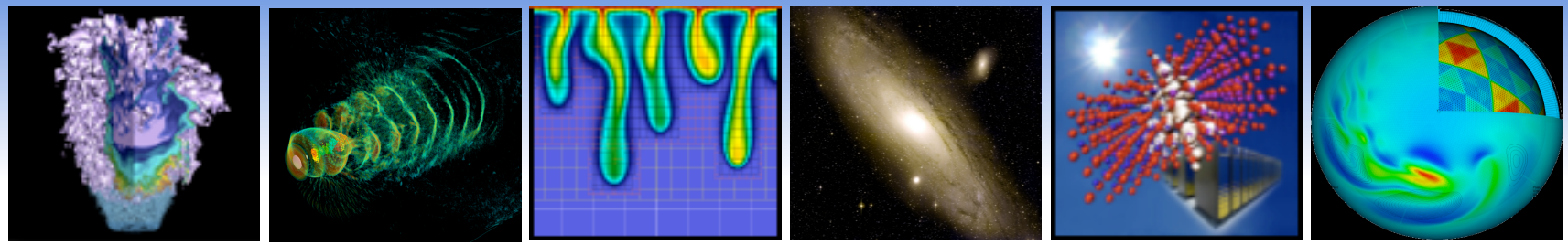


Magellan @ NERSC

Shane Canon
Lavanya Ramakrishnan
Iwona Sakrejda

HEPIX
November 5th 2010
Cornell University, Ithaca

U.S. DEPARTMENT OF ENERGY | Office of Science

NERSC

BERKELEY LAB

NERSC
NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

Magellan – Exploring Cloud Computing

- National Energy Research Scientific Computing Center (NERSC)
- Argonne Leadership Computing Facility (ALCF)
- Funded by DOE under the American Recovery and Reinvestment Act (ARRA)

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

Magellan Mission

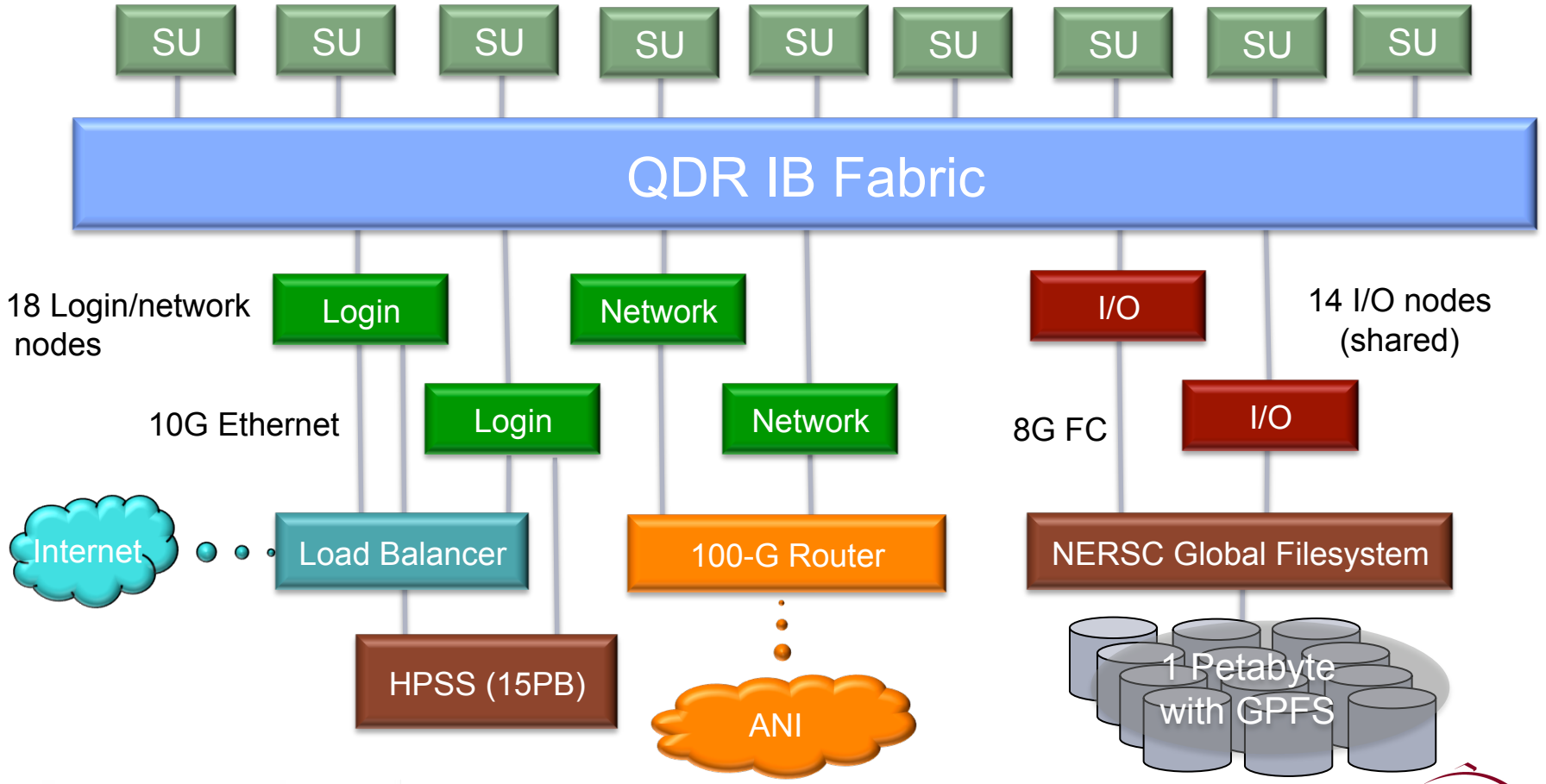
- Determine the appropriate role for commercial and/or private cloud computing for DOE/SC midrange workloads
- Deploy a test bed compute and data cloud to serve the needs of mid-range scientific computing.
- Evaluate the effectiveness of this test bed for a wide spectrum of DOE/SC applications in comparison with other platform models.

