OSiRIS (NSF CC*DNI DIBBs)

March 14 2016

USATLAS Meeting at the Clemson OSG AHM

Shawn McKee

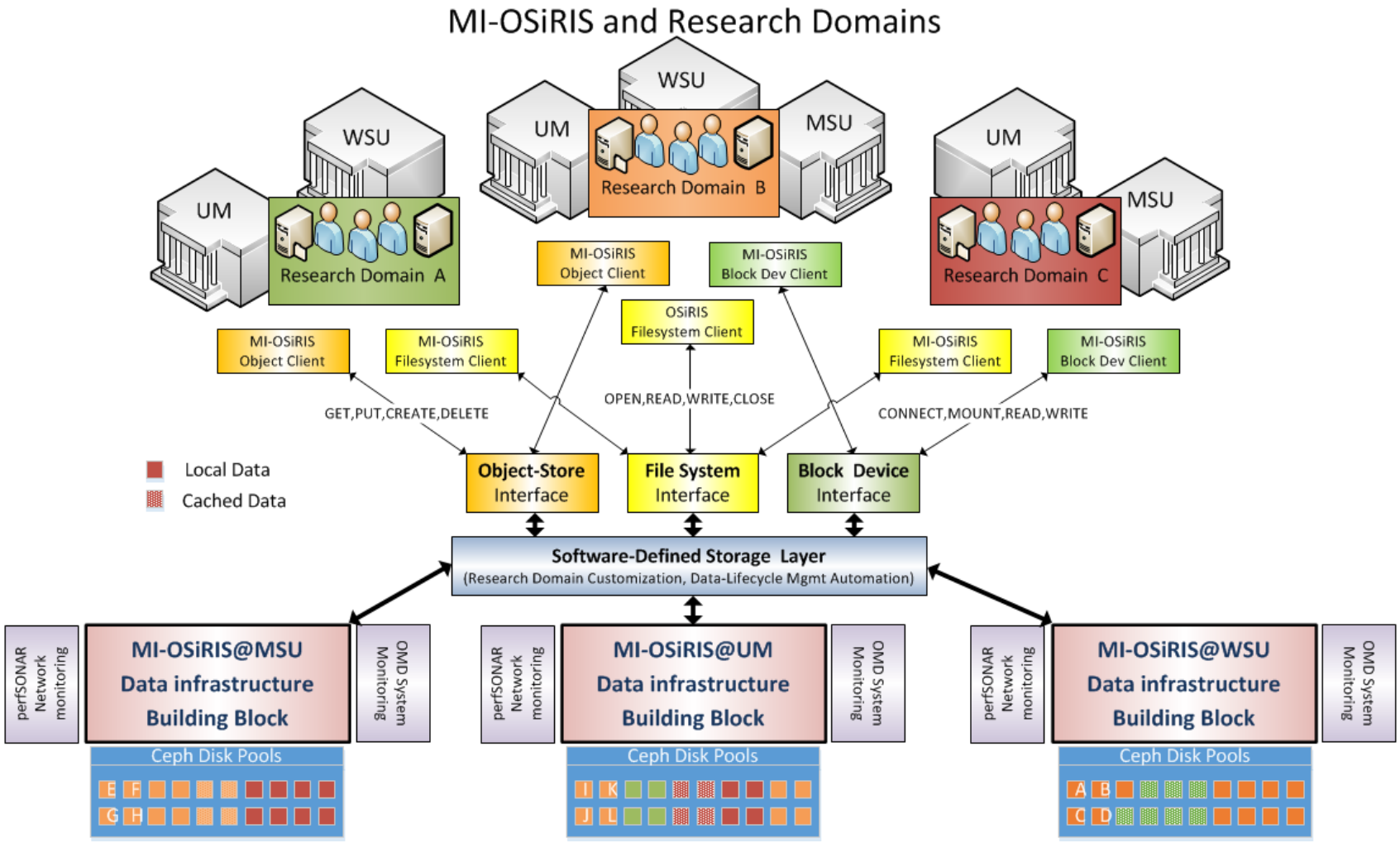
OSiRIS (NSF CC*DNI DIBBs)

- OSiRIS (Open Storage Research Infrastructure) is a NSF CC*DNI DIBBs grant awarded to the University of Michigan, Michigan State, Wayne State and Indiana University.
  - See https://nsf.gov/awardsearch/showAward?AWD_ID=1541335
  - One of 4 funded nationally by this program
  - Approximately $5M (~$1M/year)
- OSiRIS will provide a distributed, multi-institutional storage infrastructure that lets researchers write, manage, and share data from their own computing facility locations.
- The goal is transparent, high-performance access to the same storage infrastructure from well-connected computing locations on any participating campus.
  - It includes network discovery, monitoring and management(SDN) tools as well as the creative use of Ceph features.
  - The project will provide data sharing, archiving, security and life-cycle management, implemented and maintained with a single distributed service.

