Deciding Which Technologies to Adopt, and When

Choosing an HPC Parallel File System

Peter Mills, Assistant Director
Center for Institutional Research Computing
Washington State University

June 2, 2020

Which Resources to Use for HPC

- Cloud service providers (Azure, AWS)
  - Regulated data (HIPAA, CUI)
  - Hardware 10x more expensive than on-prem
- Leadership class facilities (XSEDE)
- High-end workstations (Nvidia DGX)
- On-premise HPC facilities

HPC Leading-Edge Technologies

- Looking at a concrete scenario
- Upgrading 4000 core cluster, standing up another
- Facing choice of parallel file systems, 1PB range
- Evolving landscape with new technologies and companies
- Advanced technologies with proven fundamentals, but not widely adopted
- What is the decision process

Selection Criteria

- Research the available technologies, in depth
- Desired feature set
  - Distributed erasure coding over commodity hardware
  - Tiering to object store using S3 (vs staging to burst buffer)
  - Cloud interoperability
  - NVMe flash storage
  - Directory quotas, snapshots
  - High-performance (io500)

Choosing a Finalist

- Candidates selected based on desired feature set
- Refinement based on
  - Best match
  - Product maturity
  - Customer base
  - Company solvency

Protecting Yourself

- Careful on-site evaluation of
  - Feature set
  - Performance
  - High availability
  - Migration
- Fallback plan
  - If purchased technologies do not pan out