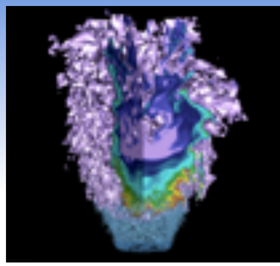
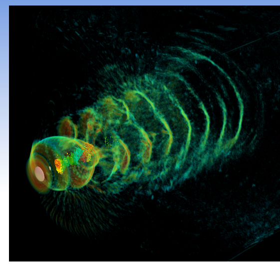
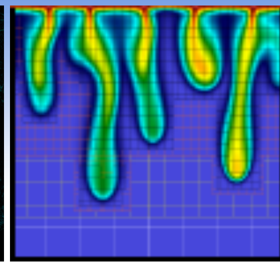
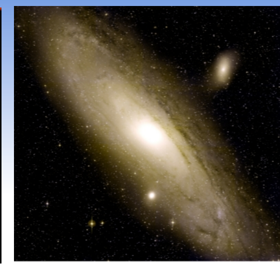
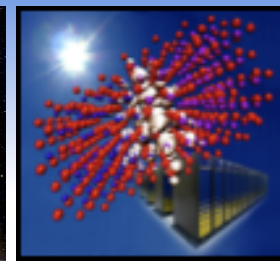
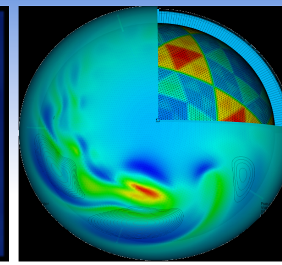
     

Evaluating Cloud Computing for Science

Lavanya Ramakrishnan

Lawrence Berkeley National Lab

June 2011

U.S. DEPARTMENT OF ENERGY | Office of Science

CRD computational research division

NERSC

Magellan Research Questions

- Are the *open source* cloud software stacks ready for DOE HPC science?
- Can DOE cyber security requirements be met within a cloud?
- How usable are cloud environments for scientific applications?
- Are the new cloud programming models useful for scientific computing?
- Can DOE HPC applications run efficiently in the cloud? What applications are suitable for clouds?
- When is it cost effective to run DOE HPC science in a cloud?
- What are the ramifications for data intensive computing?

Cloud Deployment Models

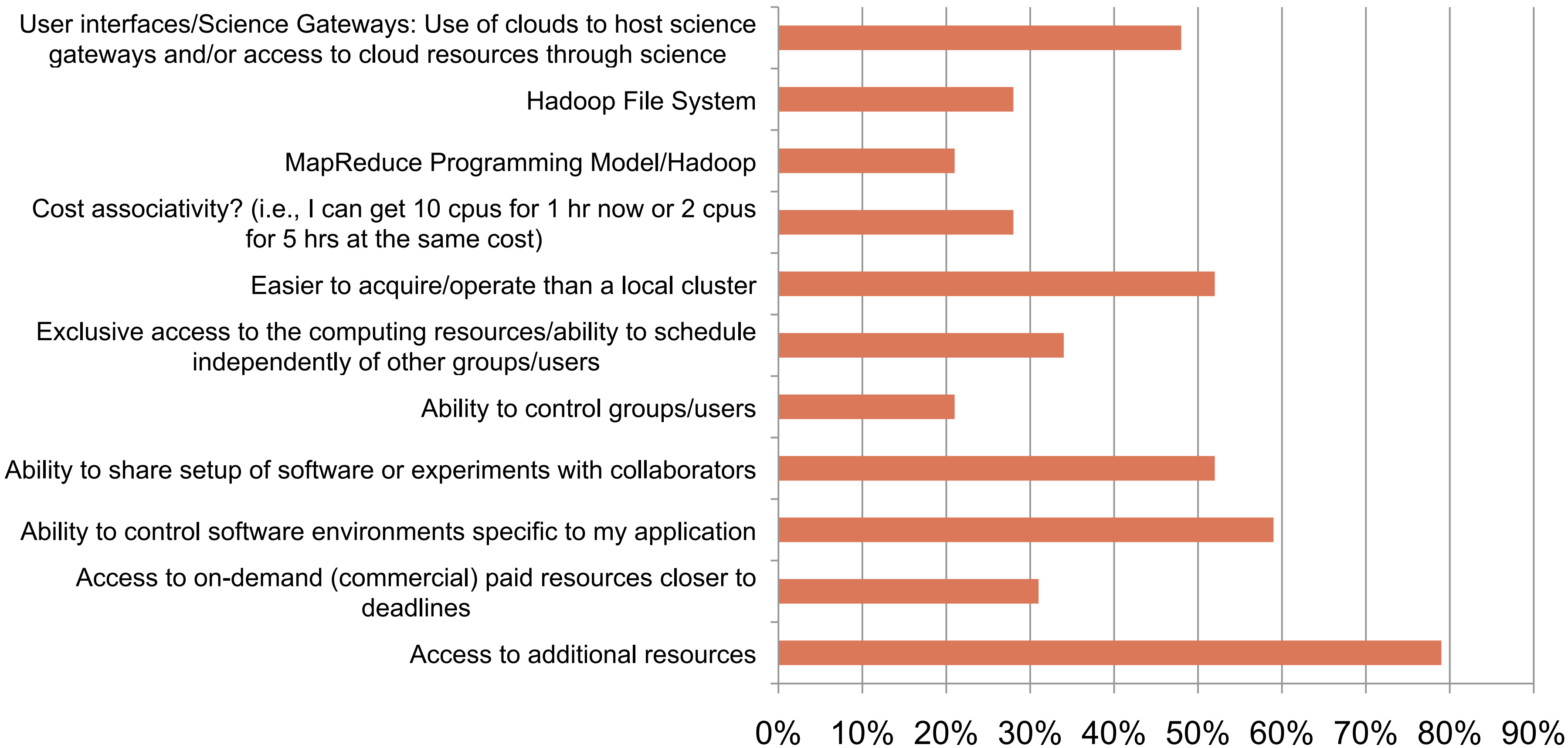
Software as a Service (SaaS)

Platform as a Service (PaaS)

Physical Resource Layer

Infrastructure as a Service (IaaS)

Magellan User Survey



Program Office	
Advanced Scientific Computing Research	17%
Biological and Environmental Research	9%
Basic Energy Sciences -Chemical Sciences	10%
Fusion Energy Sciences	10%

Program Office	
High Energy Physics	20%
Nuclear Physics	13%
Advanced Networking Initiative (ANI) Project	3%
Other	14%

Magellan Software

U.S. DEPARTMENT OF ENERGY | Office of Science

CRD computational research division

NERSC

Are the *open source* cloud software stacks ready for DOE HPC science?

Can DOE cyber security requirements be met within a cloud?

Amazon Web Services

- Web-service API to IaaS offering
- Non-persistent local disk in VM
- Simple Storage Service (S3)
  - scalable persistent object store
- Elastic Block Storage (EBS)
  - persistent, block level storage
- Offers different instance types
  - standard, micro, high-memory, high-cpu, cluster computer, cluster GPU

Private Cloud Software

- Eucalyptus
  - open source IaaS implementation, API compatible with AWS
  - KVM and Xen can be used as hypervisors
  - Walrus & Block Storage
    - interface compatible to S3 & EBS
  - experience with 1.6.2 and 2.0
- Other options
  - OpenStack, Nimbus, etc

Experiences with Eucalyptus (1.6.2)

- **Scalability**
  - VM network traffic is routed through a single node
  - limit on concurrent VMs due to messaging size
- **Requires tuning and tweaking**
  - co-exist with services such as DHCP
  - advanced Nehalem CPU instructions
- **Allocation and Accounting**
  - hard to ensure fairness since first come first serve
- **Limited Logging and Monitoring**

Security in the Cloud

- Trust issues
  - User provided images uploaded and shared
  - Root privileges by untrained users opens the door for mistakes
- Effective Intrusion Detection System (IDS) strategy challenging
  - Due to the ephemeral nature of virtual machine instances
- Fundamental threats are the same, security controls are different

Can DOE HPC applications run efficiently in the cloud? What applications are suitable for clouds?