Magellan Research Agenda

- **What are the unique needs and features of a science cloud?**
  - NERSC Magellan User Survey
- **What applications can efficiently run on a cloud?**
  - Benchmarking cloud technologies *Hadoop, Eucalyptus) and platforms (Amazon EC2, Azure)
- **Are cloud computing Programming Models such as Hadoop effective for scientific applications?**
  - Experimentation with early applications such as JGI and Supernova Factory
- **Can scientific applications use a data-as-a-service or software-as-a-service model?**
  - Identifying use cases with user engagement
- **What are the security implications of user-controlled cloud images?**
  - Detailed analysis by NERSC Security Group
- **Is it practical to deploy a single logical cloud across multiple DOE sites?**
  - JGI pipeline
  - Engagement with Argonne on running multi-site Eucalyptus setup
- **What is the cost and energy efficiency of clouds?**
  - ***Future work***

Magellan Cloud
Purpose-built for Science Applications

720 nodes, 5760 cores in 9 Scalable Units (SUs) ➔ 61.9 Teraflops
SU = IBM iDataplex rack with 640 Intel Nehalem cores



NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

Current Magellan Node Allocation

Purpose	Nodes	Comments
CLUSTER RESOURCES INC.	520 (Parallel) 40 (Serial)	Mix of node types and queues. Future: Dynamic provisioning and VMs
Eucalyptus Systems	40	Can expand based on demand. Supports: VMs, block storage
hadoop	40 (SATA) 40 (SSD)	MapReduce. Both configured with HDFS
	40	Testing provisioning, new cloud stacks

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Magellan Allocations - IPM Study

- **Profiling time is available to all users as part of the Cloud Computing Performance Study**
  - **Separate allocation pool (NOT part of MPP allocation)**
  - **IPM will be turned on by default for all jobs**
  - **IPM will be used to collect several data points for each job (CPU Counters, time in MPI calls, IO)**
  - **Hope to develop "stop-light" chart of applications suitability for Cloud systems**

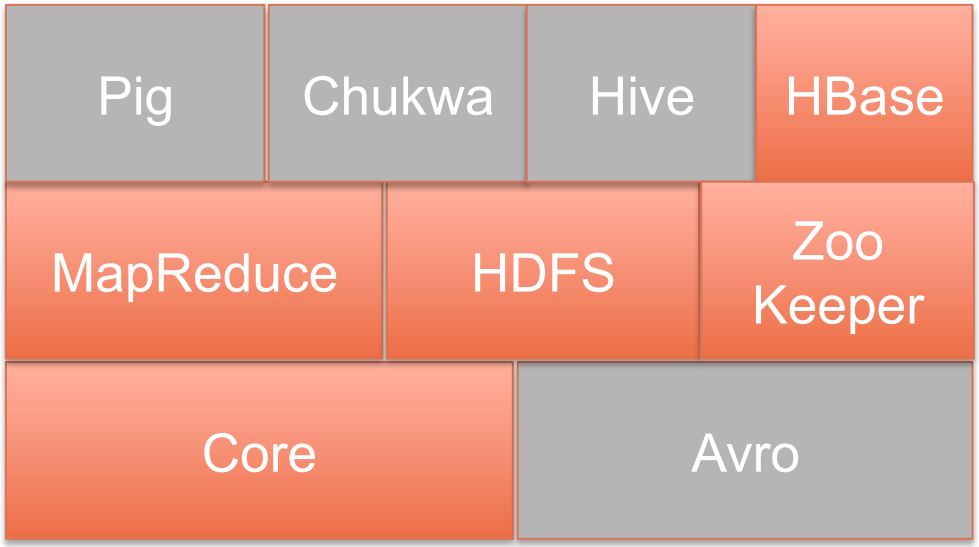
NERSC
NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

