

Future Laboratory Computing and BigPanDA

National Laboratory Research Computing Group (NLRCG)– Initiated Spring 2017

The mission of the National Laboratory Research Computing Group (NLRCG) is ***to improve the National Laboratories abilities, both independently and collectively, to effectively support research computing*** now and into the future. In addition, the NLRCG aims to ***collaborate with the SC to jointly address future research computing support challenges.***

The NLRCG is charged by the National Laboratories NLCIO and CRO working groups, and ASCR. The NLRCG will periodically report out to these stakeholders to keep them apprised of progress.

Members of the NLRCG will be made up of researchers, scientists, technologists, and other domain experts from within the DOE Lab complex. A lead POC will be appointed by each Lab's CIO and/or CRO.

The NLRCG will hold regular meetings and may bring in outside experts to obtain the information needed to complete its tasks.

Future Lab Computing Thrust

The **Future Lab Computing Thrust (FLC)** will identify best practices and research challenges leading to the creation and operation of a DOE/SCwide federated Distributed Computing and Data Ecosystem (DCDE).

Leads: Eric Lançon (BNL), Arjun Shankar (ORNL)

Steward: Richard Carlson (ASCR)

Capabilities and Background Reviewed

- Science Use Case

- NWChemEX - Theresa Windus
- LCLS – Amedeo Perazzo

- Facilities

- OLCF - Jack Wells
- NERSC - David Skinner
- EMSL - Dave Cowley
- ESNET - Eli Dart
- JLAB - Amber Boehnlein
- CADES@ORNL - Arjun Shankar
- Data@BNL - Eric Lancon
- LCRC@ANL - Ray Bair

- Middlewares/Tools

- Swift - Michael Wilde
- HEP Cloud - Panagiotis Spentzouris
- Pegasus - Ewa Deelman
- NSF Collaborative - Shantenu Jha

- Data and Data Movement

- RUCIO – Vincent Garonne
- IRODS – Jason Cposky

- Security

- Authentication and Authorization and ID Management - Von Welch

Recurring themes

- Make the SC complex computing look like a coherent whole
- Combine capabilities – surge instead of planning for peaks
- Growing scientific collaborations and a forthcoming deluge of data
- Need to collaborate with other thrust areas
 - Data management
 - Federated ID (primarily not a technical issue)
- Easier access to data at different facilities
- Consistent levels of services across providers and facilities (Labs, universities, networks,...)
- Mid-scale facilities required for development, tests, glue across facilities
- Machine-Learning
- Containers for portability

FLC-WG Preliminary recommendations

- Establish pilot(s) that implement all (most?) of following items/topics
 - ID management
 - Allocation & accounting
 - Ease of use
 - Variety of use
 - Long-term data management
 - Governance mechanisms

BigPanDA and FLC

- Alexei talked with Rich Carlson at ASCR PI meeting in September and to Eric Lancon after the meeting
 - Preliminary agreement about participation of BigPanDA in the FLC pilot project
 - I was asked to get in touch with group leader at RACF in charge of the pilot project implementation
 - I talked to Eric Lancon and John Hover who is currently tasked with pilot project implementation at BNL
 - The project is in early stages
 - Emphasis seems to be on federated identity system, WMS seems to be a distant priority
 - In terms of the project resources BNL HTC cluster and CADES@OLCF will be involved
 - Resources at ANL, PNNL and NERSC will be available later
 - In principle we can start WMS part in parallel to the main activities in the pilot project
 - We can use PanDA server at EC2 for the project
 - Proceed with Harvester installations at BNL and talked to Arjun about Harvester at CADES
 - Later on, when FedID solution will be identified, we can start work on PanDA integration with FedID system

DCDE Components

Seamless user access

Governance