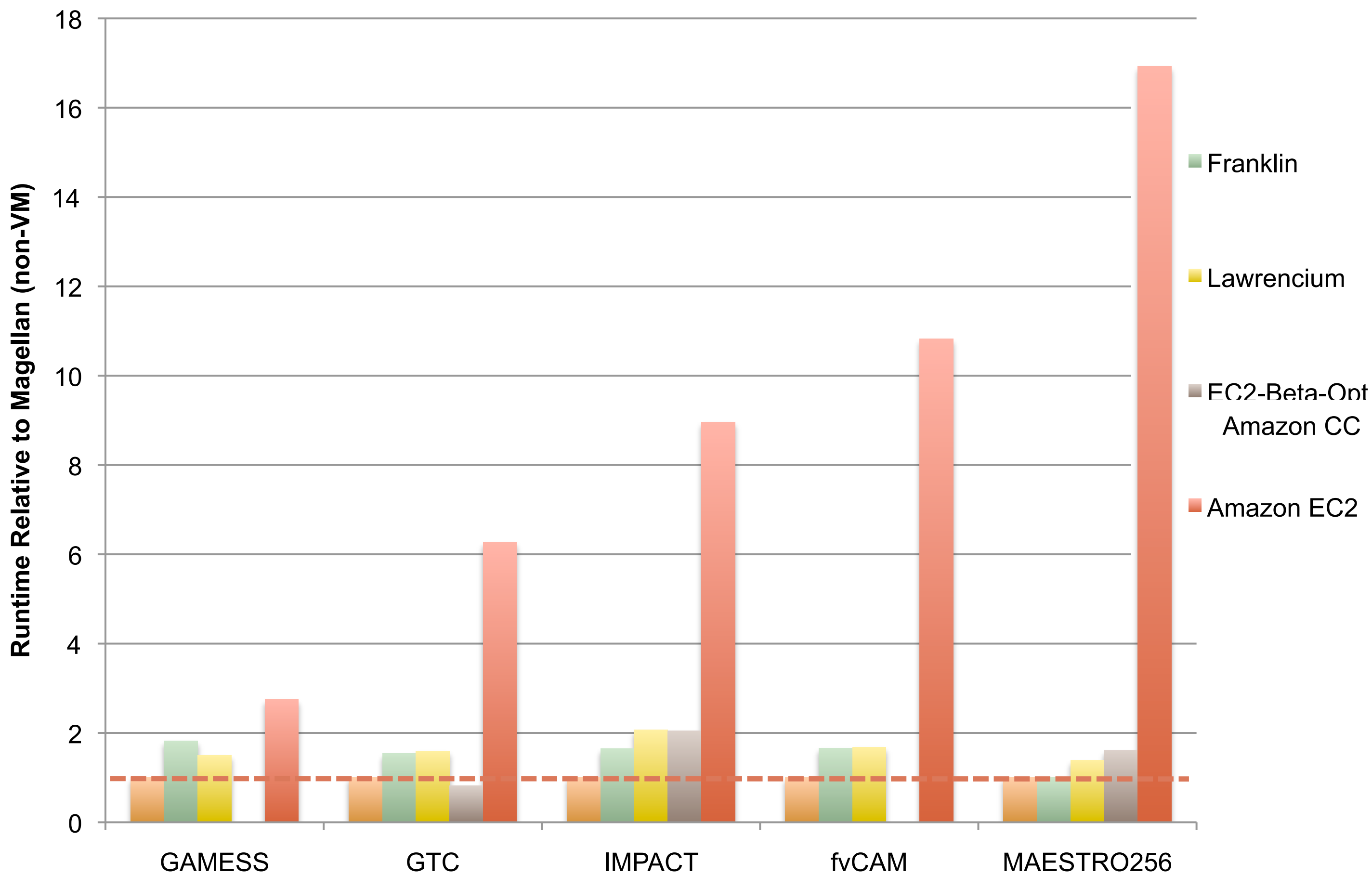
Workloads

- **High performance computing codes**
  - supported by NERSC and other supercomputing centers
- **Mid-range computing workloads**
  - that are serviced by LBL/IT Services, other local cluster environments
- **Interactive data intensive processing**
  - usually run on scientist's desktops

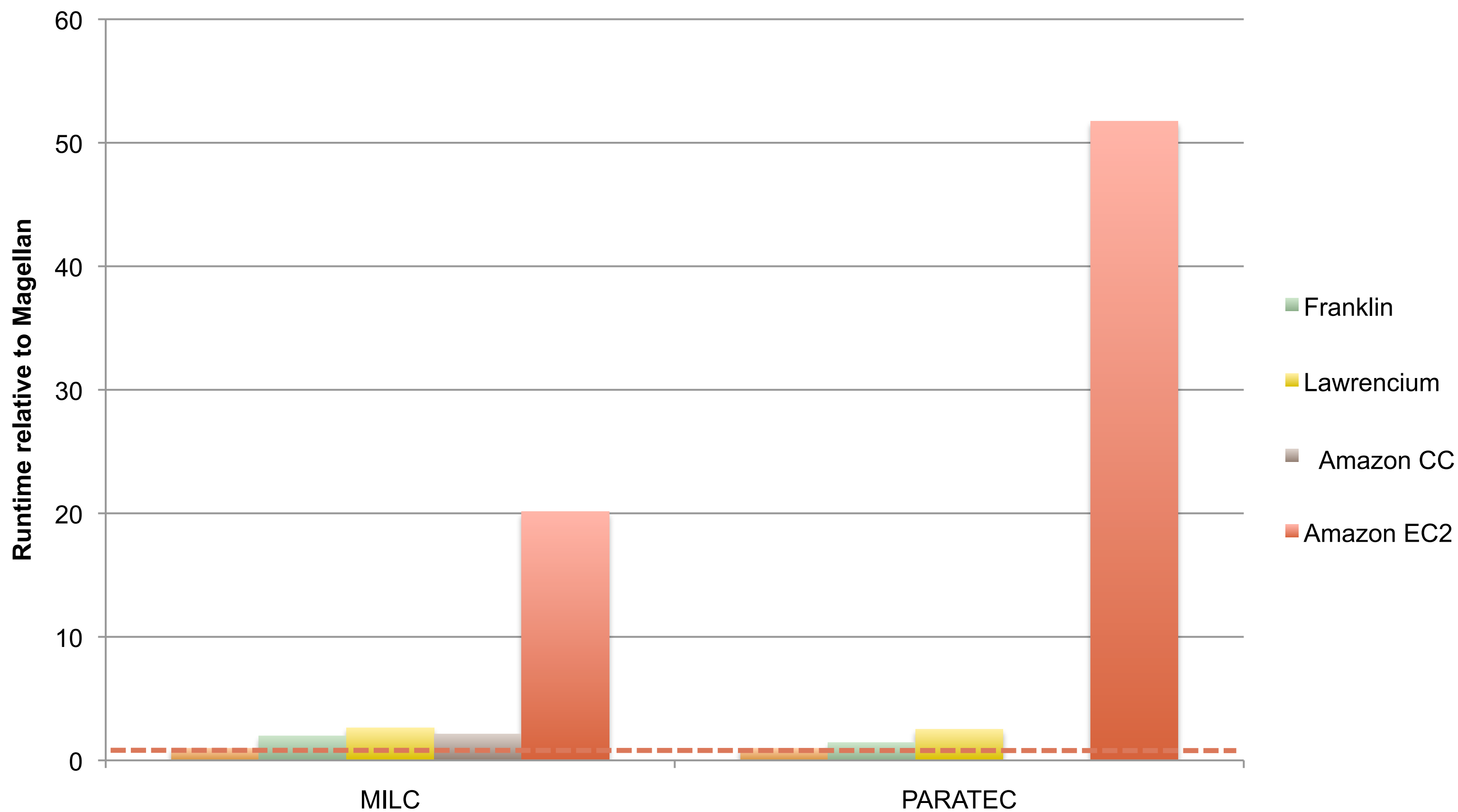
Experiment Setup

- Workloads
  - HPCC
  - Subset of NERSC-6 application benchmarks for EC2 with smaller input sizes
    - represent the requirements of the NERSC workload
    - rigorous process for selection of codes
    - workload and algorithm/science-area coverage
- Platforms
  - Amazon
  - Lawrencium (IT cluster)
  - Magellan