The Challenge We Are Meeting

- Scientists working with large amounts of data face many obstacles in conducting their research
  - Typically the workflow needed to get data to where they can process it becomes a substantial burden (along with the bookkeeping)
- The problem intensifies when adding in collaboration across their institution or especially beyond their institution.
- Institutions have sometimes responded to this challenge by constructing specialized infrastructures to support specific science domain needs.
  - This doesn't scale and can be expensive (in many ways)
- The OSiRIS team proposed a research project to investigate a possible solution

Logical View of OSiRIS



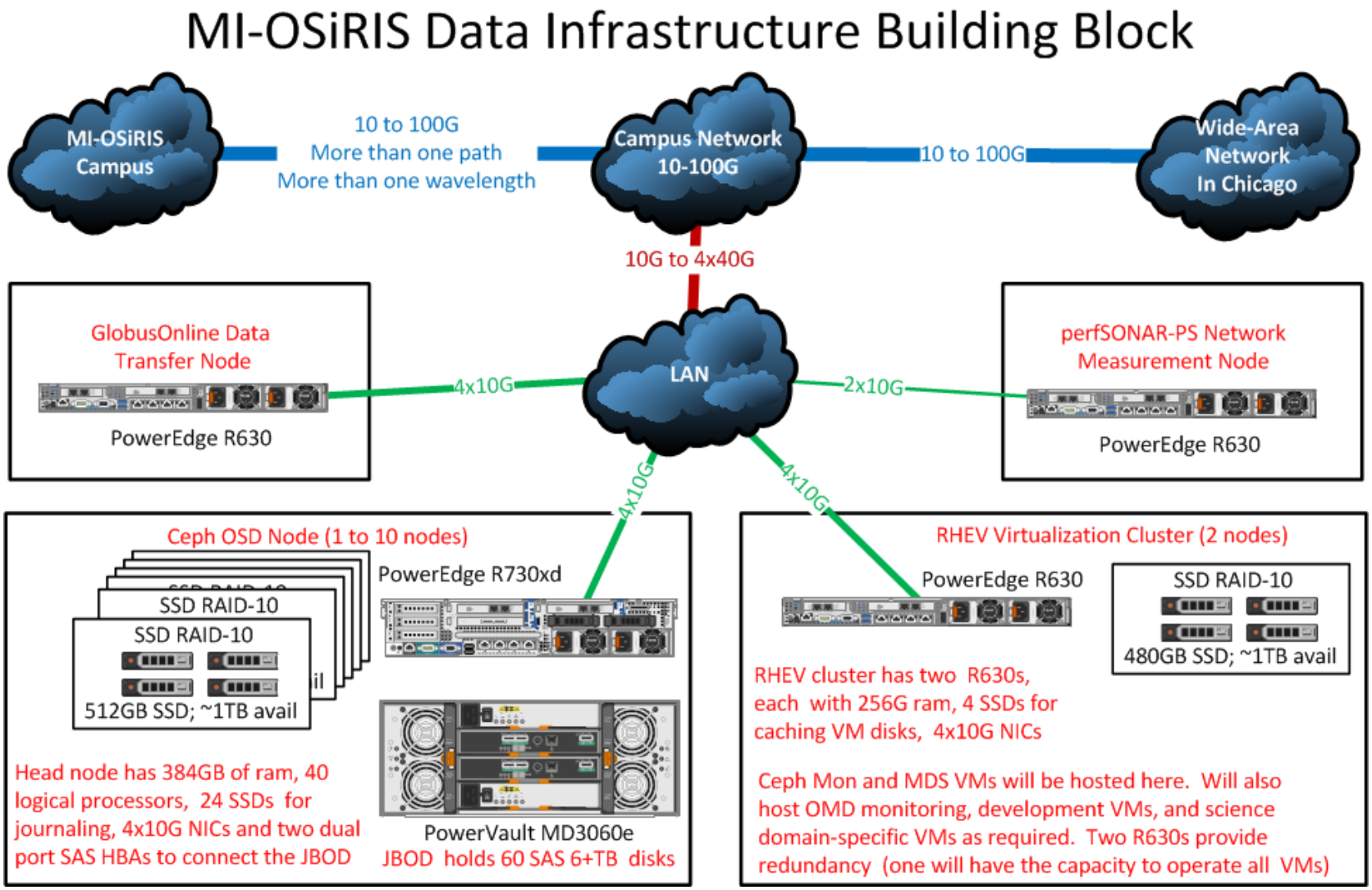
Why OSiRIS?

