CHI

Co-organized by
Cambridge Healthtech Institute

BIOTEAM
Enabling Science

The Hilton La Jolla Torrey Pines
September 17-18, 2011

Cloud Computing Training:

Amazon Web Services (AWS)

healthtech.com/cloud

Official Publication: Bio·IT World

Today's life science organizations must deal with increasingly complex network, storage and computational requirements. Next-generation lab instruments and protocols are changing faster than the underlying research IT infrastructures built to support them. Operating an efficient, scalable and agile research IT infrastructure in the face of such rapid change is a complex challenge we all encounter.

Cloud-based Infrastructure-as-a-Service (IaaS) providers such as Amazon Web Services (AWS) offer some compelling alternatives but real answers can be hard to find amongst all the hype. BioTeam and CHI offer a unique hands-on training course covering the applied use of AWS for research, science and engineering applications.

Taught by two instructors in a limited-size classroom setting, attendees will gain a rock solid understanding of AWS and how it can best be applied to use in research applications. Training topics are covered with a split focus on enabling legacy applications while also covering in detail the emerging body of cloud best-practices and reference architectures, leaving attendees prepared for both today and tomorrow.

Training Information

Who Should Attend

Researchers, IT management, systems administrators & software developers looking for a practical & solid understanding of current AWS capabilities and how best to leverage them for research, scientific or engineering uses. By necessity the broad scope of course topics means that areas of particular interest to scientists, IT operations staff and software developers will be covered. We expect (and are tailoring the course for) a diverse attendee mix of scientists, IT professionals and software developers.

Level of Interaction

Materials are presented in a dynamic lecture format with frequent instructor-led demonstrations, discussions and highlighted examples. Recorded screen-casts may be used on topics that are difficult to orchestrate live. Attendees will have dedicated access to cloud-resident training systems along with example code, scripts and self-paced activity worksheets that can be used for individual exploration and experimentation at any time during the course. It is important to note that a conscious decision has been made to favor the inclusion of more content & topics at the expense of interactive lab exercises which can consume significant amounts of class time. Other courses are currently being prepared that are biased towards hands-on, highly-interactive training labs and exercises.

Level of Difficulty

Familiarity with Linux and shell scripting is expected. Attendees should generally be comfortable using SSH clients and operating from the Linux command line. No programming experience is required but an understanding of software development practices will help when deployment & architecture strategies are discussed. Basic familiarity with Amazon Web Services is expected (see prerequisites below) to minimize the amount of introductory materials that need to be covered. As a general rule we will use AWS-aware utilities, wrappers, command-line tools and GUIs to show and orchestrate AWS actions rather than directly manipulating the Amazon APIs. Any questions or concerns should be addressed to chris@bioteam.net.

Instructors

The class will be taught by two dedicated instructors:
- Chris Dagdigian chris@bioteam.net
- Adam Kraut kraut@bioteam.net

Prerequisites

Attendees should have wireless-capable laptops with SSH clients. The Mozilla Firefox web browser is recommended for attendees interested in exploring the various AWS-aware browser plugins & extensions. Attendees will be provided with remote access to Linux systems containing the necessary AWS software, utility, library & resource requirements needed for the course.

Attendees with Mac OS X or Linux systems should have SSH & Java 1.5 (JDK or SDK) installed and available if they want to locally install the AWS command-line utilities on their machines. This is not required. Attendees may also wish to have personal accounts set up with Amazon Web Services (http://aws.amazon.com). This is not required for training as shared credentials belonging to BioTeam will be used for exercises, labs and demonstrations.

Attendees new to Amazon Web Services are encouraged to follow the Amazon self-paced "Getting Started With EC2" tutorial online (http://docs.amazonwebservices.com/AWSEC2/latest/GettingStartedGuide/) prior to attending the class. The tutorial can be completed in a short time and presents an excellent introduction to the core EC2 service.

Comments or questions can be addressed directly to Chris Dagdigian chris@bioteam.net

Course-at-a-Glance

Saturday, September 17

9:00 am	Registration and Coffee
8:30am - 9:00	Registration and Coffee
9:00-12:00pm	General Session
12:00-1:00	Lunch
1:00-5:00	General Session
5:00-6:00	Optional Self-Paced Learning and One-on-One Mentoring

Sunday, September 18

8:30am	Coffee
9:00-1:00pm	General Session
1:00	Course Ends

Saturday, September 17

Objective: Progress iteratively through the topics essential for building out larger or more production-focused workstreams on the AWS platform. Day One will focus on the basic foundations and will use the Maq assembler as an example use case for building out a more traditional (or 'legacy') workflow on Amazon AWS.

I. Intro & Logistics

II. AWS Overview

There are a huge number of AWS service and product offerings. We'll cover the ones most of interest to people involved in informatics and high performance computing.

III. Mapping Informatics to the Cloud

Cover the major environmental, performance and architecture differences between HPC, grid and cluster environments and the AWS cloud environment.

IV. AWS: Billing & Credential Management

Briefly cover the logistics and mechanisms behind organizational billing and credential management with focus on the newly announced AWS 'consolidated billing' offering.

V. AWS: EC2 Overview

Light introduction to Amazon EC2 to cover definitions & capabilities before we start making heavy use of EC2 instances in live demos and recorded screen-casts.