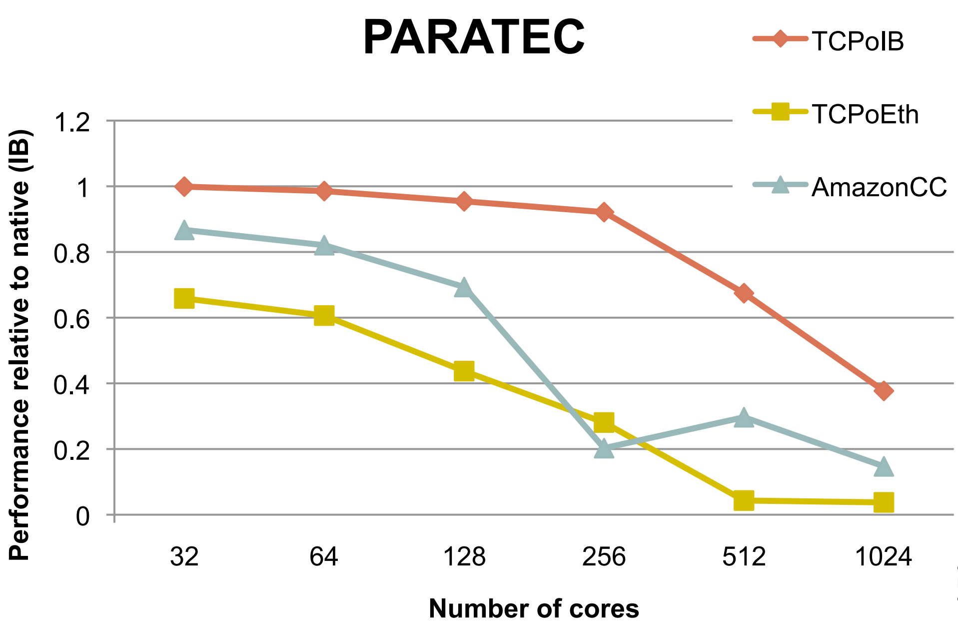
Application Benchmarks

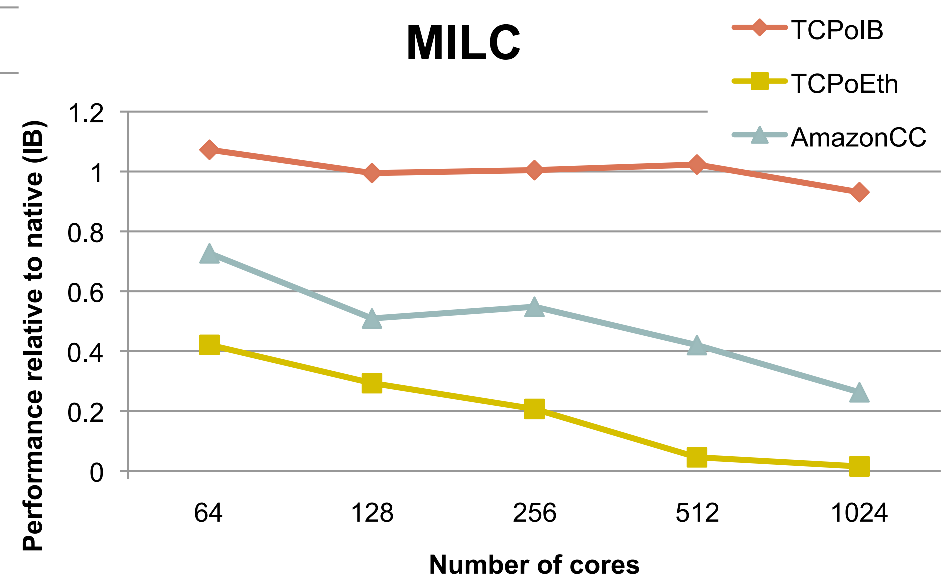


Application Benchmarks

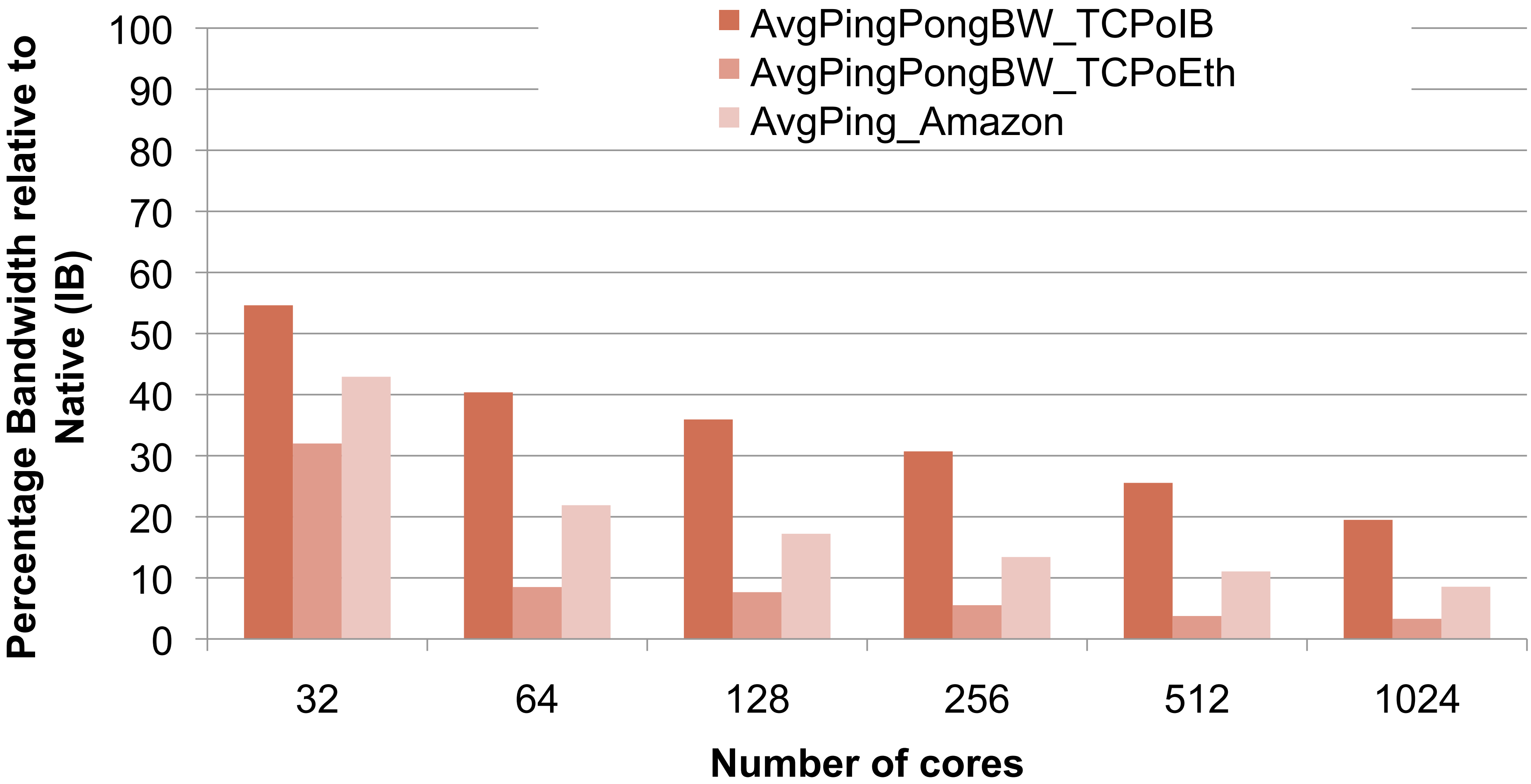


Application Scaling

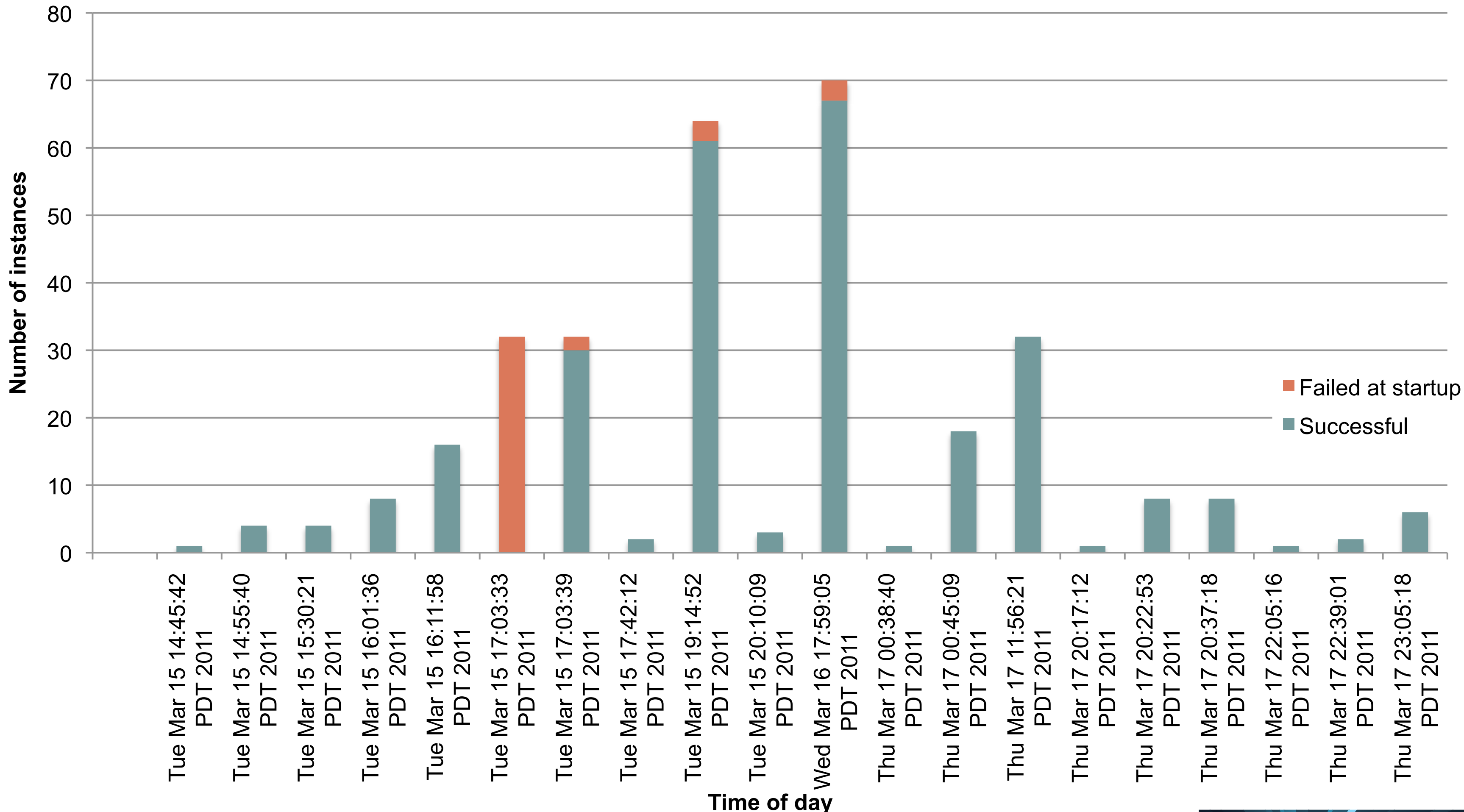




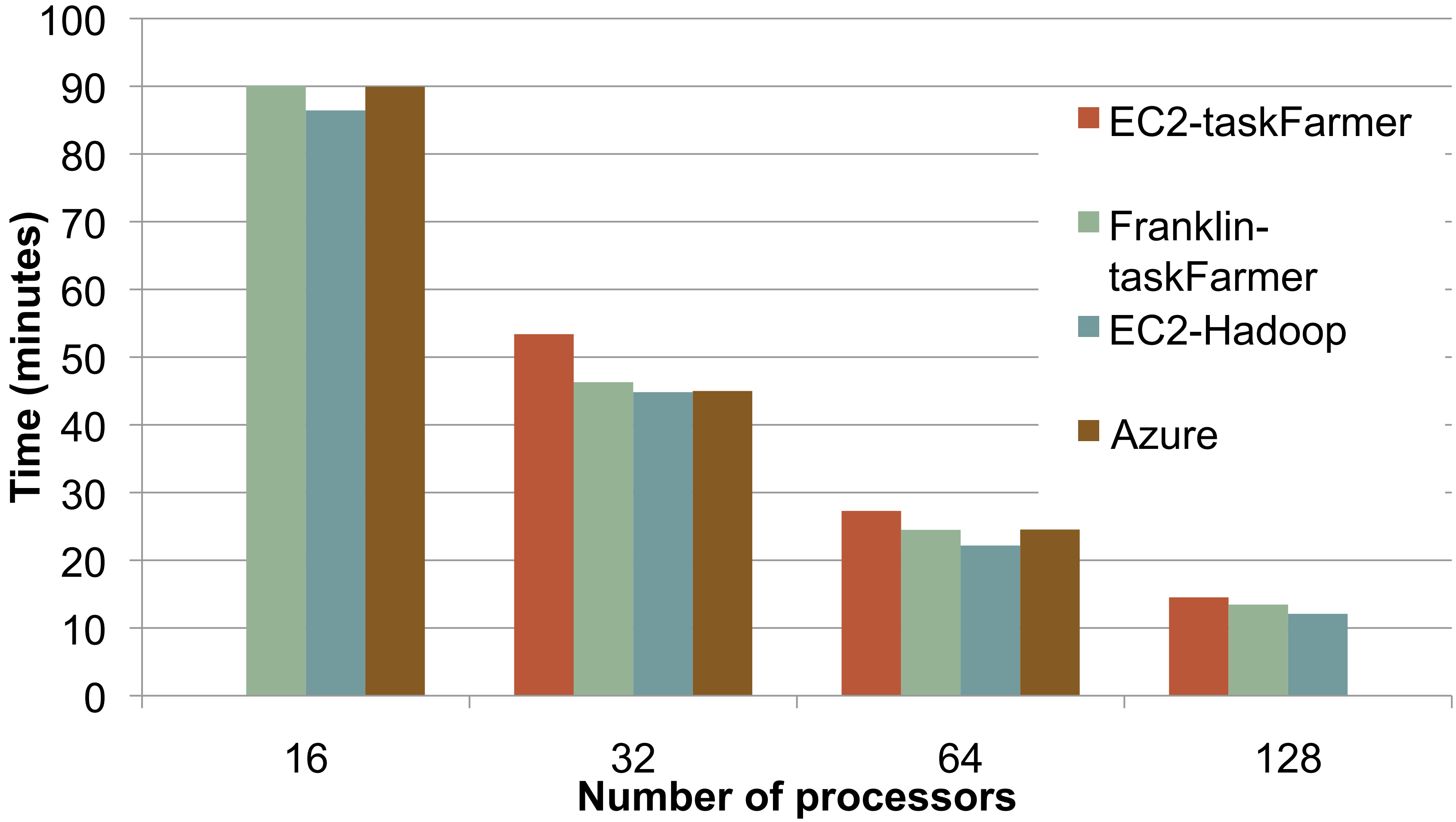
Comparison of PingPong BW



Amazon Reliability: A Snapshot



BLAST Performance

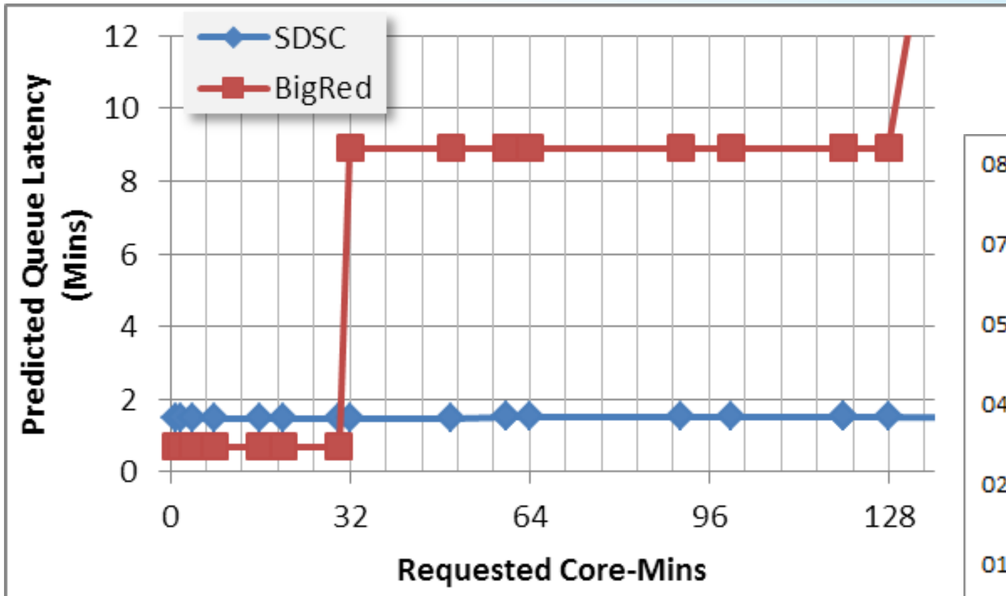


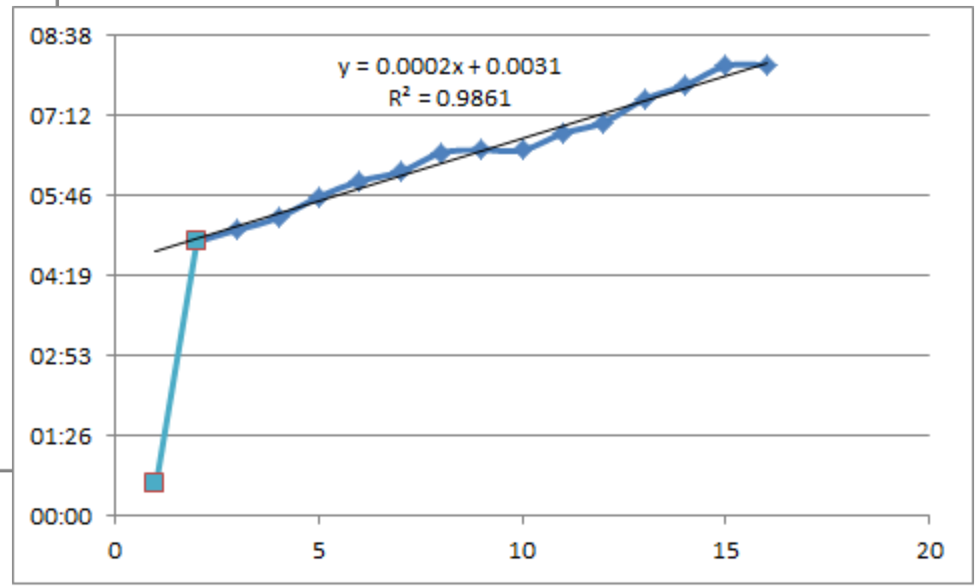
U.S. DEPARTMENT OF ENERGY | Office of Science

CRD computational research division

NERSC

Queue Wait Time Vs VM Startup Overhead





Batch queue prediction times from QBETS service on NSF TeraGrid resources, 2010

BReW: Blackbox Resource Selection for eScience Workflows
collaboration w/ Emad Soroush, Yogesh Simmhan, Deb Agarwal, Catharine van Ingen

Windows Azure VM startup time,2010
