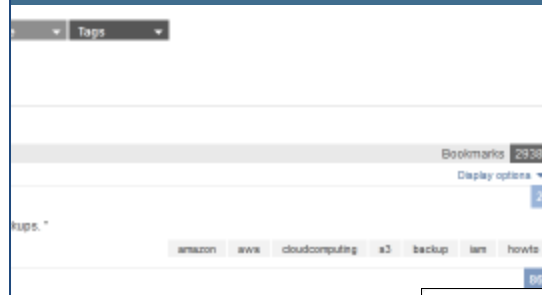
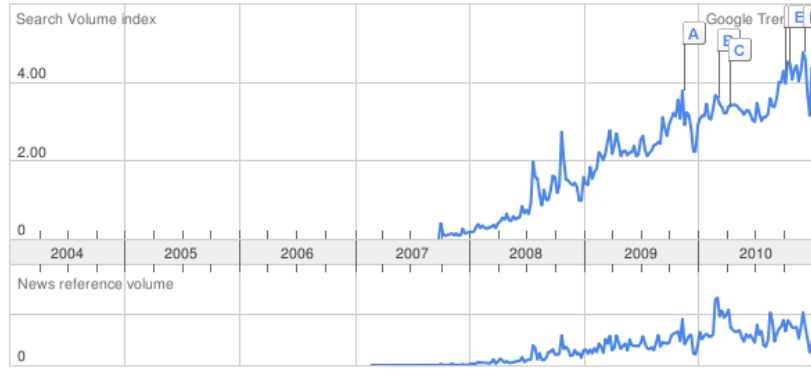


The Cloud is Suddenly Everywhere



Scale is based on the average worldwide traffic of **cloud computing** in all years. [Learn more](#)

cloud computing 1.00

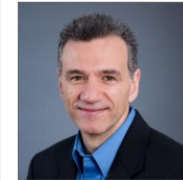


Commentary

Cutting Through U.S., EUROPE, ASIA

Vittorio Viarengo, 07/15/10, 12:40 PM EDT

The term "cloud computing" may be overused, but the cloud has lots to offer.



Palo Alto, Calif. -- "Cloud" has become a charged, and sometimes polarizing, word within enterprise IT organizations. It is the topic of many conversations and has become a priority on most chief information officer agendas, but often when we ask

people within the IT ranks whether they are using cloud computing or implementing private clouds, they tell us "no"--even many of those who are far along in their virtualization journeys.

I believe that overuse (and abuse) of the term, coupled with a

demands. The framework also takes care of scheduling, load management, and data transport, so that the genomic workflow can be executed locally available to the EC2 worker instance."

EDIT | SHARE | DELETE

HPC in the Cloud: Cloud to Improve Genomic Research at Spanish National Cancer Research Centre

"The vast potential benefits of the cloud will enable the Spanish National Cancer Research Centre to speed up its pace of innovation and bring them a current research efforts."

EDIT | SHARE | DELETE

11 JAN 11 Cloud infrastructure: Soon there'll be just one that counts | Cloud Computing - InfoWorld

"One company dominates the infrastructure services cloud space: Amazon.com through its Amazon Web Service (AWS). When you talk to other infrastructure providers, they describe their position in the market relative to AWS and even mimic the way AWS deploys its technology."

EDIT | SHARE | DELETE

Post Tech - Treasury moves to the cloud

"The U.S. Department of the Treasury is moving four existing sites into the Amazon Web Services cloud and will work with the company to host a new 4

EDIT | SHARE | DELETE

08 JAN 11 Five Under-the-Radar Amazon Web Services Cloud Features - Cloud Computing - News & Reviews - eWeek.com

"Think you know all there is to know about Amazon Web Services (AWS) and its cloud computing strategy? Think again."

EDIT | SHARE | DELETE

05 JAN 11 aantix/turkee - GitHub

"Seamlessly convert your Rails forms for use on Mechanical Turk. Then, easily import the data posted by the Mechanical Turk workers back into your data models."

EDIT | SHARE | DELETE



Current Research Challenges

There is never enough:

- Time
- Money
- CPU power
- Storage
- Physical space
- Power or cooling

Massive Funding Cuts for Higher Ed in Governor's 2011-13 (FY12 & FY13) Budget

----- Forwarded message -----

From: **Richard Coffey** <richard.coffey@phys.washington.edu>

Date: Jun 24, 2009 2:54 PM

Subject: [Astro-majors] /astro/net/safe1 is full!!!!

