HEPCloud: An Elastic Hybrid HEP Facility using an Intelligent Decision Support System

Eric Vaandering on behalf of the HEPCloud team

CHEP 2018

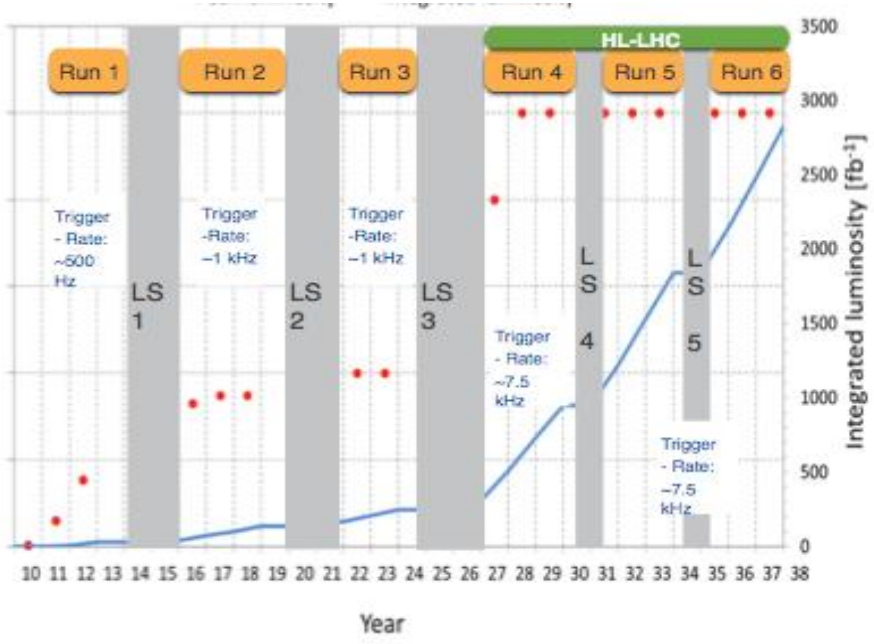
July 9, 2018

Disclaimer

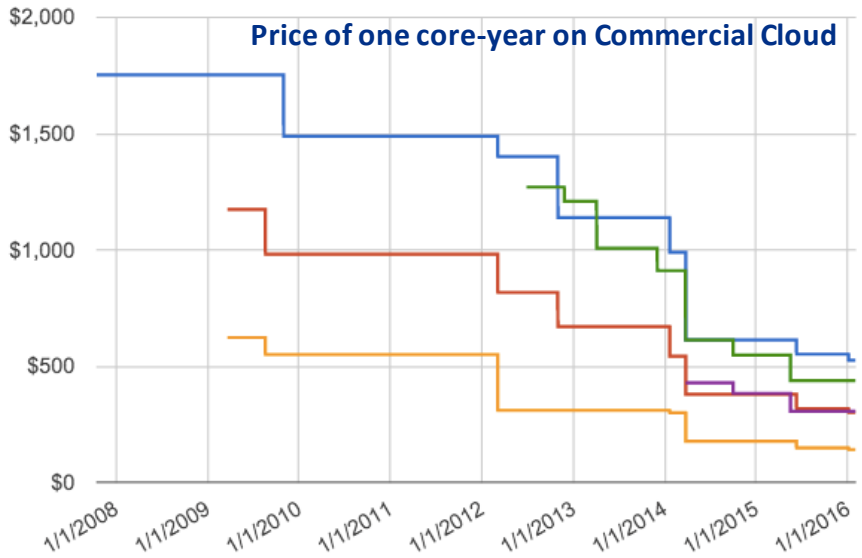
Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof.

