Future Laboratory Computing

WORKING GROUP – GOALS AND STATUS

Eric Lançon, Ph.D., BNL
Arjun Shankar, Ph.D., ORNL

DOE SC Steward: Richard Carlson

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.

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

Four defined thrusts for labs to work together:

- Economies of Scale
- Blockers to Collaboration
- Scientific Data Management
- Future Laboratory Computing – DOE SC Interest

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/SC wide federated Distributed Computing and Data Ecosystem (DCDE).

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

Steward: Richard Carlson (ASCR)

Charge – Part I

The National Laboratory Research Computing Group (NLRCG) and the Advanced Scientific Computing Research (ASCR) program office jointly establish the Future Lab Computing Working Group (FLC-WG) thrust to identify best practices and research challenges leading to the creation and operation of a DOE/SC wide federated Distributed Computing & Data Ecosystem (DCDE). The FLC-WG will accomplish this by:

- Reviewing past ASCR research activities
- Reviewing current lab, scientific community, academic, and commercial solutions
- Synthesizing past and current technical activities into a set of best practice documents
- Identify current and future challenges that need to be addressed
- Determine the activities needed to pilot an operational deployment of a DCDE

Charge – Part II

- The outcome of the FLC-WG will be a report describing the current state of lab computing, a set of research challenges, and plans for achieving a federated DCDE.

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

- The NLRCG FLC-WG thrust will coordinate its activities with other NLRCG thrust working groups to obtain relevant background information and to ensure SC wide integration of the practices and recommendations

Future Lab Computing Thrust

Plans – Initiated Summer 2017

- **Near term**
  - Assemble working group members
  - Circulate questionnaire including key topics in the area of DCDE
    - Distributed resource management, Workflow/workload management, Interactive access, Surge computing, Virtualization/Containers, Streaming Data, User Model, Data Storage
  - Assemble list of experts and PIs on topics for DCDE
- **Mid term**
  - Analyze and summarize survey / questionnaire responses, and synthesize past and current technical activities into review of state-of-the-art
  - Invite and set up talks by guest speakers and PIs
- **Report phase**
  - Face-to-face meeting at Supercomputing 2017
  - Determine gaps and activities needed to pilot an operational deployment of a DCDE
  - Incorporate content and best practices from guest speakers and findings into report

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
  - HEPCloud- Panagiotis Spentzouris
  - Pegasus - Ewa Deelman
  - NSF Collaborative - Shantenu Jha
- Data and Data Movement
  - RUCIO – Vincent Garonne
  - IRODS – Jason Coposky
- Security
  - Authentication and Authorization and ID Management - Von Welch

LCLS-II Data Systems Requirements and Planning

Sep 21st 2017
FLC-WG

Amedeo Perazzo
LCLS Controls & Data Systems

Identity Management for Scientific Collaborations

Von Welch

Future Laboratory Computing Working Group (FLC
November 30th

CENTER FOR APPLIED CYBERSECURITY RESEARCH
INDIANA UNIVERSITY
Pervasive Technology Institute

NWChemEx Project

Theresa L. Windus, Ames Laboratory
Thom H. Dunning, Jr., Northwest Institute for Advanced Computing
Robert J. Harrison, Institute for Advanced Computational Science

Swift Parallel Scripting

Presented to:
ture Laboratory Computing Working Group
Nov. 2, 2017

h#p://swi*-lang.org
h#p://parsl-project.org

iRODS - An Overview

Jason Coposky
@jason_coposky
Executive Director, iRODS Consortium

Rucio: Scientific Data Management

Vincent Garonne
on behalf of the Rucio team

ESnet

ENERGY SCIENCES NETWORK

Science Networks and Lab Computing – Supporting Science

Oak Ridge Leadership Computing: Responses to FLC-WG Questions

s (for NCCS/OLCF Team)
cience
eadership Computing Facility
ational Laboratory

ratory Computing Working Group (FLC-WG) tele
017

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. Some of the work presented here is from the TQ1M and Oak Ridge National Laboratory collaboration which is done under the CRADA agreement NFE-14-05227. Some of the experiments were supported by an allocation of advanced computing resources provided by the National Science Foundation. The computations were performed on Nautilus at the National Institute for Computational Sciences.

OAK RIDGE National Laboratory

Computing at JLab

Future of Lab Computing
October 5, 2017

Amber Boehnlein
IT Division Director

Jefferson Lab
Thomas Jefferson National Accelerator Facility

EMSL 20YRS

The EMSL Scientific User Facility and the Proposed DCDE

David Cowley
EMSL Capability Lead
Molecular Science Computing Operations

Scientific Innovation Through Integration ▪ www.emsl.pnnl.gov

Pegasus Workflow Management System

Ewa Deelman, Ph.D.

USC Viterbi

http://pegasus.isi.edu

DCDE Benefits and Challenge

- Right compute resources available to DOE researchers regardless of location
- Data accessible from anywhere
- Data can be moved easily
- Unified seamless computing environment with information technology hurdles
- Challenges
  - Enable self-support
  - Enable platforms as service abstraction
  - Improve protocols for enabling growing number of workflows
  - Create on-the-fly intelligent data access and data management

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

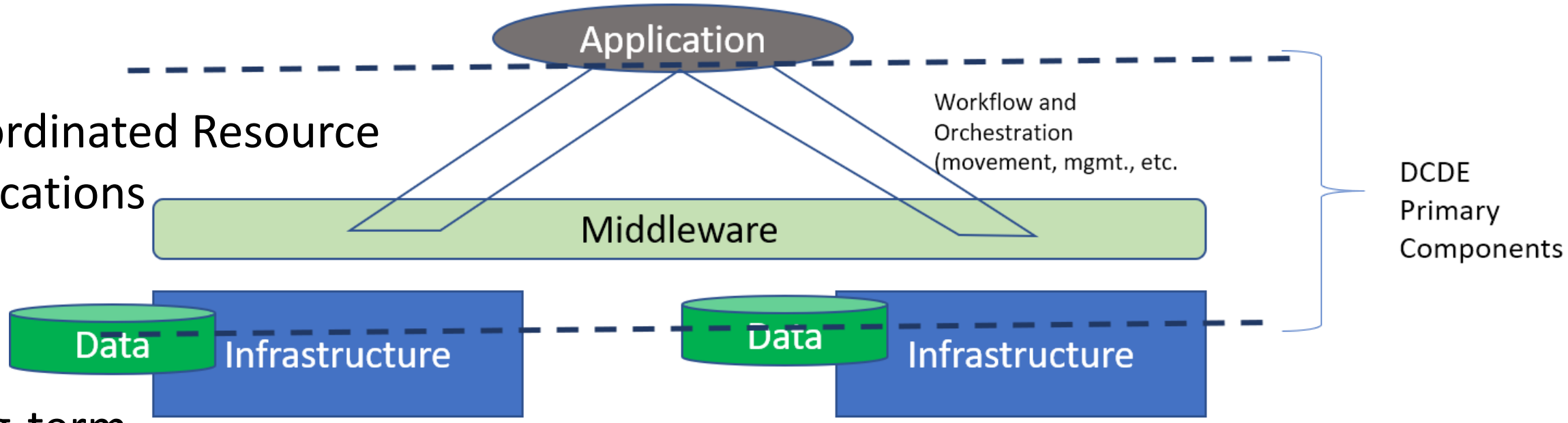
FLC-WG Preliminary recommendations

- 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

DCDE Components

Seamless user access

Governance



Coordinated Resource allocations

Long-term Data

Variety, Portability: containers, etc.

Scaling from institutional to DOE facilities

Thank you!

Questions?