To: astro-all@astro.washington.edu, astro-majors@astro.washington.edu

Hi all,

Our automated messages do not apply to safe1 because we have not enabled quotas on this file system. Please clean up any unnecessary files from safe1 and stop writing file to this archive space.

```
astrolab05.astro #] df -h /astro/net/safe1
```

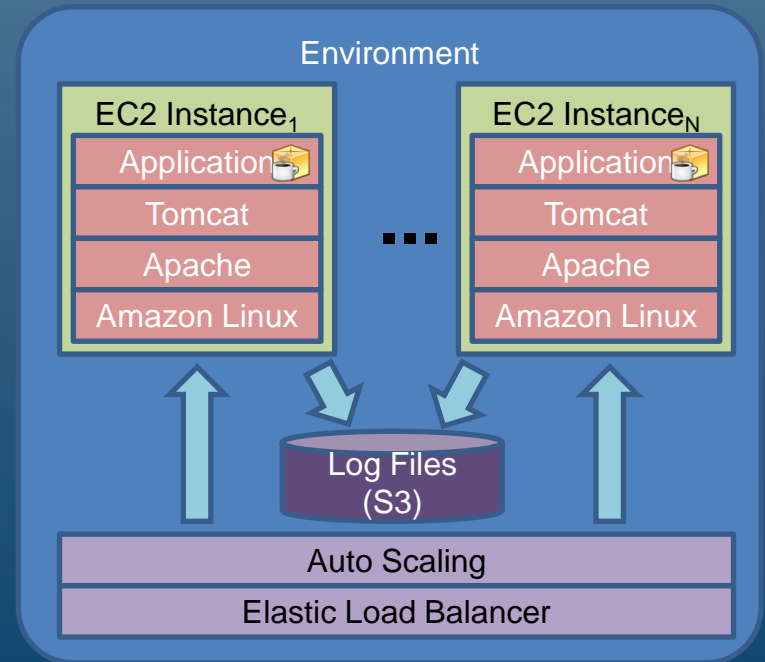
Filesystem	Size	Used	Avail	Use%	Mounted on
pexport1...	5.5T	5.5T	13G	100%	/astro/net/safe1