Drivers for Evolving the Facility: Capacity, Functionality and Cost

- High Energy Physics computing needs will be 10-100x current capacity
  - Two new programs coming online (DUNE, High-Luminosity LHC), while new physics search programs (Mu2e) will be operating

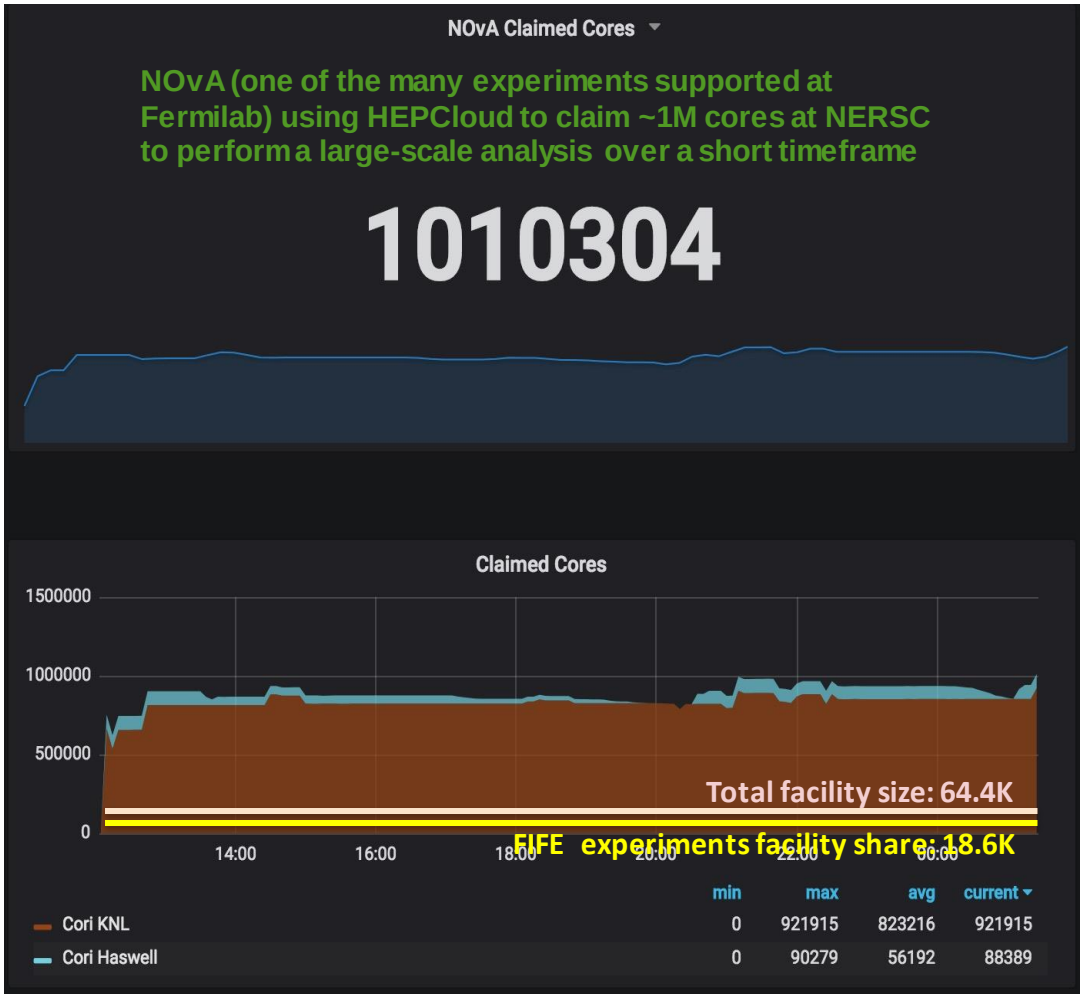


- Scale of industry at or above R&D
  - Commercial clouds offering increased **value** for decreased **cost** compared to the past
- Industry is offering variety of hardware configurations that are becoming relevant to the HEP
  - GPUs for Machine Learning



Drivers for Evolving the Facility: Elasticity

- Usage is not steady-state
- Computing schedules driven by real-world considerations (detector, accelerator, …) but also ingenuity – this is research and development of cutting-edge science