VI. AWS: Configuration Management

Configuration management of EC2 AMIs is a major component in deploying cloud applications in a reliable, repeatable and easy to manage process. For this topic, we will be using Chef Server (http://www.opscode.com/chef/) to demonstrate configuration management of cloud-based server AMIs.

VII. AWS: Identity Management

There are some cases where individual access via SSH keys may not be sufficient (such as with web applications). Topic will be covered with a demonstration of either LDAP server integration or OpenID integration.

VIII. AWS: Monitoring & Reporting

Discuss and demonstrate a number of different monitoring & reporting options. Specific focus on Amazon Cloudwatch (AWS product offering), Server Density (commercial solution from www.serverdensity.com), Hyperic HQ Open Source Edition (open source solution from www.hyperic.com) and SyslogNG (www.balabit.com) for logfile consolidation.

IX. Putting it all together: Maq Assembler

Using the Maq assembler algorithm as our demonstration use-case, we will discuss and show several different legacy deployment methods utilizing Amazon Web Services. The "legacy" methods are for supporting existing applications and workstreams that may have been built for HPC clusters and compute farms. Day One will showcase the "legacy" methods while Day Two will showcase a more traditional cloud architecture using current AWS best practices.

Time permitting, we will discuss and demonstrate each of the following:

• "Cloud Bursting" - Local persistent servers capable of harnessing EC2 nodes on-demand. Demo may involve commercial products from www.univa.com ("UnivaCluster & UniCloud") or Sun Microsystems (Grid Engine SDM & Cloud Services Adapter).

• Standalone large Grid Engine instance - Single server solution for completing a particular workload. Grid Engine will be used as the task/job scheduler.

• Self-assembling Grid Engine clusters - Multi-node self organizing Grid Engine clusters within AWS. Of particular interest for development, testing and supporting of legacy HPC workstreams and applications

X. Wrap-up & Discussion

Discuss and review the topics of the day with particular focus on identifying attendee interest in areas that were not covered or were not covered enough. Time is being left open in the "Day Two" schedule to handle inclusion of additional topics or demonstrations.

Sunday, September 18

Objective: Continue progressing iteratively through the topics essential for building out larger or more production-focused workstreams on the AWS platform. Day Two will focus on continued use of the Maq assembler as our example use case. The focus today will be on architecting solutions using current AWS products and best practices. Due to estimated session size, this training will be lecture, discussion and live demo driven.

I. Intro & Logistics

II. AWS: S3 Overview

Coverage of the object-based AWS storage service.

III. AWS: EBS Overview

Coverage of the block-based AWS storage service.

IV. AWS: Data Movement

Data movement in and out of "the cloud" is problematic for data heavy fields like life science informatics. Cover known issues, alternatives such as the Amazon physical ingest/outgest services and where to "draw the line."

V. AWS: SQS Overview

Review and demonstrate the AWS SQS service, often a central component of cloud-resident workstreams.

VI. AWS: Additional Topics

Placeholder topic for areas identified during Day One as needing more depth, discussion or demonstrations.

VII. Putting it all together: Maq Assembly Revisited

Review the "legacy" Maq solutions shown in Day One and discuss the pros and cons of those approaches. Continue on with discussion of current-day AWS best practices culminating in a revised/revisited Maq demonstration using more traditional cloud workflow methods.

VIII. Wrap-up & Discussion

Hotel & Registration Information

The Hilton La Jolla Torrey Pines
10950 North Torrey Pines Road
La Jolla, CA 92037
Tel: 858-558-1500

CHI has a block of guest rooms at a reduced rate of $179. Please call the hotel directly to reserve your sleeping accommodations. Identify yourself as a Cambridge Healthtech Institute conference attendee to receive the reduced room rate. The cutoff date for the block will be August 22, 2011, and afterwards registrations will be accepted based on availability.

For Directions and Transportation info:
http://www1.hilton.com/en_US/hi/hotel/SANTPHH-Hilton-La-Jolla-Torrey-Pines-California/directions.do

---

Click Here to Register, or fax the form below to 781-972-5425

☐ Cloud Computing Training Pricing $1395

First: _____ Last: _____

Title: _____ Div/Dept.: _____

Company:_____

Address: _____

City:_____

State: _____ Zip/Postal Code:_____

Country: _____

Email: _____

Telephone: _____ Fax:_____

Yes! I would like to receive a FREE subscription to:

☐ Bio-IT World - a monthly magazine focused on technologies for the life sciences

☐ eCliniqua - Innovative management in clinical trials

☐ Bio-IT World Weekly Update - Industry news, commentary, and feature highlights from Bio-IT World

☐ YES! I would like more information on CHI's Insight Pharma Reports – Expert Intelligence for Better Decisions

Billing Information

Cardholder's Name: _____

Credit Card Number:_____ Exp. Date: _____

Address: _____

City:_____

State: _____ Zip/Postal Code:_____

Cambridge Healthtech Institute
250 First Avenue, Suite 300
Needham, MA 02494
www.healthtech.com

BioTeam Inc.
Cambridge, MA USA
http://www.bioteam.net

BIOTEAM
Enabling Science

Primary Contact: Olga Pavleyeva
Email: opavleyeva@healthtech.com

Primary Contact: Chris Dagdigain
Email: chris@bioteam.net