Eucalyptus

- **Open source Infrastructure as a Service implementation**
  - **API compatible with Amazon AWS**
  - **Virtual Machines, Object and Block Store**
- **Private virtual clusters**
  - **scripts to manage dynamic virtual clusters**
  - **NFS/Torque etc**
  - ***Coming soon: customized hooks for user/community extensions***

Hadoop Stack

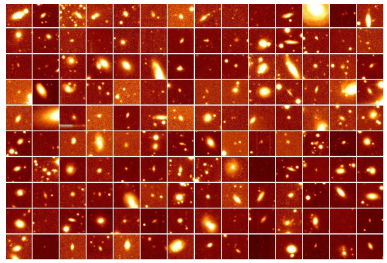
- **Open source reliable, scalable distributed computing**
  - Implementation of MapReduce
  - HDFS distributed file system
- **Number of applications**
  - DeNovo Assembly, Kbase, large databases, image analysis, etc
- ***Coming soon: Simple templates to plug in applications***

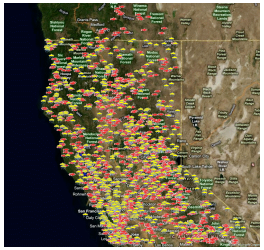


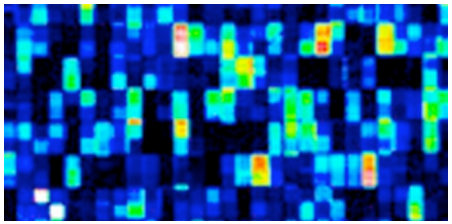
Source: Hadoop: The Definitive Guide

Attractive Features of the Cloud

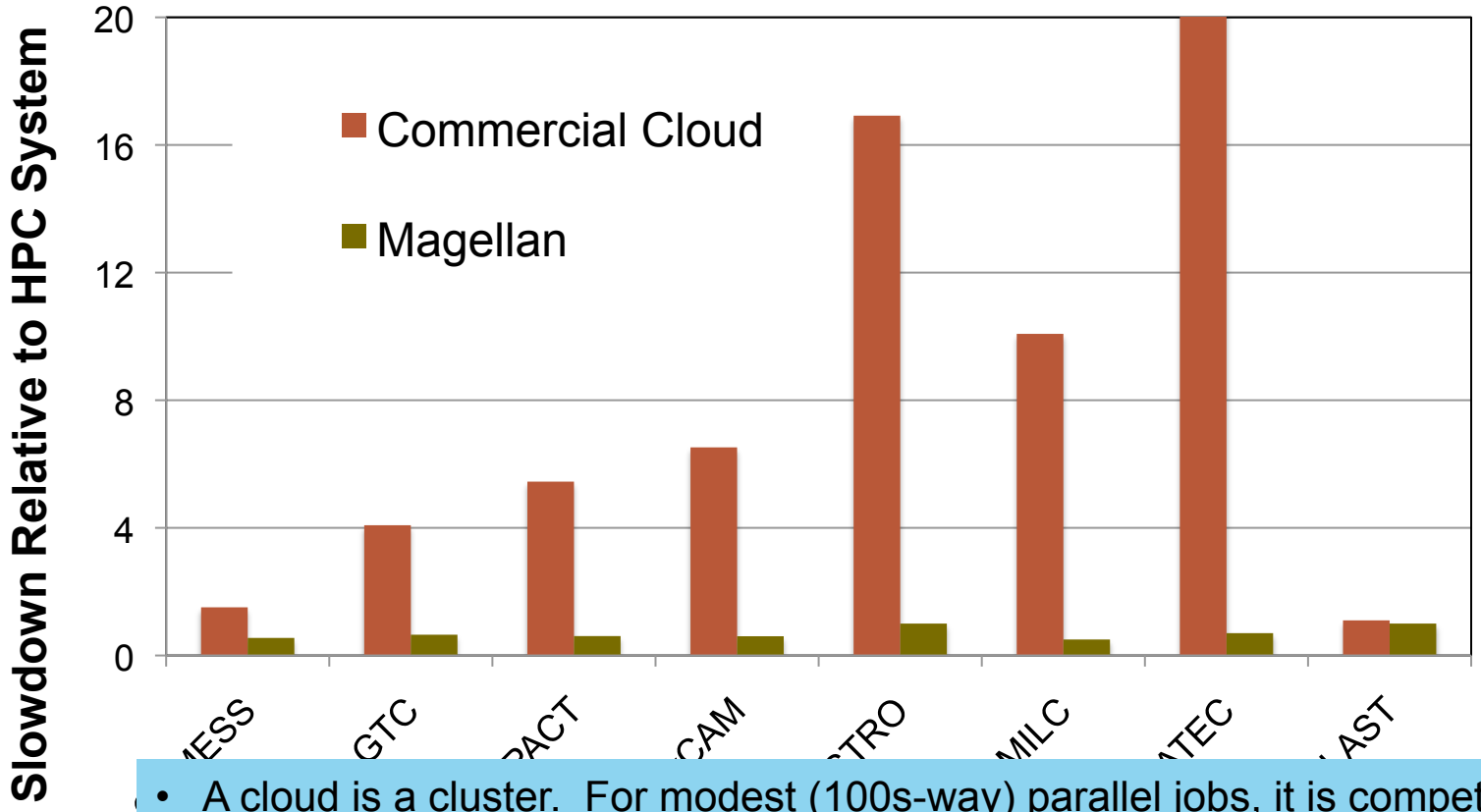
- **On-demand access to compute resources**
  - Cycles from a credit card! Avoid lengthly procurements.
- **Overflow capacity to supplement existing systems**
  - Berkeley Water Center has analysis that far exceeds the capacity of desktops
- **Customized and controlled environments**
  - Supernova Factory codes have sensitivity to OS/compiler version
- **Parallel programming models for data intensive science**
  - Hadoop (data parallel, parametric runs)
- **Science Gateways (Software as a Service)**
  - Deep Sky provides an Astrophysics community data base