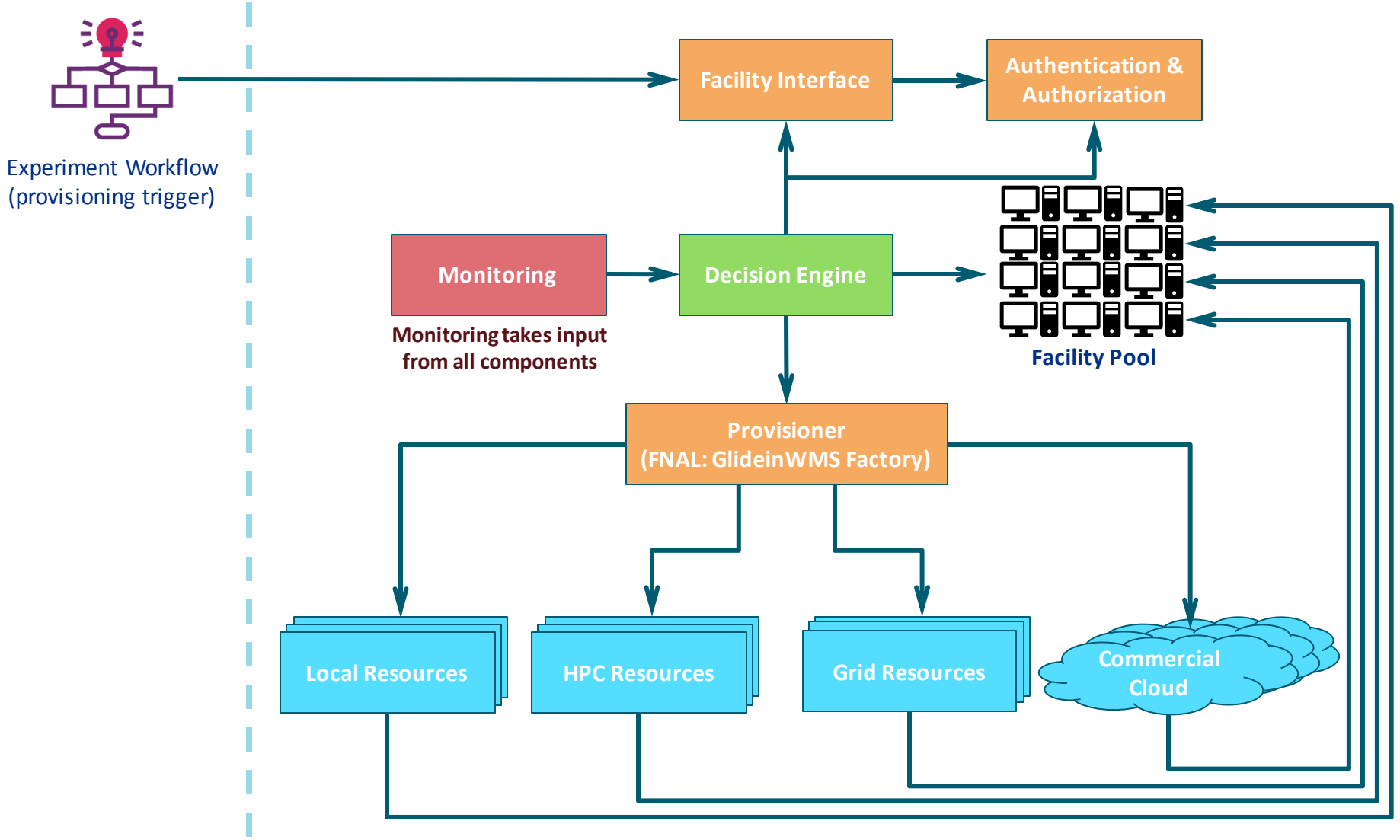


News article:
http://news.fnal.gov/2018/07/fermilab-computing-experts-bolster-nova-evidence-1-million-cores-consumed/

HEPCloud: the Evolved Facility

- Started as a pilot project in 2015 to explore feasibility, capability of HEPCloud
  - Goal of moving into production by end of 2018
  - Seed money provided by industry
- A portal to an ecosystem of diverse computing resources commercial or academic
- Provides “complete solutions” to users, with agreed upon levels of service
- The Facility routes to local or remote resources based on workflow requirements, cost, and efficiency of accessing various resources
- Manages allocations of users to target compute engines

