OSG All Hands 2022 Report

Brian Bockelman, Brian Lin
Center for High Throughput Computing
Morgridge Institute, UW–Madison

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

HT CENTER FOR HIGH THROUGHPUT COMPUTING

OSG All Hands

- Annual meeting for OSG users, resource providers, and staff
- Fully virtual this year
- Nine ~1.5h sessions spread over five days
  - ~1h discussion sessions each day
  - User and admin office hours
- https://indico.fnal.gov/event/53029/
- 251 attendees from 149 institutions
- 46 speakers

State of OSG

Institutions contributing to OSCF

- **64 US institutions contributed compute resources** during the year ending on March 9th 2022.
  - We count institutions, organizations inside institutions & "clusters"
    - E.g. UCSD is an institution with 2 "green dots": SDSC & CMS group in Physics. 6 clusters contributed resources in 2021, incl. one entirely inside the commercial cloud, and another that itself has hardware across 30 institutions.
- Out of these 64 institutions in the USA
  - **9 are Minority Serving Institutions (MSI)**
    - 1 CC*, 2 EPSCoR, 1 Non-R1
  - 10 are in EPSCoR states
  - 17 received a CC* award for a compute cluster
  - 13 are non-R1
  - 31 are none of the above

26 of the 64 US institutions are either MSI, EPSCoR, or non-R1

OSG is Democratizing Access to Cyberinfrastructure

11

Top Users of Data Federation

Project	Data Read	Working Set	Reread multiplier
LIGO (Private)	10PB	38TB	264
Minerva	5.6PB	3.1TB	1,789
NOVA	2.6PB	1.9TB	1,348
LIGO (Public)	2.4PB	38TB	67
Tufts_Hempstead	2.0PB	380GB	5,321
DUNE	1.6PB	185GB	8,658
Steward	1.0PB	11TB	92
REDTOP	874TB	95TB	9.2
Molcryst	570TB	5GB	115,650
BiomedInfo	530TB	66GB	8,090

17 projects have >TB working sets
11 of these are OSPool users

Tufts Computer Architecture Lab

Steward Observatory Data Analytics

R&D towards future particle physics experiment

Quantum chemical and machine learning insights into supra-molecular organization of molecular crystals

Development and application of software tools for performing large-scale biomedical informatics on microbial genome sequence data.

26

Integrating a Diversity of Resources

- Access Points (APs) play a central role in harnessing capacity of a diverse set of resources
- Green resources through providers, like Lancium
- Kubernetes-based resources, like the National Research Platform (NRP)

PATh Facility

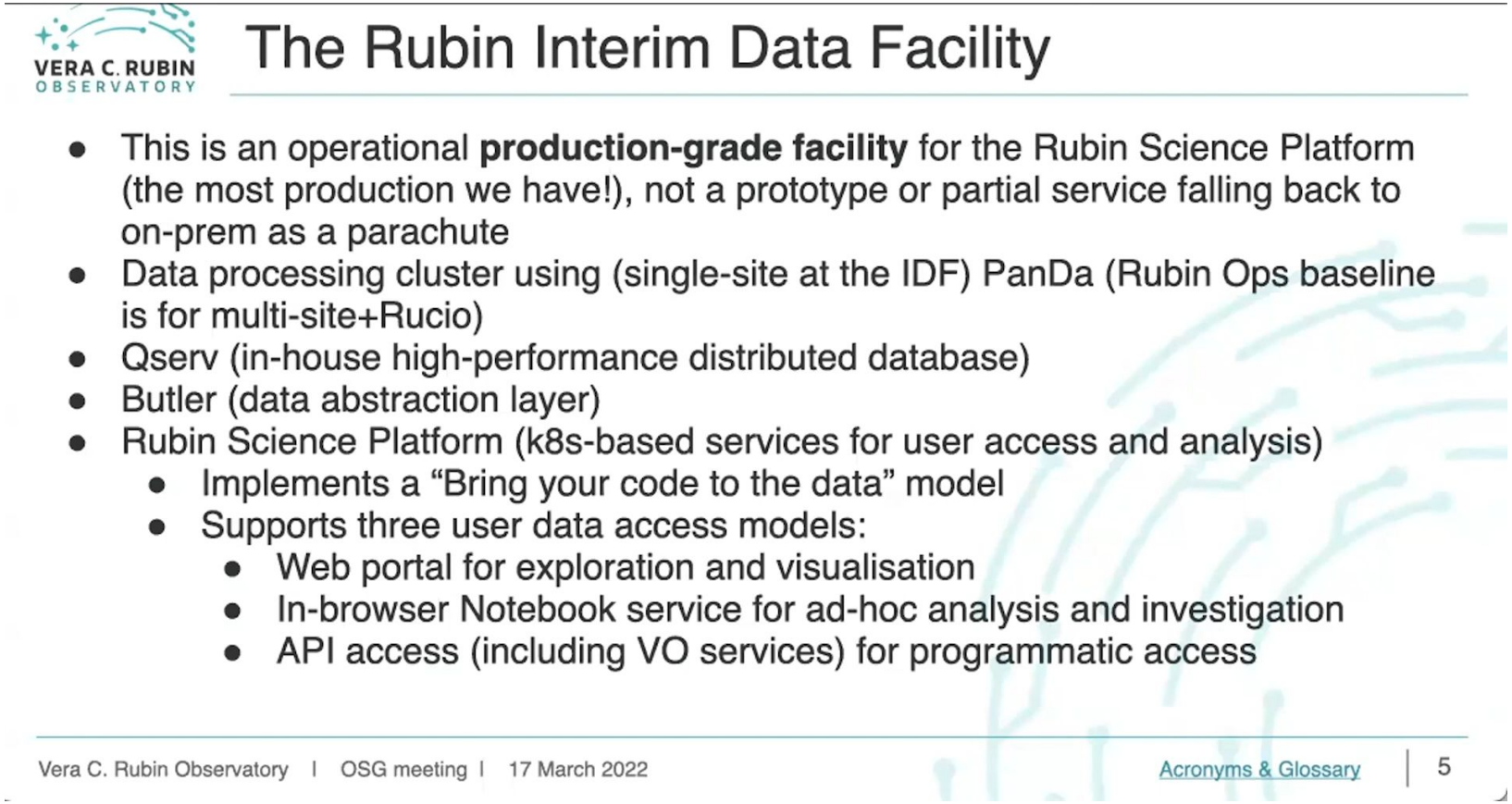
- Six distributed sites make up this dHTC-dedicated facility
- Researchers given access via HTC credit accounts
- Each site is a GitOps-managed Kubernetes cluster
- Soon: Open Science Data Federation (OSDF) caches