How the Cloud Can Help

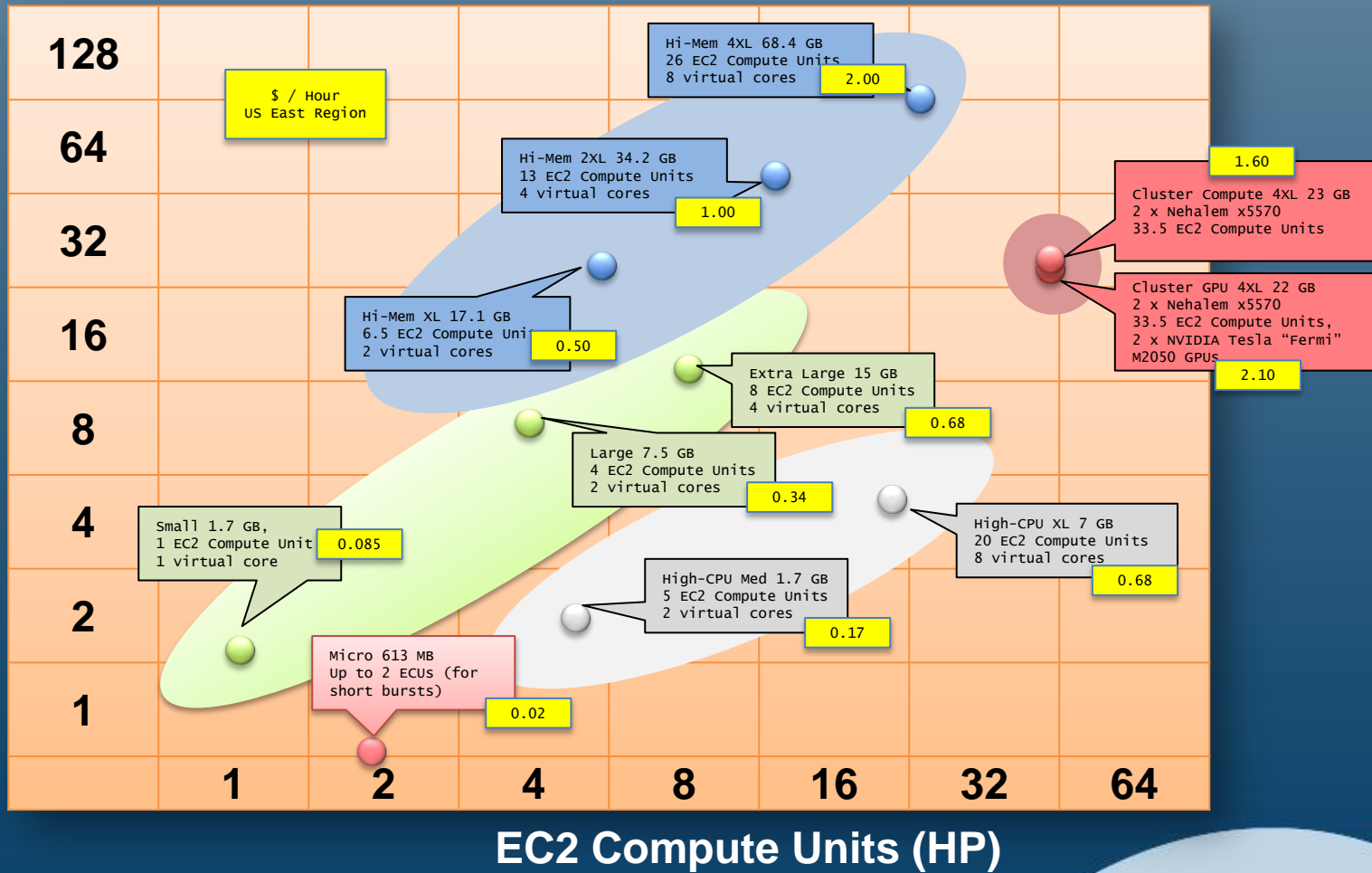
- 📦 Time
 - Less time spent on system & infrastructure issues.
 - More time for research.
- 📦 Money
 - Economical, pay-as-you-go access to resources.
 - AWS in Education program.
- 📦 CPU power
 - As little or as much as you need (0 – 10⁴ cores and beyond).
- 📦 Storage
 - GB, TB, PB on demand.
- 📦 Physical space
 - No longer “the final frontier.”
- 📦 Power or cooling
 - Our problem, not yours.

Amazon Web Services (AWS)

- 📦 Compute
 - Elastic Compute Cloud (EC2)
 - Auto Scaling + Elastic Load Balancing
 - AWS Elastic Beanstalk
- 📦 Storage
 - Simple Storage Service (S3)
 - Elastic Block Storage (EBS)
- 📦 Database
 - Relational Database Service (RDS)
 - Amazon SimpleDB
 - Third-Party offerings
- 📦 Messaging
 - Simple Queue Service
 - Simple Notification Service



Memory (GB)



AWS Management Console

The screenshot shows the AWS Management Console interface in a Mozilla Firefox browser. The main content area displays the 'My Instances' page for the 'US East' region. A table lists two instances: 'Sage Build Machine' (running) and 'Sage on T1 Micro' (stopped). Below the table, the details for the selected instance 'EC2 Instance: i-3f9d4953' are shown.

Name	Instance	AMI ID	Root Device	Type	Status	Security Groups	Key Pair Name	Monitoring	Vir
<input checked="" type="checkbox"/> Sage Build Machine	i-3f9d4953	ami-2272864b	ebs	m2.xlarge	running	BasicLinux	keys-jbarr-us-east	detailed	par
<input type="checkbox"/> Sage on T1 Micro	i-e371e8f	ami-5a46b733	ebs	t1.micro	stopped	BasicLinux	keys-jbarr-us-east	detailed	par

Description		Monitoring	Tags
AMI ID:	ami-2272864b		
Security Groups:	BasicLinux		
Status:	running		
VPC ID:	-		
Virtualization:	paravirtual		
Reservation:	r-eb676681		
Platform:	-		
Kernel ID:	aki-427d952b		
AMI Launch Index:	0		
Root Device:	/dev/sda1		
Block Devices:	/dev/sda1=vol-4c813f24:attached:2011-01-15T01:15:11.000Z:true		
Lifecycle:	normal		
Zone:	us-east-1b		
Type:	m2.xlarge		
Owner:	348414629041		
Submit ID:	-		
Placement Group:			
RAM Disk ID:	-		
Key Pair Name:	keys-jbarr-us-east		
Monitoring:	detailed		
Elastic IP:	-		
Root Device Type:	ebs		