HEPCloud Architecture

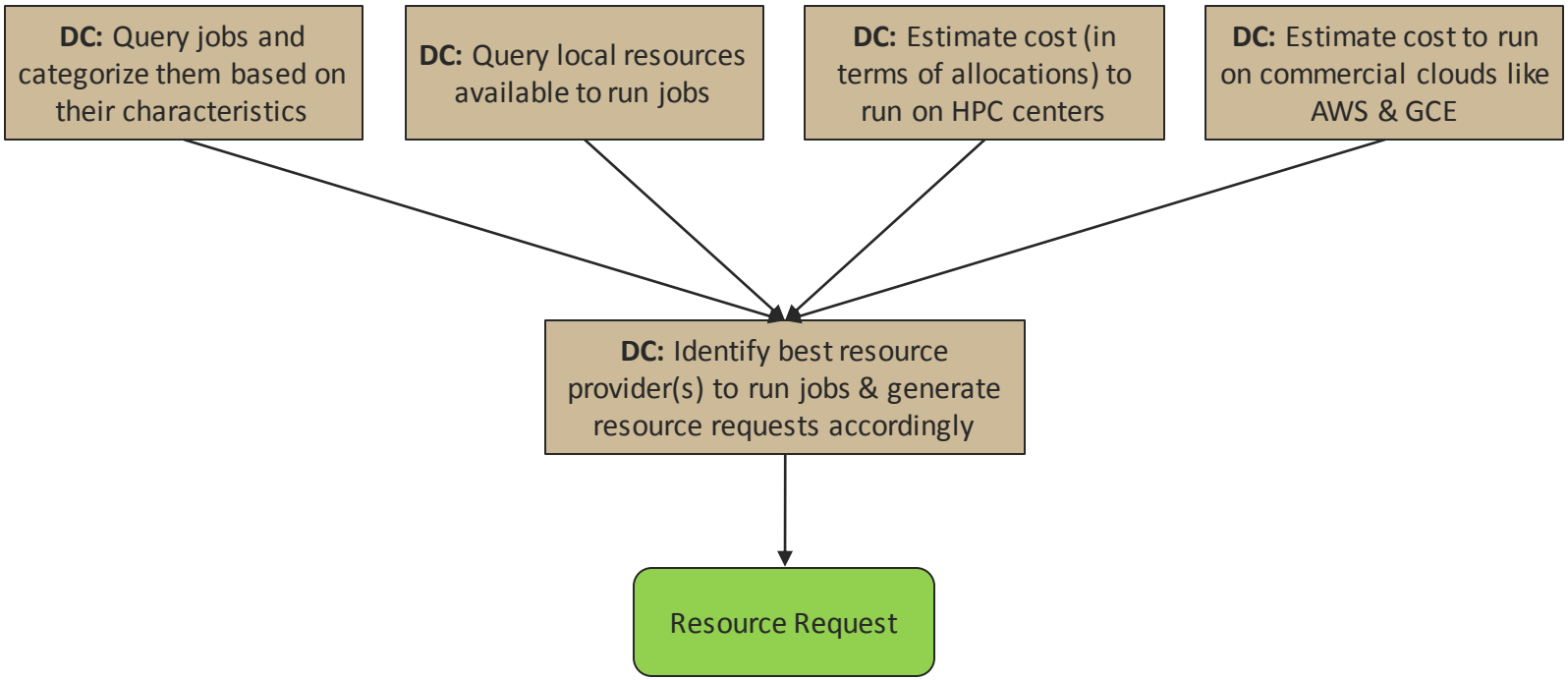


Decision Engine (DE)

- A modular intelligent decision support system (IDSS)
  - Makes decisions that aid in the automatic provisioning of resources
  - Supports multiple types of resources
    - Cloud providers (Amazon AWS, Google GCE)
    - HPC centers (NERSC)
    - Grid computing federations.
- Design Drivers
  - Framework supports user supplied plugins
    - Allows the injection of user-supplied code and expert knowledge
  - Powerful configuration functionality
  - Validity of data used during decision making process
  - Ability to replay decisions