Yogesh Simmhan, Microsoft

What codes work well?

- **Minimal synchronization, modest I/O requirements**
- **Large messages or very little communication**
- **Low core counts (non-uniform execution and limited scaling)**
- **Generally applications that would do well on midrange clusters**
  - ***Future: Analyzing data from our batch queue profiling (through IPM)***

How usable are cloud environments for scientific applications?

Application Case Studies

- **Magellan has a broad set of users**
  - various domains and projects (MG-RAST, JGI, STAR, LIGO, ATLAS, Energy+)
  - various workflow styles (serial, parallel) and requirements
- **Two use cases discussed today**
  - MG-RAST - Deep Soil sequencing
  - STAR – Streamed real-time data analysis

MG-RAST: Deep Soil Analysis

Background: Genome sequencing of two soil samples pulled from two plots at the Rothamsted Research Center in the UK.

Goal: Understand impact of long-term plant influence (rhizosphere) on microbial community composition and function.

Used: 150 nodes for one week to perform one run (1/30 of work planned) and used NERSC for fault tolerance and recovery

Observations: MG-RAST application is well suited to clouds. User was already familiar with the Cloud

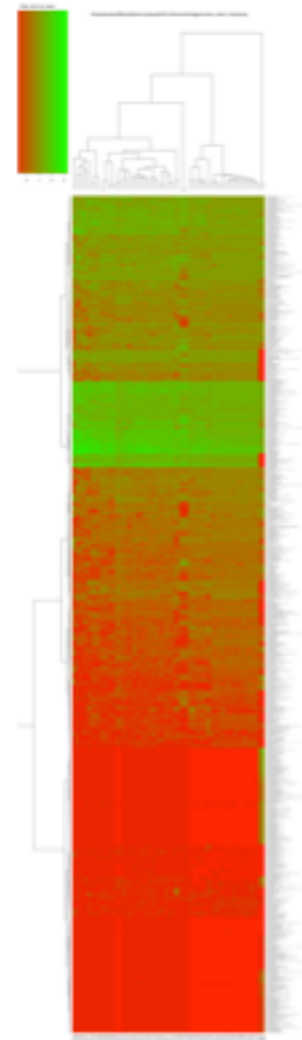
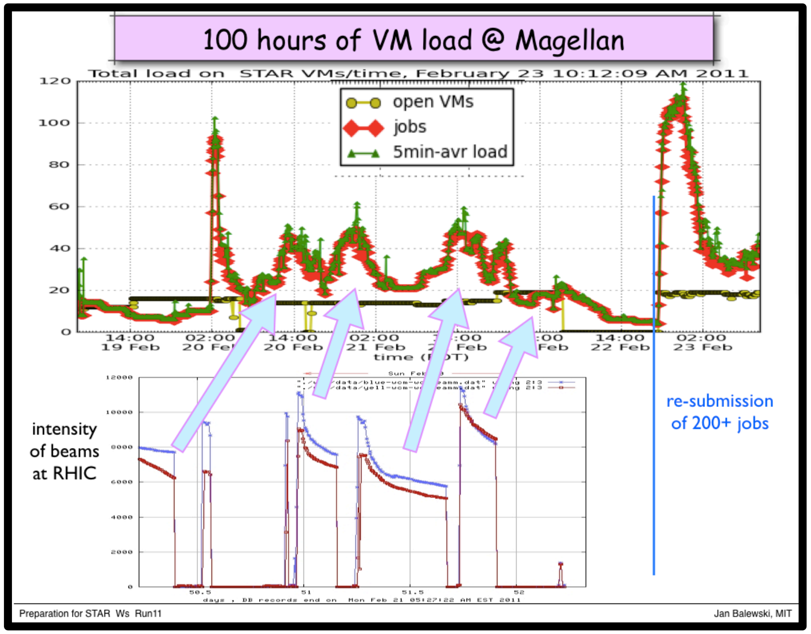


Image Courtesy: Jared Wilkening

Early Science - STAR



Details

- **STAR performed Real-time analysis of data coming from Brookhaven Nat. Lab**
- **First time data was analyzed in real-time to a high degree**
- **Leveraged existing OS image from NERSC system**
- **Started out with 20 VMs at NERSC and expanded to ANL.**