AWS Sample Pricing

- 📦 Micro server 24x7, 10 GB disk, 15 GB in, 15 GB out
 - \$0 (AWS Free Tier for new customers)
- 📦 Small server, 24x7, 100 GB Disk, 5 GB in, 5 GB out
 - \$97.12 / month
- 📦 Large server, 24x7, 250 GB Disk, 20 GB in, 20 GB out
 - \$338.88 / month *On-Demand*
 - \$177.84 / month + \$1400 3 Year *Reserved Instance*
- 📦 8 Cluster Compute for 24 hours, 1 GB in, 1 GB out
 - \$153.60 *On-Demand*
- 📦 All prices in US East as of 2011-02-01

AWS in Education

Enable the worldwide academic community to easily leverage the benefits of Amazon Web Services for teaching and research.

- ❏ **Teaching Grants** for educators using AWS in courses (plus access to selected course content resources).
- ❏ **Research Grants** for academic researchers using AWS in their work.
- ❏ **Project Grants** for student organizations pursuing entrepreneurial endeavors; Tutorials for students that want to use AWS for self-directed learning.
- ❏ **Solutions** for university administrators looking to use cloud computing to be more efficient and cost-effective in the university's IT Infrastructure.

AWS in Education Success Stories

University of Maryland, College Park

"Thanks to the generous support of Amazon Web Services, students were provided AWS access through my teaching grant which they applied to coursework. The ability to provision Hadoop clusters on-demand gave students hands-on experience with utility computing and provided a vehicle for completing coursework and a final project."

—Jimmy Lin, Associate Professor, University of Maryland, College Park

University of California at Berkeley

"Using AWS for our Web 2.0 Application Development courses has been a phenomenal resource. Administration was so easy that students were able to get their projects deployed quickly, and venture capitalists attending the final project demos were impressed at the level of polish and creativity that a small student team could produce in just a few weeks."

—Armando Fox, Adjunct Associate Professor, University of California at Berkeley

University of California, San Diego

"The generous grant we have received from the AWS in Education program will dramatically increase the reach of my student group's collaboration with the National Federation of the Blind (NFB) to develop distributed assistive computer vision technology for the visually impaired."

—Serge Belongie, Associate Professor, University of California, San Diego



Use of Cloud Computing for Bioinformatics

A collaboration between the San Francisco State University CS Department, SFSU Center for Computing and Life Sciences and Stanford Helix Group
Jinesh Lalan, Mike Wong and Dr. Dragutin Petkovic

Use of cloud computing for bioinformatics: case study of Amazon Elastic Compute Cloud EC2

Supported in part by NIH grant LM05652, SFSU Center for Computing for Life Sciences and Amazon Educational Grant

Background

- Genomic and proteomic sequencing data have grown tremendously in size because of advancing in sequencing technology
- Consequently, analysis is intractably computationally intensive for common PC architectures and building and operating even a relatively small cluster can be a formidable undertaking.
- Recently, virtual computing clouds are drawing attentions for its flexibility since users can create a high performance cluster of any number of virtual computing units

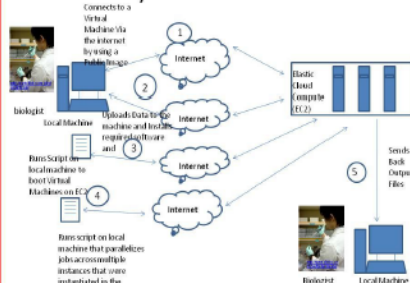
Purpose

- Investigate applicability of Amazon Cloud Computing EC2 for typical bioinformatics applications, specifically for performance and ease of use
- Provide tutorials and setup SW for non-expert users

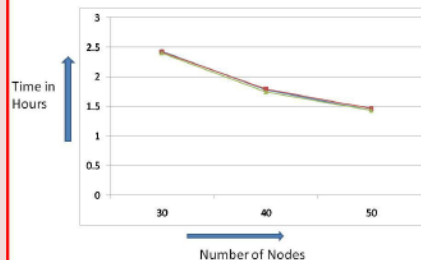
Methods

- We downloaded the Anopheles Gambiae Gene data from Vector Base.
- Installed BLAST [1] Software on the Amazon Machine Image. Already installed on the Cluster
- We ran BLAST on EC2 with 30, 40 and 50 instances and measured the run time of each. At least three runs of each number of instances were carried out. Instance type used for the research was "Small" (1.7 GB of Memory, 1-1.2 GHz Opteron or Xeon Processor, 1 virtual core)
- We also carried out BLAST runs on a local cluster of 36 cores, 152 GB of Memory and 72 GB of Disks