Slowdown of Clouds Relative to an HPC System



- A cloud is a cluster. For modest (100s-way) parallel jobs, it is competitive with an HPC system.
- It needs a good network (e.g., Infiniband) and scheduler (batch) without virtualization

Study by Jackson, Ramakrishnan, Muriki, Canon, Cholia, Shalf, Wasserman, Wright

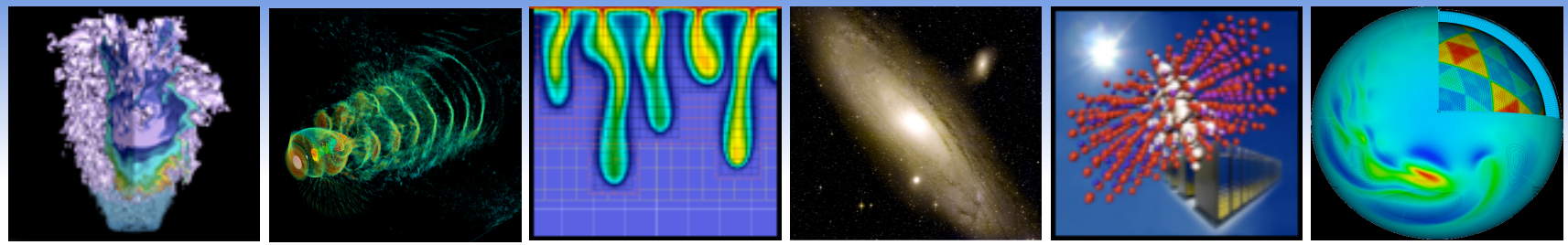
Dark Side of Clouds

- **Difficult to scale up HPC in the cloud**
  - Fine-grained / tightly-coupled MPI applications are a poor fit
  - Large scale jobs difficult to marshall
  - Long runs subject to node instability
- **Some assembly required**
  - Flexibility of Eucalyptus/EC2 comes with a price
  - Need mechanisms to distribute data and work (no batch, no parallel file system)
- **Frameworks like Hadoop can be difficult**
  - Designed to process large amount of unstructured data
  - Legacy apps can be difficult to convert
  - Not a replacement for most MPI-based alogithms

NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

Closing Remarks

- Magellan is a test bed. Help us explore!
- Try out virtual clusters, hadoop, and flash storage
- The goal is to understand the potential role of Cloud computing for DOE Science.
- We need users help to answer that question



Thank you!

Contact Info:
Shane Canon
SCanon@lbl.gov

Iwona Sakrejda
Isakrejda@lbl.gov

U.S. DEPARTMENT OF ENERGY | Office of Science

NERSC

BERKELEY LAB