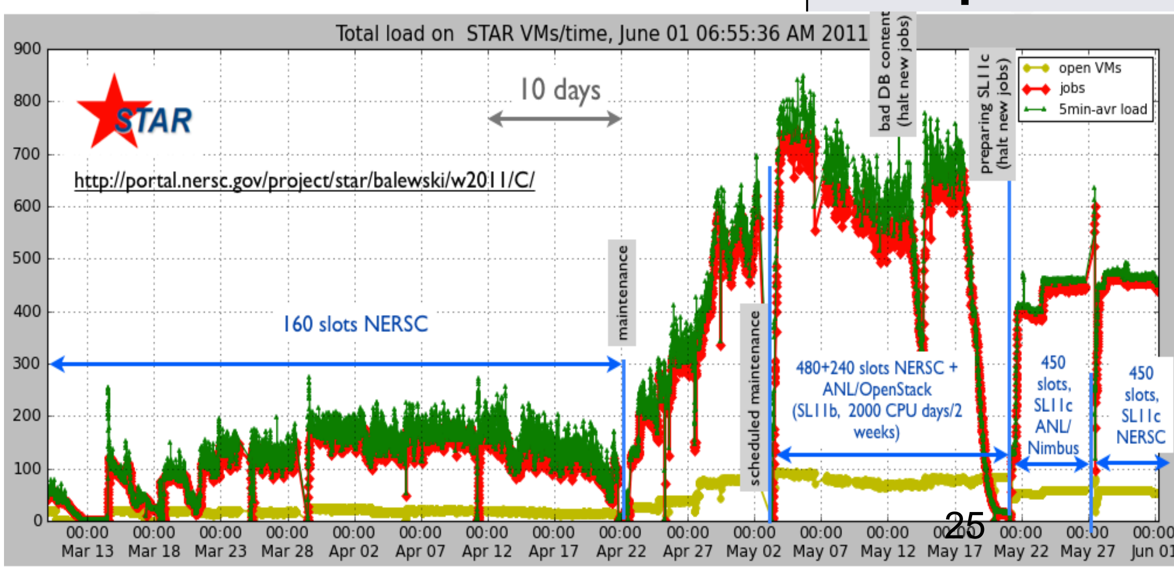


Image Courtesy:
Jan Balewski, STAR collaboration

CRD computational research division

NERSC

Application Design and Development

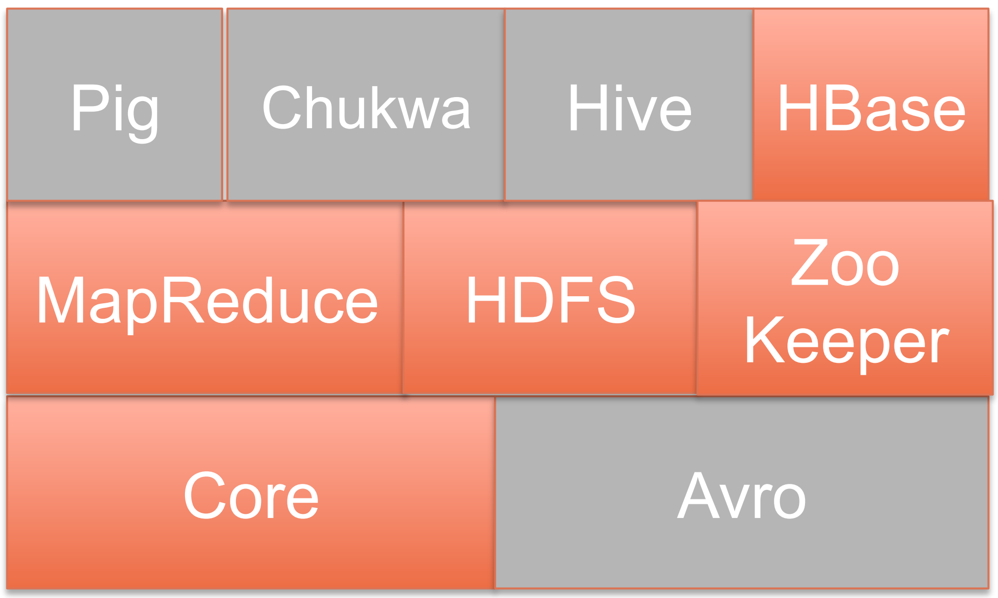
- Image creation and management
  - system administration skills
  - determining what goes on image etc
- Workflow and data management
  - need to manage job distribution and data storage, archiving explicitly
- Performance and reliability needs to be considered

Are the new cloud programming models useful for scientific computing?

What are the ramifications for data intensive computing?

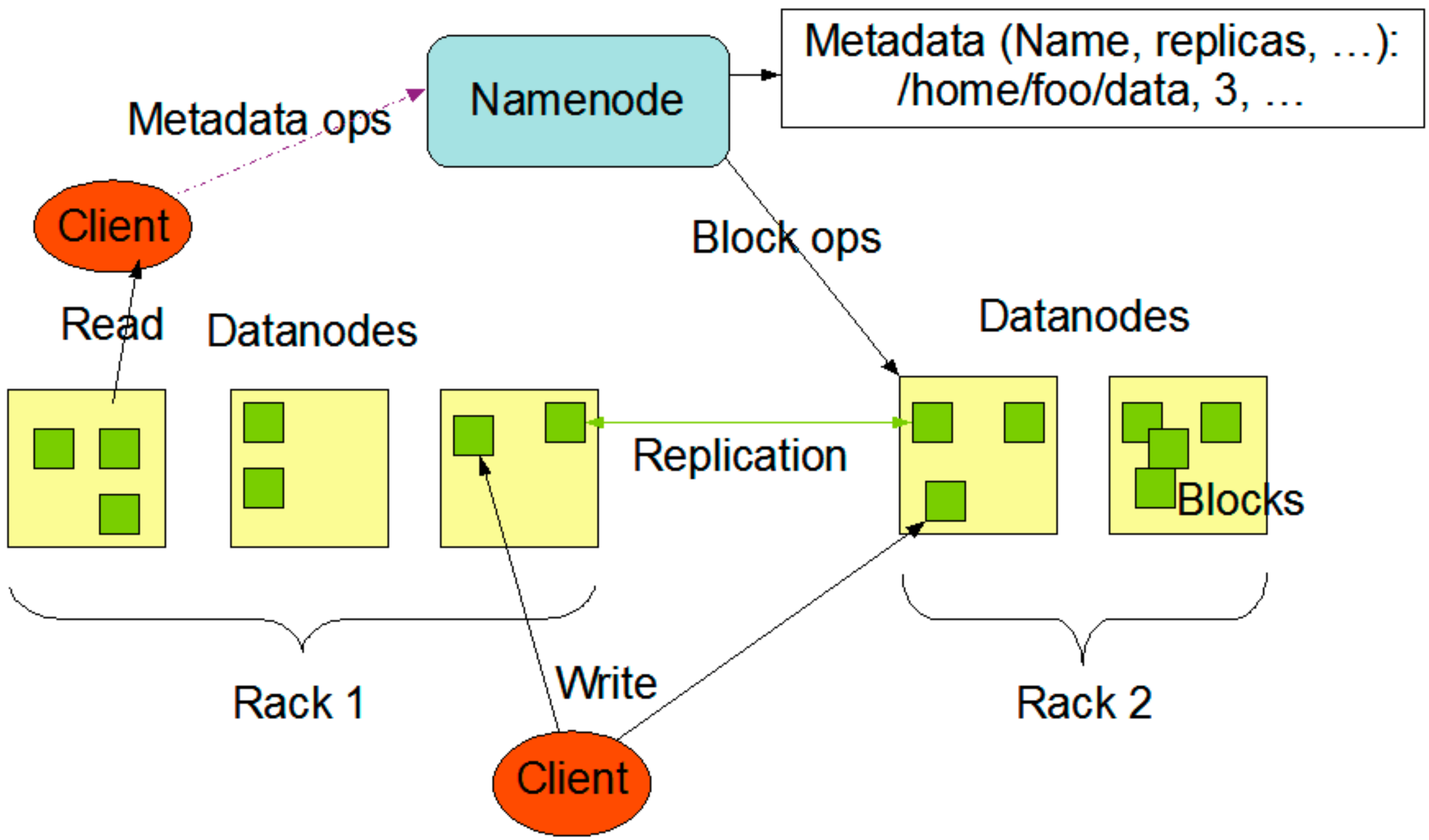
Hadoop Stack

- **Open source reliable, scalable distributed computing**
  - **implementation of MapReduce**
  - **Hadoop Distributed File System (HDFS)**



