Dynamic, on-demand resources and provisioning: Demonstrations at the RENCI Booth, #2942

renci

SC11 Seattle, WA Connecting communities through HPC

Bandwidth and Cyberinfrastructure Created As Needed

Scientists with compute or data-intensive problems often need varied cyber resources at multiple sites. They also need the bandwidth to run their jobs on these resources. However, it's not practical or even desirable to maintain resources and high bandwidth pipelines once a job is completed. In three demonstrations, a RENCI and Duke University research team will show how the **Open Resource Control Architecture (ORCA)** creates and provisions infrastructure on demand—including bandwidth and computational resources at fixed sites and in multiple clouds—to execute a scientific workflow.

Demonstration 1: Provisioning Cloud Infrastructure to Enable Alternative Energy Research

This demonstration will use ORCA to allocate a slice of computational resources and bandwidth-provisioned network connections between sites in order to execute a scientific workflow. The workflow will use seven applications that communicate using files. Six are serial applications that will run on virtual machines (Condor clusters) provisioned from clouds at RENCI (Chapel Hill, NC) and Duke University (Durham, NC). The final large MPI application will run on several thousand processors on Hopper, a Cray XE6 system at NERSC (Berkeley, CA).

Cross-continental network provisioning. The demonstration will be the first to show ORCA's ability to provision network resources across the U.S., starting with the **Breakable Experimental Network (BEN)**, which connects clouds at RENCI and Duke. Parts of the job will then use **National Lambda Rail (NLR)** and a router located in the **StarLight** facility in Chicago to connect to **ESnet**, the Energy Sciences Network, where the workflow applications will access Hopper.

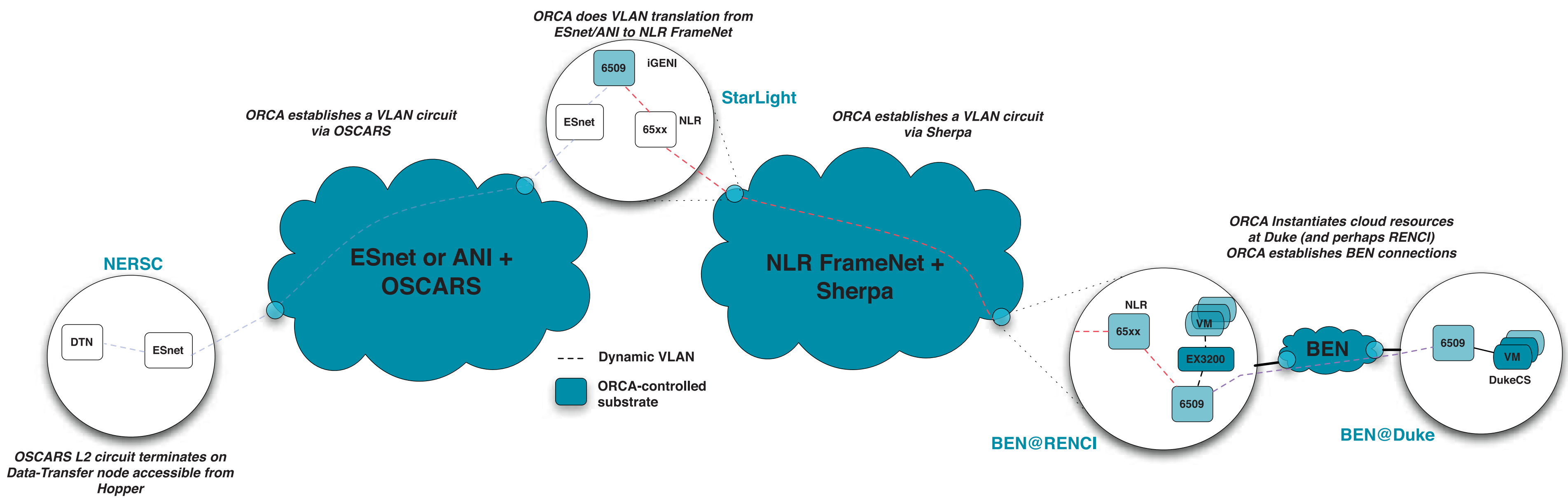
The science: new materials for solar energy. The demonstration will offer a simplified version of a workflow used to apply effective inverse-design strategies to the discovery of materials for solar energy. The electronic structure of a highly relevant Ru(IV) moiety will be computed using both ensemble-like scheduling and parallel processing concepts to compute the total energy. The multistep workflow will begin with several single core steps followed by a highly parallel final step. The work aims to build computational capability to enable research at the DOE-sponsored Solar Fuels and Next Generation Photovoltaics Energy Research Frontier Center at UNC-Chapel Hill.
This work was partially supported under U.S. DOE SciDAC-e award DE-FC02-06ER25764, "Enhancing Productivity of Materials Discovery Computations for Solar Fuels and Next Generation Photovoltaics."

Demonstration 2: Multi-cloud Hadoop Deployments for Scientific Applications

This demonstration will use the ORCA framework to execute a Hadoop workflow—with master-slave topology of various sizes—on multiple clouds connected through bandwidth-provisioned network pipelines. It will show how virtual machines at RENCI and UNC-Chapel Hill cloud sites are allocated and will provision network links to support HDFS traffic between the sites. ORCA will automatically launch Hadoop daemons on the VMs to set up a Hadoop "slice' for the user. The slice will be used to run the Hadoop-BLAST application and to run HDFS benchmarks under different provisioned configurations.

Demonstration 3: Provisioning a Condor Cluster From Multiple Clouds

This demonstration will take a closer look at a part of the first demonstration: the on-demand provisioning of computational infrastructure to stand up a Condor cluster in the cloud. The cluster will include VMs at Duke and RENCI connected over BEN. The demonstration will walk through the step-by-step process of standing up a Condor slice using ORCA and will execute a Pegasus workflow on the cluster.



Demonstration Times

Monday, Nov. 14: 7 p.m. – 9 p.m.
Tuesday, Nov. 15: 10:30 a.m. (Demo 1), 11:30 a.m. (Demo 2) and 1 p.m. (Demo 3)
Wednesday, Nov. 16: 10:30 a.m. (Demo 1), 2 p.m. (Demo 2) and 2:30 p.m. (Demo 3)
Thursday, Nov. 17: 10:30 a.m. – 12:30 p.m. (Demos 1, 2 and 3)

Research Team:

Ilia Baldine, Anirban Mandal, Paul Ruth and Yufeng Xin, RENCI
Jeff Chase, Aydan Yumerefendi Duke University

The research team wishes to express gratitude to Lavanya Ramakrishnan (NERSC), the ESnet team, National Lambda Rail, and the StarLight facility in Chicago for helping to make these demonstrations possible.

Funding: DOE Advanced Scientific Computing Research award DE-SC0005286, NSF Software Development for Cyberinfrastructure awards 1032573 and 1032873, and awards from the Global Environment for Network Innovation (GENI) Project Office . ORCA was originally developed by the New Internet Computing Lab at Duke University by Jeff Chase and his students under NSF sponsorship and is a control framework candidate for GENI.

SC11 Seattle, WA Connecting communities through HPC

renci