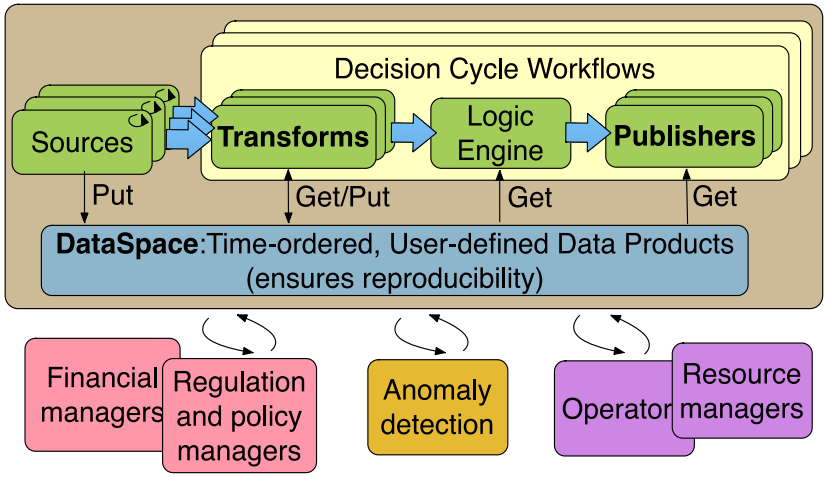
Decision Engine: Typical Provisioning Workflow

- ***Decision Channel (DC)*** is a grouping of tasks that generates decision
- ***Decision*** consists of recommendation of one or more actions that should be executed, actions that are directly executed, or both



Decision Channel Components

- ***Source*** modules are responsible for communicating with an external system (via native APIs to that system) to gather data that acts as input to the system
- ***Transform*** modules contain algorithms to convert input data into new data. A transform consumes one or more data products (produced by one or more sources, transforms or both) within a Decision Channel and produces one or more new data products
- ***Logic Engine*** is a rule-based forward chaining inference engine that helps in the decision making process
- ***Publisher*** modules are responsible for publishing results to external systems
- The ***DataSpace*** acts as a Knowledge Management system



Decision Engine with VC3 as Resource Provisioner

- VC3 (Virtual Clusters for Community Computation) is a project that allows groups of researchers to **aggregate** their **resource allocations** to run fully dynamic virtual clusters of multiple types (HTCondor, WorkQueue, Jupyter, Spark, etc.)
- BNL wrote **DE modules to enable VC3** as a mechanism **to provision resources**
- **Key benefit:**
  - DE (or any system using DE, like HEPCloud) can leverage all resources already defined and curated within VC3 by delegating the provisioning
- This work also demonstrates
  - VC3 can be used programmatically by external tools (previously only web and command-line control was supported)
  - VC3 can be enhanced to allow external head nodes (previously only VC3-created head nodes were supported)

BROOKHAVEN NATIONAL LABORATORY

Fermilab

Conclusion

- The Fermilab HEPCloud facility will be the primary system for provisioning compute resources across HPC, Clouds and Grids, for all Fermilab-affiliated experiments
  - HEPCloud will be in production by the end of 2018
- HEPCloud will be using the DE to provide elasticity to the facility
- The DE architecture is flexible and modular
  - Facilities/Experiments can write their own DE modules (source, transform, logic, and publishers) to do non-trivial functions and use different provisioners
  - Fermilab's instance of the HEPCloud uses GlideinWMS factory as the resource provisioner
  - BNL was able to develop Decision Engine modules to use VC3 to provision resources

Backup Slides

Changing Roles of HEP Facilities

- Strategic Plan for U.S. Particle Physics (P5 Report)

  Rapidly evolving computer architectures and increasing data volumes require effective crosscutting solutions that are being developed in other science disciplines and in industry. Mechanisms are needed for the continued maintenance and development of major software frameworks and tools for particle physics and long-term data and software preservation, as well as investments to exploit next-generation hardware and computing models.

