, group, and data type**
- **Inability to profile user I/O behavior in real time**

If I had three wishes (or 30)...

When dealing with wishes for metadata and archival at scale there are three perspectives. That of the user, the administrator, and the acquisition team (show me the money). Let's look into the wishes of each perspective...

What a user wants

- **Metadata**

- Ability to find physical location of all copies of data.
- How many files and total capacity being used (how much is left?)
- I/O characteristics of the physical residence
- Staging of data to specific resources
- Chain of custody (user and digital access)
- Extended attributes that are easily searchable
- Ability to enhance/add/change attributes as environment and science change

- **Archive**

- Ability to create transportable archives (physical/cloud)

What an admin needs

- **Real time I/O characteristics**
- **Easily queried sensitivity levels for data**
- **Encryption capabilities and key management**
 - Data at Rest Encryption
 - Data in Flight Encryption
 - Data Encryption per Object
 - User defined Encryption
 - Group defined Encryption

What and admin also needs

- **Chain of Custody**
- **User, Project, Organizational level reporting for utilization and data location**
- **Data collection and movement for user/project/organization**
- **Data curation for lifetime of project**
- **Heat maps of metadata operations during processing**
- **Efficient means to purge data**
- **Ability to prove Qos of I/O subsystem**

Balancing needs within an acquisition

- **Detailed reports (how/what/when/where)**
 - User level reports, System level reports (high level), Physical reports on data at rest (low level), Program level reports (enterprise level), Reports on data in flight (how is data moving)
- **Auditing of lifecycle of data**
- **Analysis of metadata types to data capacity for cost analysis**
- **Amount of time file/object spends on media type (SSD, HDD, tape)**
- **Heat maps of I/O traffic and utilization across the Enterprise**