Source: Hadoop: The Definitive Guide

HDFS Architecture



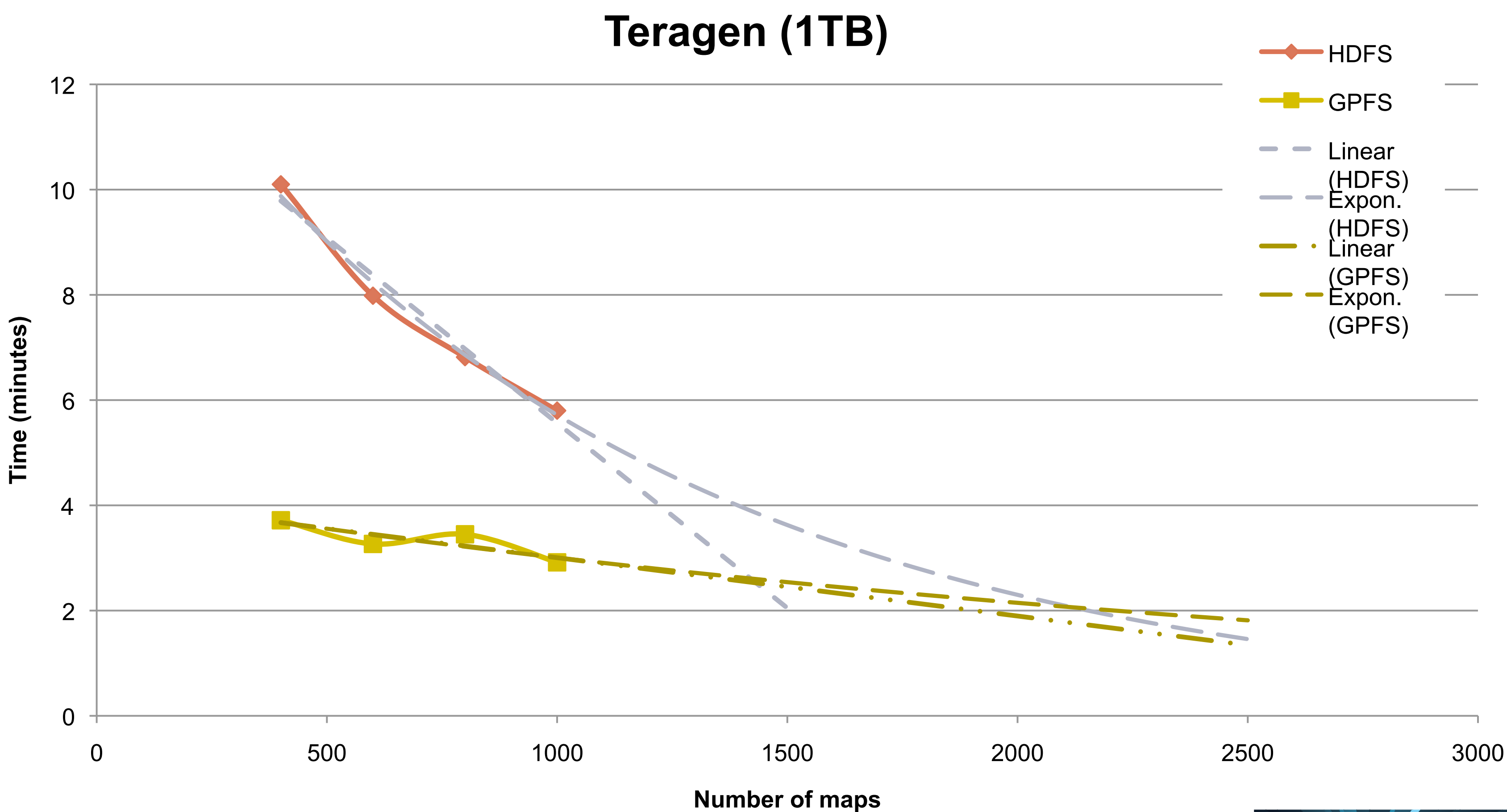
Hadoop for Science

- Advantages of Hadoop
  - transparent data replication, data locality aware scheduling
  - fault tolerance capabilities
- Hadoop Streaming
  - allows users to plug any binary as maps and reduces
  - input comes on standard input

Application Examples

- Bioinformatics applications (i.e., BLAST)
  - parallel search of input sequences
  - Managing input data format
- Tropical storm detection
  - binary file formats can't be handled in streaming
- Atmospheric River Detection
  - maps are differentiated on file and parameteC

HDFS vs GPFS (Time)



U.S. DEPARTMENT OF ENERGY | Office of Science

CRD computational research division

NERSC

Challenges

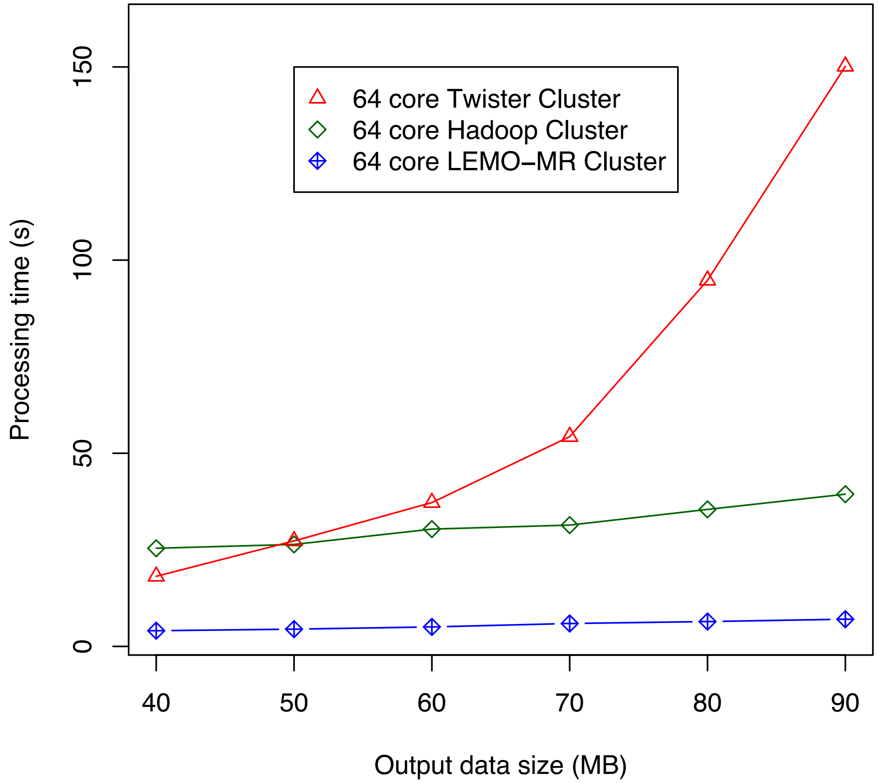
- **Deployment**
  - all jobs run as user "hadoop" affecting file permissions
  - less control on how many nodes are used - affects allocation policies
- **Programming: No turn-key solution**
  - using existing code bases, managing input formats and data
- **Additional benchmarking, tuning needed, Plug-ins for Science**