The Research

Analyzing Public LIGO Data Using OSG

- Analyzed using two workflows
  - Open Science Pool
  - Cluster-specific submits using allocations (contact: Matyas Selmeci)
- Needed order 10 * more disk space for multiple users
- 5 days of data typically ran to completion in < 3 days under ideal circumstances on open pool

The Research

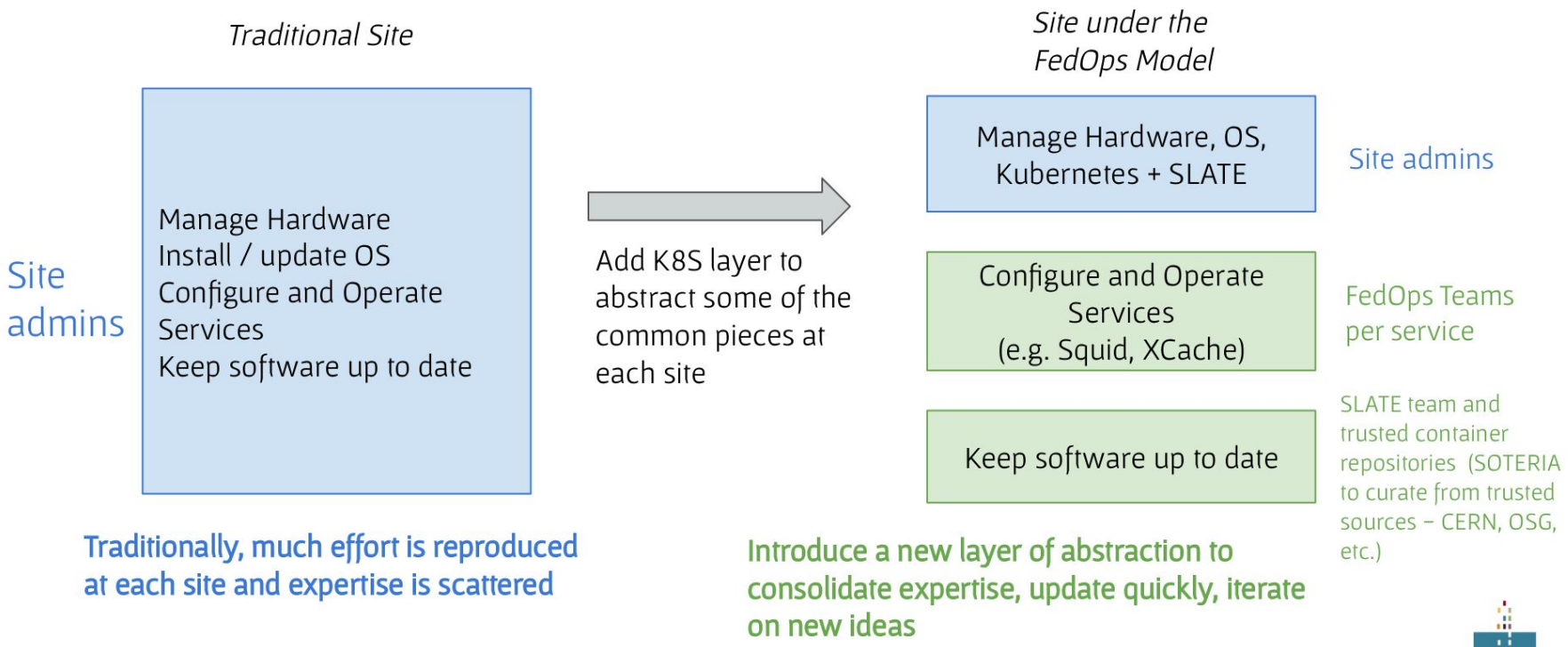


The Technology

- Token transition
  - OSG 3.5 (last of GCT support) ending on May 1, 2022
  - HTCondor 9.0 support extended to Feb 1, 2023
  - Concerns about token-readiness of VOs using DIRAC and how sites can bridge the gap
  - Vault: a pluggable solution for user tokens
- OSDF monitoring (https://osdf.osg-htc.org)
  - Validation tests uncovered XRootD monitoring packet loss at scale
  - Solution: design a site “shoveler” to turn UDP packets into a resilient format (Message Bus)
  - Shoveler deployed at several sites
  - Next step: distribute shoveler “side cars” alongside caches

The Technology

Evolving our T2 model



17

LHC

A full day dedicated to discussions on the following topics:

- XRootD development
- Network activities and plans
- HPC and clouds
- Analysis Grand Challenge
- Token transition and GSI retirement
- Caching and data lake infrastructure

Questions?

This material is based upon work supported by the National Science Foundation under Grant Nos. 1148698, 1836650 and 2030508. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.