- Need to evolve the facility beyond present infrastructure

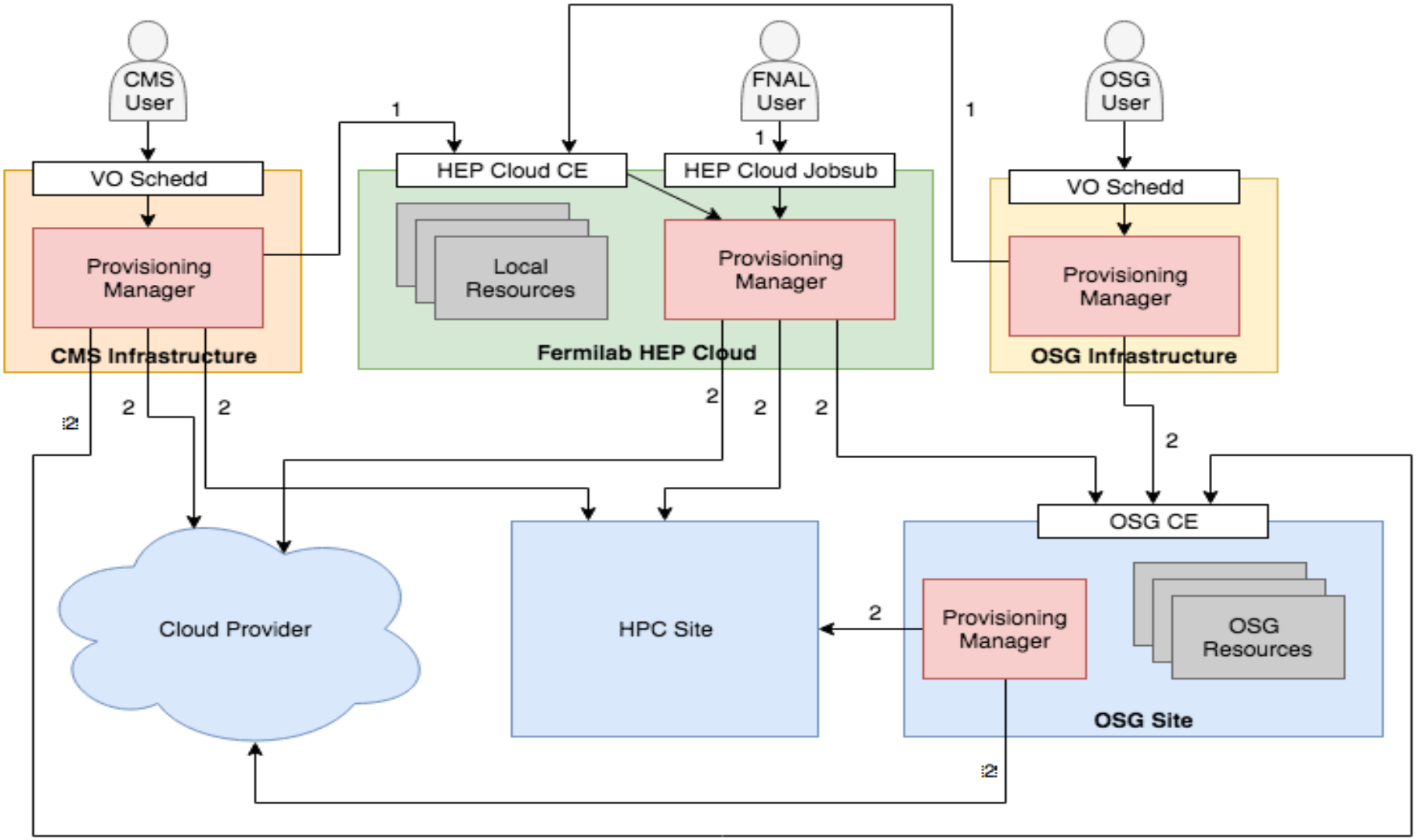
HTC, HPC Cores
64.4K total site

Tape
102 PB total site

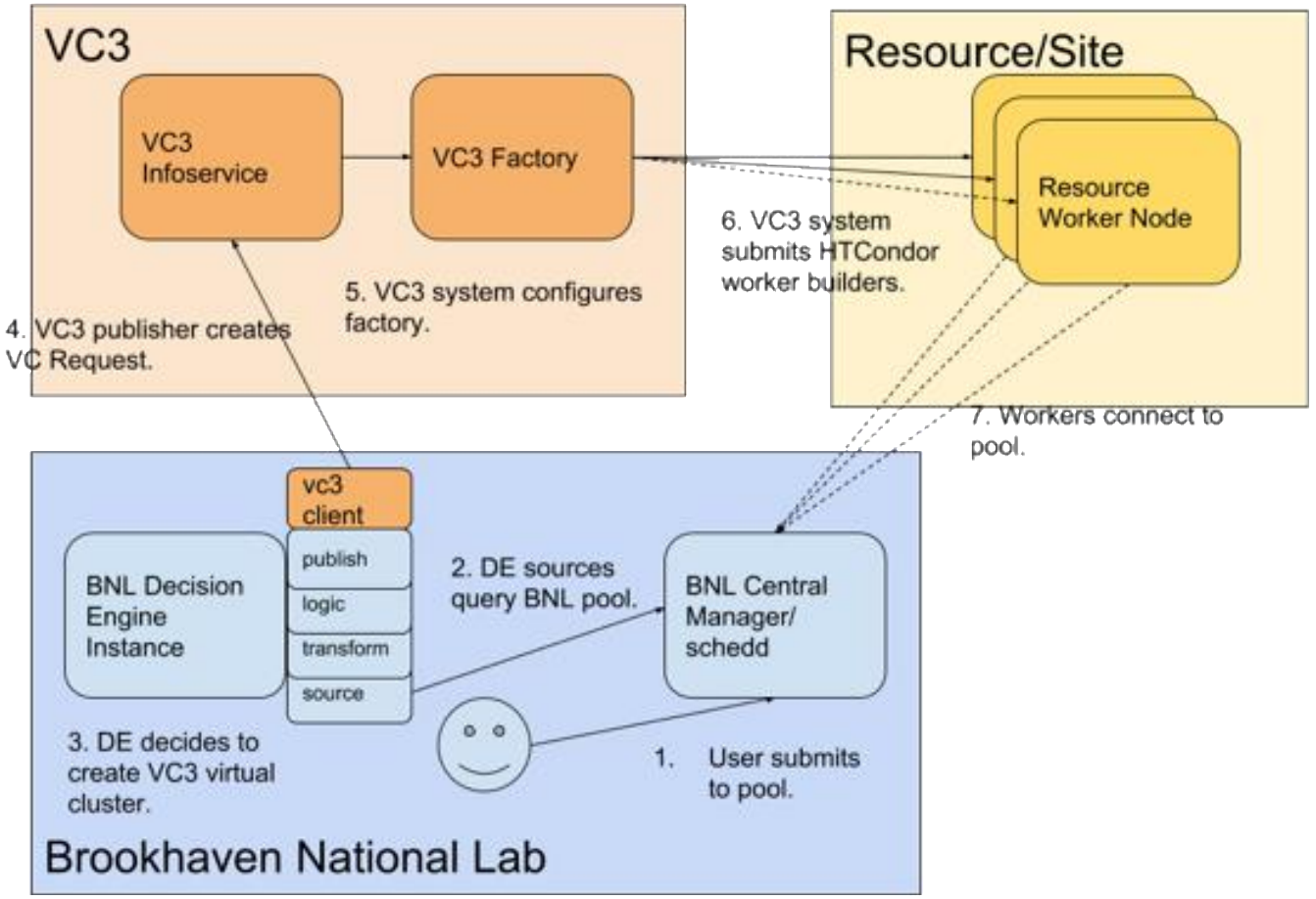
Disk Systems
47.6 PB total site

Networking
10/100 Gb
~5k internal network ports

User's View of the HEPCloud Facility



Decision Engine with VC3 as Resource Provisioner



Decision Engine Architecture

- A ***Configuration Manager*** acts as the Configuration Factory
- ***Task Managers*** are responsible for the scheduling of Decision channels tasks
- A single ***Engine*** is responsible for coordinating the Task Managers