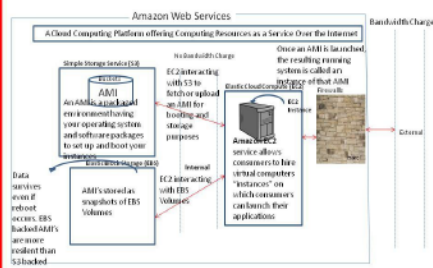
System Overview



Results



Amazon Web Services



Discussion

- Looking at the graphs, the **performance** seems to be stable and **scales linearly**.
- Noisy Neighbor Problem**. A noisy neighbor here means a Virtual Machine that uses the shared resources disproportionately. Sadly, you cannot choose your neighbors
- Network Bandwidth issues (Internal as well as external)
- Provides Simple and Easy tutorials to get started using AWS services
- Various alternative Instance types. Instance types specially designed for bioinformatics. Various instance types have different compute capacity
- Public Datasets** on AWS can be accessed with no charge

Cloud Computing Vs. Local Cluster

- Cheap** → Zero cost of equipments, physical space, cooling, power, maintenance
- Highly Scalable and Elastic** → Can scale Resources up and down as an when you want
- Your Jobs don't have to wait in the Queue
- Root access to your Nodes
- Researcher with no access to a local cluster can have their results cheaply within hours.
- Expensive** → Heavy cost of equipments, physical space, cooling, power, highly paid system admin
- Fixed Number of Resources** → Cluster Owners have to take care that the Cluster is not over or under utilized
- Your Jobs have to share the resources with other cluster users
- No root Access**

Genome Analysis Resources on AWS

- Publicly available Amazon Machine Images and Volumes for Bioinformatics
 - ami-4e57a227 -- 64 bit image
 - ami-0a91263 -- 32 bit image
 - snap-84e771ef -- Data volume [5]
- Some of the tools installed on the above AMI's :
 - PRIMER3, MUSCLE, BLAST2, EMBOSS, HMMER, READSEQ, BOWTIE, NOVAALIGN [5]
- Databases available on the above mentioned volume
 - Genome sequences pre-indexed for search with next-gen aligners like Bowtie, Novaalign
 - UniRef protein databases, indexed for searching with BLAST+ [5]

References

- NCBI/ BLAST [<http://blast.ncbi.nlm.nih.gov/Blast.cgi>]
- Amazon Web Services. [<http://aws.amazon.com>]
- EC2. [<http://aws.amazon.com/ec2/>]
- S3. [<http://aws.amazon.com/s3/>]
- CloudBioLinux [<http://www.cloudbiolinux.com/>]