U.S. DEPARTMENT OF ENERGY | Office of Science

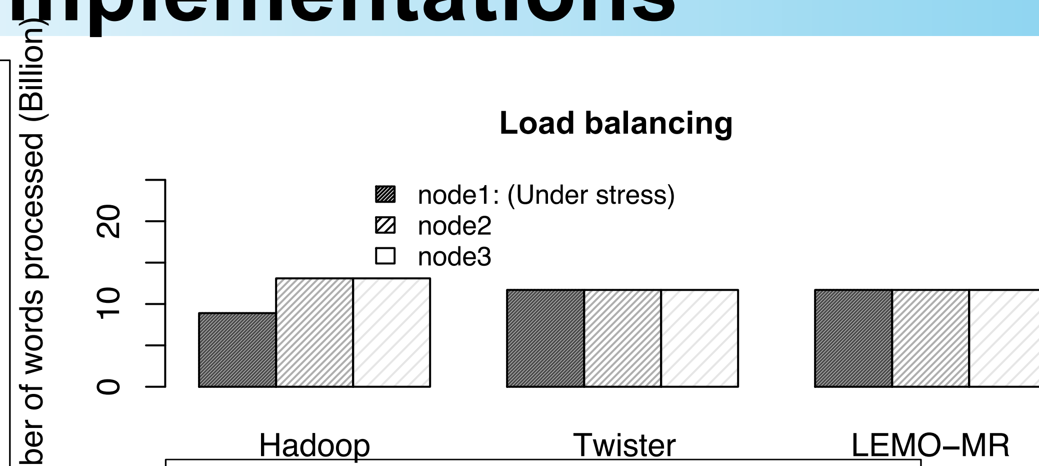
CRD computational research division

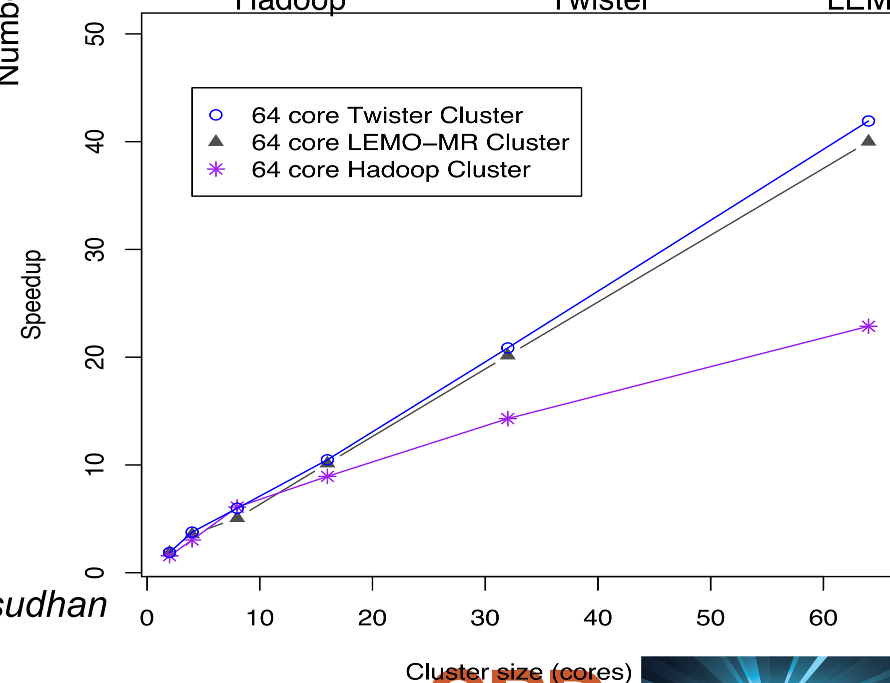
NERSC

Comparison of MapReduce Implementations



Producing random floating point numbers

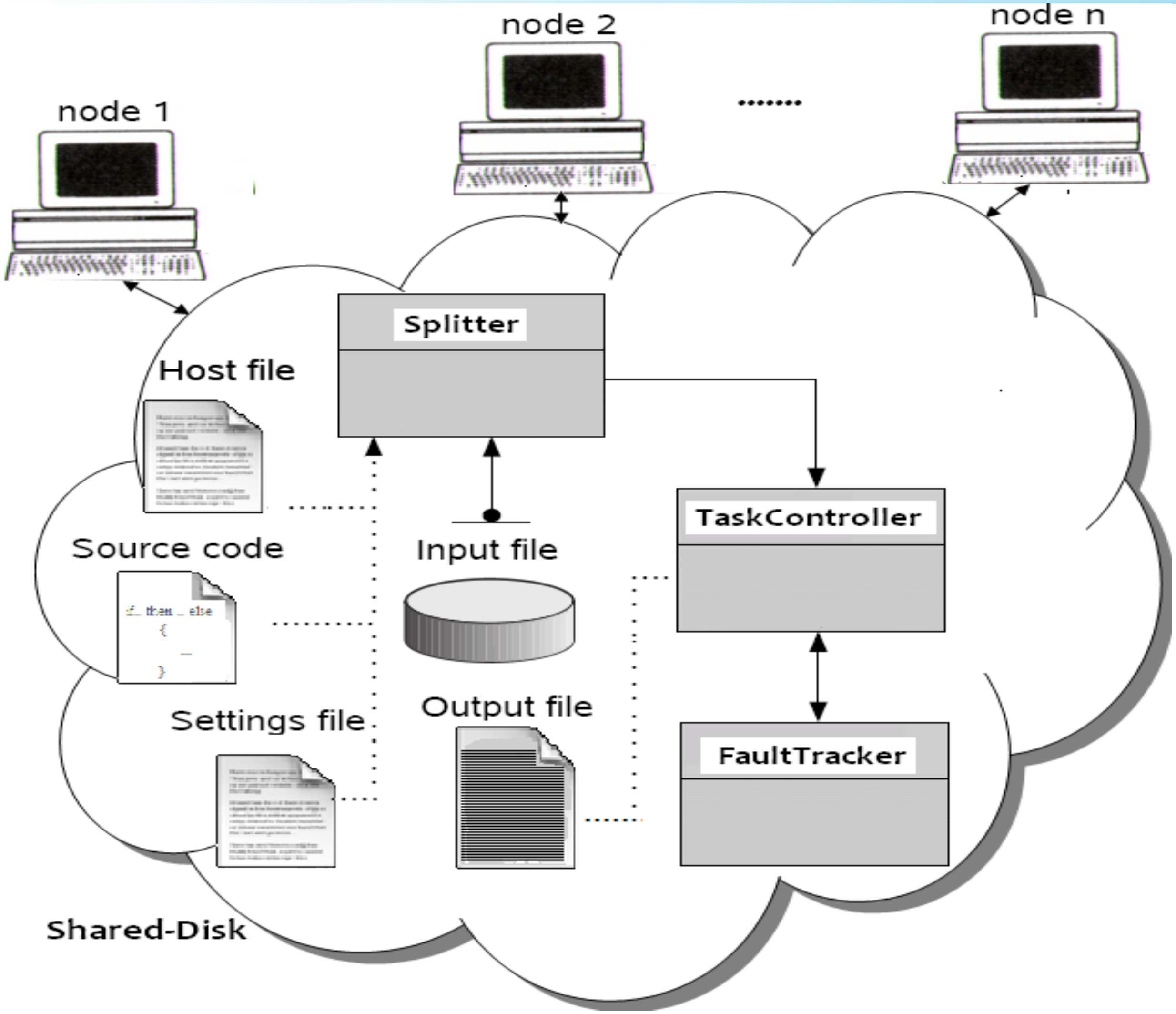




Processing 5 million 33 x 33 matrices

Collaboration w/ Zacharia Fadika, Elif Dede, Madhusudhan Govindaraju, SUNY Binghamton

MARIANE



Collaboration w/ Zacharia Fadika, Elif Dede, Madhusudhan Govindaraju, SUNY Binghamton

Data Intensive Science

- Goal: Evaluating hardware and software choices for supporting next generation data problems
- Evaluation of Hadoop
  - using mix of synthetic benchmarks and scientific applications
  - understanding application characteristics that can leverage the model
    - data operations: filter, merge, reorganization
    - compute-data ratio

(collaboration w/ Shane Canon, Nick Wright, Zacharia Fadika)

Tools for managing code ensembles/UQ

- Code ensembles
  - problem that is decomposed into a large number of loosely coupled tasks
    - Running VASP on 125K crystals in the Materials Genome database
    - Uncertainty Quantification
- Evaluate Hadoop for managing CEs
  - compare with database (MySQL, MongoDB), workflow tools, Message Queues

(collaboration w/ Dan Gunter, Elif Dede)

Unique Needs and Features of a Science Cloud

- Access to parallel filesystems and low-latency high bandwidth interconnect
  - access to legacy data sets
- Bare metal provisioning for applications that require custom environments
  - that cannot tolerate the performance hit from virtualization
- Preinstalled, pre-tuned application software stacks
  - specific libraries and performance considerations
- Customizations for site-specific policies
  - authentication, fairness
- Alternate MapReduce implementations
  - account for scientific data and analysis methods

Conclusions

- **Current day cloud computing solutions have gaps for science**
  - **performance, reliability, stability**
  - **programming models are difficult for legacy apps**
  - **security mechanisms and policies**
- **HPC centers can adopt some of the technologies and mechanisms**
  - **support for data-intensive workloads**
  - **allow custom software environments**
  - **provide different levels of service**

Acknowledgements

- US Department of Energy DE-AC02-05CH11232
- Magellan
  - Shane Canon, Tina Declerck, Iwona Sakrejda, Scott Campbell, , Brent Draney
- Amazon Benchmarking
  - Krishna Muriki, Nick Wright, John Shalf, Keith Jackson, Harvey Wasserman, Shreyas Cholia
- Magellan/ANL
  - Susan Coghlan, Piotr T Zbiegiel, Narayan Desai, Rick Bradshaw, Anping Liu, , Ed Holohan
- NERSC
  - Jeff Broughton, Kathy Yelick
- Applications
  - Jared Wilkening, Gabe West, Doug Olson, Jan Balewski, STAR collaboration, K. John Wu, Alex Sim, Prabhat, Suren Byna, Victor Markowitz

Questions?

Lavanya Ramakrishnan
LRamakrishnan@lbl.gov