U.S. DEPARTMENT OF ENERGY | Office of Science

DOE Distributed and Collaborative Computing R&D activities

Richard Carlson
richard.carlson@science.doe.gov

LSN-MAGIC

Jan 7, 2015

Grids and Clouds

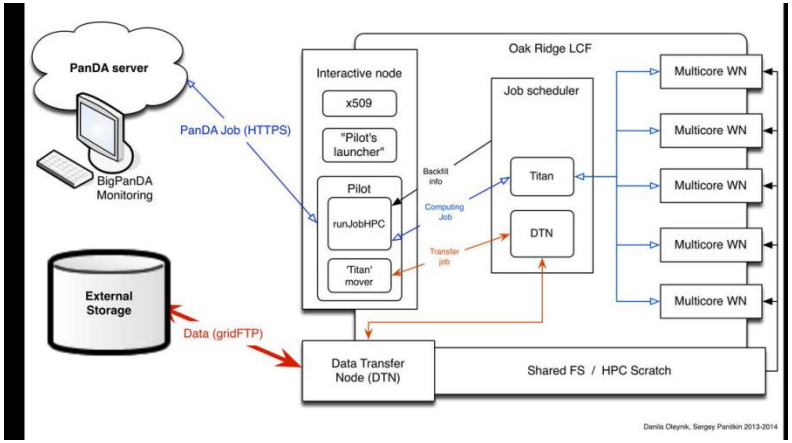
- **Grid research and standardization activities**
  - Globus and Grid pilot projects
  - Partnerships with HEP, FES, NP, Climate
  - Distributed computing for DOE scientists
- **Magellan Project**
  - Explore if/how DOE science communities can best use cloud based technologies
  - http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Magellan_Final_Report.pdf

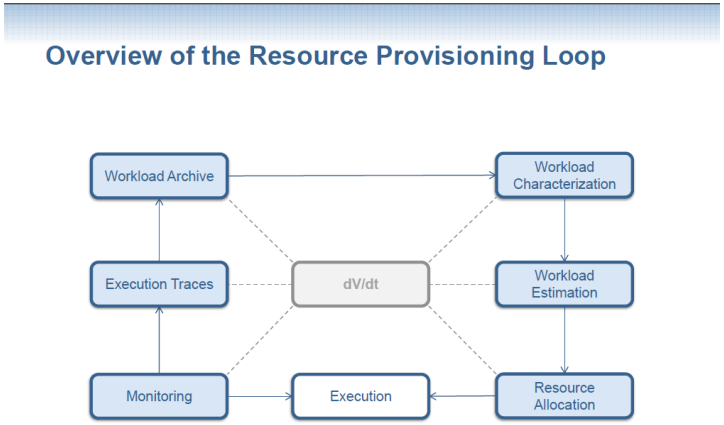
Hardware vs Software

- **ASCR Facilities Division**
  - Funds the purchase and operation of leadership class supercomputers and networks
  - Supports open science research by multiple domain science communities
  - Staff knowledge essential in porting codes and supporting scientists
- **ASCR Research Division**
  - Funds R&D activities that advance scientific discovery through the use of HPC
  - Collaborative science communities use unique instruments and a range of compute/storage resources
  - Current focus is developing a usable Exascale computer

Existing Research Program Highlights

- **Big Panda**
  - Explore using US ATLAS developed Workload Management Software on OLCF Supercomputer (Titan)
- **Data driven computing requirements**
  - Explore the future of Distributed Computing
- **dV/dt**
  - Goal: “make it easier for scientists to conduct large-scale computational tasks that use the power of computing resources they do not own to process data they did not collect with applications they did not develop”





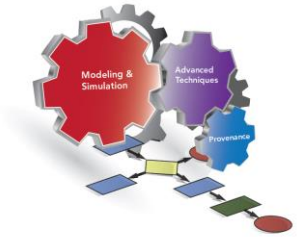
Scientific Workflows

- **Science applications have evolved from simple jobs to large complex workflows**
  - Multiple parameter sweeps to explore potential outcomes
  - Multiple components developed by different experts
  - Geographically distributed teams
  - Relying on data and or analysis codes generated elsewhere
- **ASCR CS/NGNS workflows workshop**
  - Develop concepts to deal with massive amounts of data generated by simulations and instruments
- **Need for Workflow Scientists**
  - Experimental, Theory, and Computational scientists

CS – Computer Science, NGNS – Next Generation Networking for Science

Analytical Modeling for Extreme-Scale Computing Environments

- **Four large projects started in FY14**
  - IPPD: Integrated End-to-End Performance Prediction and Diagnosis for Extreme Scientific Workflows – Kerbson, et.al. (PNNL)
  - Panorama: Predictive Modeling and Diagnostic Monitoring of Extreme Science Workflow – Deelman, et.al. (USC/ISI)
  - Ramses: Robust analytical models for science at extreme scales – Foster, et.al. (ANL)
  - X-Swap: Extreme-Scale Scientific Workflow Analysis and Prediction – Strohmaier, et.al. (LBNL)



PANORAMA

ramses
robust analytical models for science at extreme scale

X-SWAP PROJECT

http://science.energy.gov/ascr/research/next-generation-networking/am-projects/

Potential Distributed Computing Directions

- **Near-time processing of Simulation, Experimental, and Observational data**
  - Data generated by unique science instruments
  - Comparing current and archival data
- **Interactive supercomputing with multiple simultaneous users**
  - Computational steering
  - Dynamic and adaptive workflows
- **Distributed Storage infrastructures**
  - Support Computationally, Experimentally, and Observationally collected data
  - Dynamic distributed resource discovery libraries and services