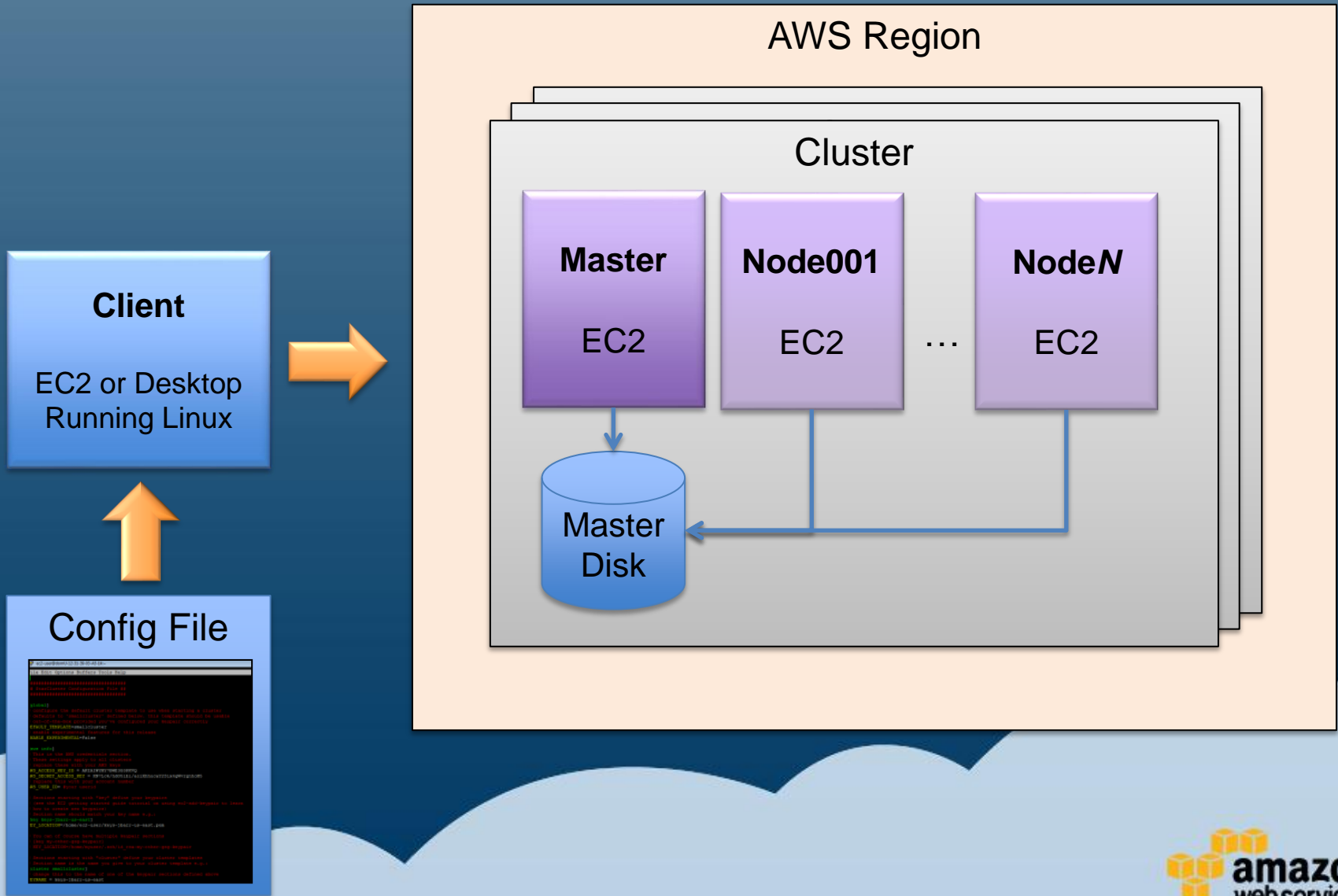
Acknowledgements

I would like to thank the gracious support and guidance of Dr. Russ Altman, Dr. Dragutin Petkovic, Mike Wong, Ljubomir Buturovic

Introducing MIT StarCluster

- ❏ Open source cluster
- ❏ Simplifies creation and management of EC2 clusters
 - OpenMPI
 - Oracle Grid Engine
- ❏ Launch a cluster of EC2 instances:
 - One command ("starcluster") to rule them all
 - Passwordless SSH pre-configured
 - Security group for SSH access
 - Shared disk volume (NFS)
 - Preinstalled libraries (NumPy, SciPy, etc.)