- Users get customized, optimized data interfaces for their multi-institutional data needs.
- Network topology and perfSONAR-based monitoring components ensure the distributed system can optimize its use of the network for performance and resiliency.
- OSiRIS, via CEPH, provides seamless rebalancing and expansion of the storage.
- **A single, scalable infrastructure is much easier to build and maintain**
- Allows universities to reduce cost via economies-of–scale while better meeting the research needs of their campus.
- Eliminates isolated science data silos on campus.
  - Data sharing, archiving, security and life-cycle management are feasible to implement and maintain with a single distributed service.
  - Data infrastructure view for each research domain can be optimized

OSiRIS Features

- Enables storage-integrated **data lifecycle management** possibilities.
- Provides a **software-defined storage service** for multi-institutional domains; tunable per science domain
- Built from inexpensive, commercial off-the-shelf (COTS) components.
- Robust, reliable infrastructure constructed to minimize single points of failure.
- **High performance** by integrating high-bandwidth network links, a large number of disk spindles, large server memory and SSDs.
- **Built using open source software components** allowing the project to customize specific components to provide required OSiRIS functionality.
- **Incorporates detailed monitoring** of the infrastructure/network topology.
- Enables cloud infrastructures, e.g., OpenStack("OpenStack," 2014).
- Collaborative storage for researchers/resources outside of our campuses.
  - Connected to the Internet at 10 to 100 Gbps
  - Uses Globus Connect(Alliance, 2014) GridFTP-based Data Transfer Nodes (each at 4x25Gbps) for wide-area data transfer

An OSiRIS Institutional Deployment



Planned OSiRIS Storage Evolution

MI-OSiRIS Storage	Year-1	Year-2	Year-3	Year-4	Year-5
Disk Size (TB)	6	8	8	12	15
Disk Count(total)	180	360	660	960	1380
Raw Space(TB)	1,080	2,520	4,920	8,520	14,820
Usable, 2-copies, 0% cache	**540**	**1,260**	**2,460**	**4,260**	**7,410**
Usable, 2-copies, 16% cache	**450**	**1,050**	**2,050**	**3,550**	**6,175**
Usable, 3-copies, 0% cache	**360**	**840**	**1,640**	**2,840**	**4,940**
Usable, 3-copies, 16% cache	**300**	**700**	**1,367**	**2,367**	**4,117**
Erasure Enc. (8+2), 16% cache	**720**	**1,680**	**3,280**	**5,680**	**9,880**

- The above table shows our assumptions for the amount of storage we might provide in OSiRIS by year.
- We have options on how we organize the disks for **each** science domain
  - The options trade-off size versus resiliency versus performance
- **We need to determine stakeholder needs and construct a plan to bring our science users into OSiRIS that will match our storage and person resources.**

Project Metrics / Evaluation

Tangible Metrics

- Used bandwidth to/from storage in total, by interface type (object-store, block and file system) and by research domain.
- Storage usage in total, by interface type, by site location and by research domain.
- Cache usage by site and science domain.
- Delivered IOPs in total, by interface type and by research domain.
- Update / Unplanned outages, both for all of OSiRIS and by research domains.
- Data access stats (bandwidth, IOPs) on local site vs. remote site vs. from outside our sites.
- New collaborations initiated because of OSiRIS existence.

Example survey questions for feedback

- What aspects of OSiRIS a most effective in supporting your research goals?
- What aspects of OSiRIS are problematic for you?
- Do you have data needs that are not met by OSiRIS?
- What customizations or tunings would make OSiRIS more effective for you?
- Are you able to collaborate internally and externally using data in OSiRIS?
- What features are missing from OSiRIS to make it better meet your needs?

OSiRIS and ATLAS

- OSiRIS targets many science domain stakeholders
  - Year 1: High-energy Physics, High-resolution Ocean Modeling
  - Year 2-5: Biosocial Methods and Population Studies, Aquatic Bio-Geochemistry, Neurodegenerative Disease Studies, Statistical Genetics, Genomics and Bioinformatics, Remaining participants, New Science Domains
- ATLAS (using AGLT2: UM and MSU) will be the first OSiRIS Science Domain incorporated
- **Two use-cases:** 1) OpenStack customized VM storage and 2) PANDA event service
  1. We deploy customized worker nodes based upon workload
  2. Use Ceph's object-store to provide a high-performance event service for HPC and Cloud resources as well as our Tier-2
- Very interested in ATLAS input and collaboration
- Homepage http://www.osris.org/

Moving Ahead with OSiRIS/ATLAS

- How best to map ATLAS workflows and needs onto OSiRIS?
  - Event service "server" seems to be a logical target…regional service for nearby HPC or Cloud resources?
  - Customized VM image server (for Openstack)
  - Other?
- We would welcome input on integrating ATLAS workflows with OSiRIS capabilities
  - Start integration this summer.

OSiRIS Links

- Homepage http://www.osris.org (couldn't get osiris.org ☹)
- GitHub: https://github.com/MI-OSiRIS
- OpenProject: https://oproj.aglt2.org/projects/os
- DokuWiki (uses Shibboleth): https://wiki.osris.org/doku.php?id=reference:start

- See OSiRIS technical talk tomorrow afternoon in OSG Technical Session

Questions?