StarCluster Architecture / Terminology



StarCluster Commands

\$ starcluster help

\$ starcluster start jb1

\$ starcluster sshmaster jb1

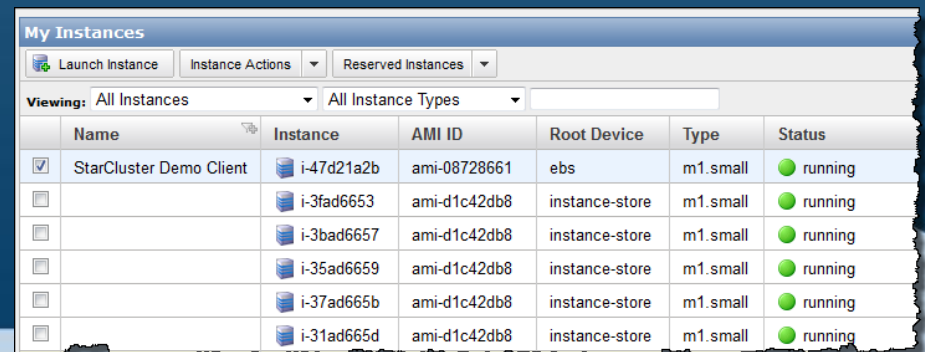
\$ starcluster listclusters

\$ starcluster stop jb1

```
<client>: starcluster listclusters
StarCluster - (http://web.mit.edu/starcluster)
Software Tools for Academics and Researchers (STAR)
Please submit bug reports to starcluster@mit.edu

-----
jb1 (security group: @sc-jb1)
-----

Launch time: 2011-01-14T05:43:44.000Z
Zone: us-east-1c
Keypair: keys-jbarr-us-east
Cluster nodes:
  master running i-3fad6653 ec2-50-16-41-160.compute-1.amazonaws.com
  node001 running i-3bad6657 ec2-184-73-107-91.compute-1.amazonaws.com
  node002 running i-35ad6659 ec2-174-129-124-218.compute-1.amazonaws.com
  node003 running i-37ad665b ec2-50-16-32-211.compute-1.amazonaws.com
  node004 running i-31ad665d ec2-50-16-31-114.compute-1.amazonaws.com
```



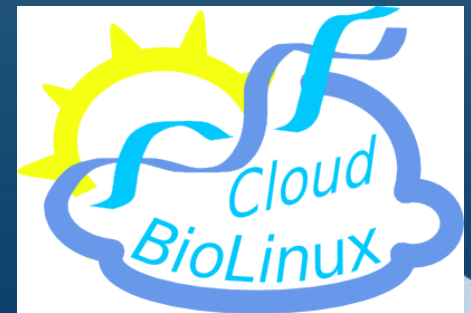
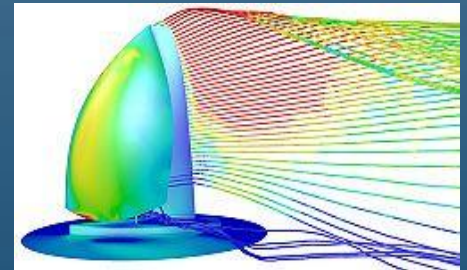
The screenshot shows the 'My Instances' page in the Amazon Management Console. It features a table with columns for Name, Instance, AMI ID, Root Device, Type, and Status. The first instance is 'StarCluster Demo Client' with ID 'i-47d21a2b'. Below it are five other instances with IDs 'i-3fad6653', 'i-3bad6657', 'i-35ad6659', 'i-37ad665b', and 'i-31ad665d', all in a 'running' state.

	Name	Instance	AMI ID	Root Device	Type	Status
<input checked="" type="checkbox"/>	StarCluster Demo Client	i-47d21a2b	ami-08728661	ebs	m1.small	running
<input type="checkbox"/>		i-3fad6653	ami-d1c42db8	instance-store	m1.small	running
<input type="checkbox"/>		i-3bad6657	ami-d1c42db8	instance-store	m1.small	running
<input type="checkbox"/>		i-35ad6659	ami-d1c42db8	instance-store	m1.small	running
<input type="checkbox"/>		i-37ad665b	ami-d1c42db8	instance-store	m1.small	running
<input type="checkbox"/>		i-31ad665d	ami-d1c42db8	instance-store	m1.small	running

AWS - HPC Use Cases

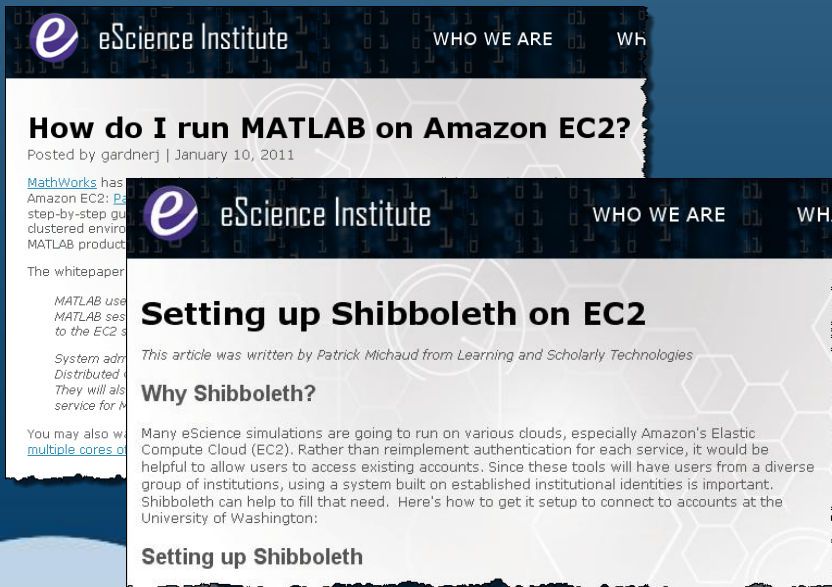
- ❏ CFD – Computational Fluid Dynamics
 - [OpenFOAM](#) on EC2
 - [CloudFlu](#)
- ❏ Molecular Modeling
 - Eli Lilly, Pfizer
- ❏ Sequence Analysis
 - [CloudBioLinux](#)
- ❏ Engineering Design
- ❏ Energy Trading & Financial Modeling
- ❏ I/O-intensive Applications
- ❏ Graphics / 3D Rendering

Open  FOAM

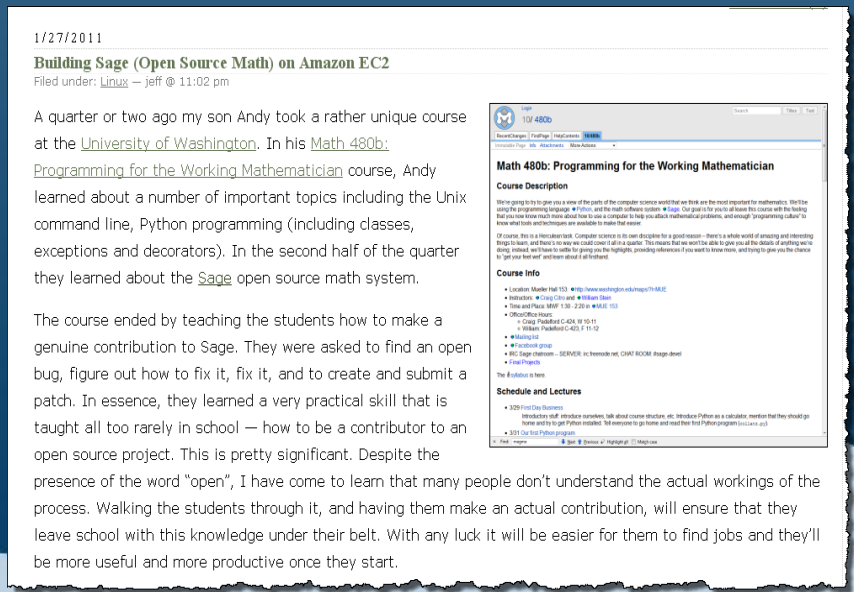


AWS at the UW

- 📦 I'm available for on-campus meetings on Wednesdays
- 📦 10 Gbit connection to AWS via CENIC/PNW Gigapop



The collage features three overlapping articles from the eScience Institute website. The top article is titled "How do I run MATLAB on Amazon EC2?" and is dated January 10, 2011. The middle article is titled "Setting up Shibboleth on EC2" and is dated 1/27/2011. The bottom article is titled "Why Shibboleth?" and is dated 1/27/2011. The eScience Institute logo and navigation links "WHO WE ARE" and "WH" are visible at the top of each article.



The screenshot shows a blog post titled "Building Sage (Open Source Math) on Amazon EC2" dated 1/27/2011. The author is identified as "Linux - jeff @ 11:02 pm". The post discusses a course at the University of Washington where students learned about the Unix command line, Python programming, and the Sage open source math system. A sidebar on the right contains a link to "Math 480b: Programming for the Working Mathematician" with a "Course Description" and "Course Info" section. The "Course Info" section lists the location (Seattle, WA), instructor (Greg Chaitin), and other details. The "Course Description" section provides a brief overview of the course content.



Getting Started

- ❏ AWS Home page: <http://aws.amazon.com>
- ❏ AWS in Education: <http://aws.amazon.com/education/>
- ❏ AWS Security: <http://aws.amazon.com/security>
- ❏ AWS Blog: <http://aws.typepad.com>
- ❏ AWS Calculator: <http://calculator.s3.amazonaws.com/calc5.html>
- ❏ StarCluster: <http://web.mit.edu/stardev/cluster/>
- ❏ Me:
 - jbarr@amazon.com
 - @jeffbarr on Twitter
 - <http://www.jeff-barr.com>