Blair Boyd
Product Manager

Circe Tsui
Associate Director, Solutions Architecture

Agenda

- What is AWS?
- AWS at Emory
- Research Use Cases
- Demo
- Service Offering
- Questions

EMORY
LIBRARIES &
INFORMATION
TECHNOLOGY

What is Amazon Web Services

- Market leader of Infrastructure as a Service (IaaS) providers
- Over 100 Products and Services
- No up-front investment
- Pay-as-you-go pricing
- Fast, elastic, scalable and with high availability

Amazon EC2
i.e. Virtual Computer

Amazon RDS
i.e. Relational Database Service

Amazon S3
i.e. "Box" on AWS

AWS at Emory

- In the near future, we will have an Emory-specific environment that our users should move to

- Until that time, we are assisting researchers set up an interim AWS account

- The interim account will require the use of a Pcard or credit card

- At this time, you cannot upload any PHI or sensitive data.

EMORY
LIBRARIES &
INFORMATION
TECHNOLOGY

Pricing Examples as of March 2017

Resource	Capacity	Cost *
On-demand EC2 Server (t2.2xlarge) for General Purpose	8 Cores with 32 GiB	$3 for 8 hours
On-demand EC2 Server (r4.8xlarge) for memory intensive applications	32 Cores with 244 GiB	$17 for 8 hours
General purpose EBS Volume	100 GB	$7/month
On-demand small Oracle database (db.t2.small)	100 GB	$32.89 / month
S3 Storage standard	100 GB	$2.18 / month
Glacier storage	100 GB	$0.4 / month

* Data Transfer out and request fees may apply.
See http://calculator.s3.amazonaws.com/ for details

EMORY
LIBRARIES &
INFORMATION
TECHNOLOGY

Research Use Cases

School of Nursing Project

- Needed a robust Oracle database to host 100,000 de-identified Emory Health Care patient records for both research and teaching. The data included hospital stays, patient visits, labs, diagnoses, etc.

AWS Relational Database Service

- Configure

Ready in 10 to 30 minutes

Locally Hosted

- Acquire a server
- Install Operating System
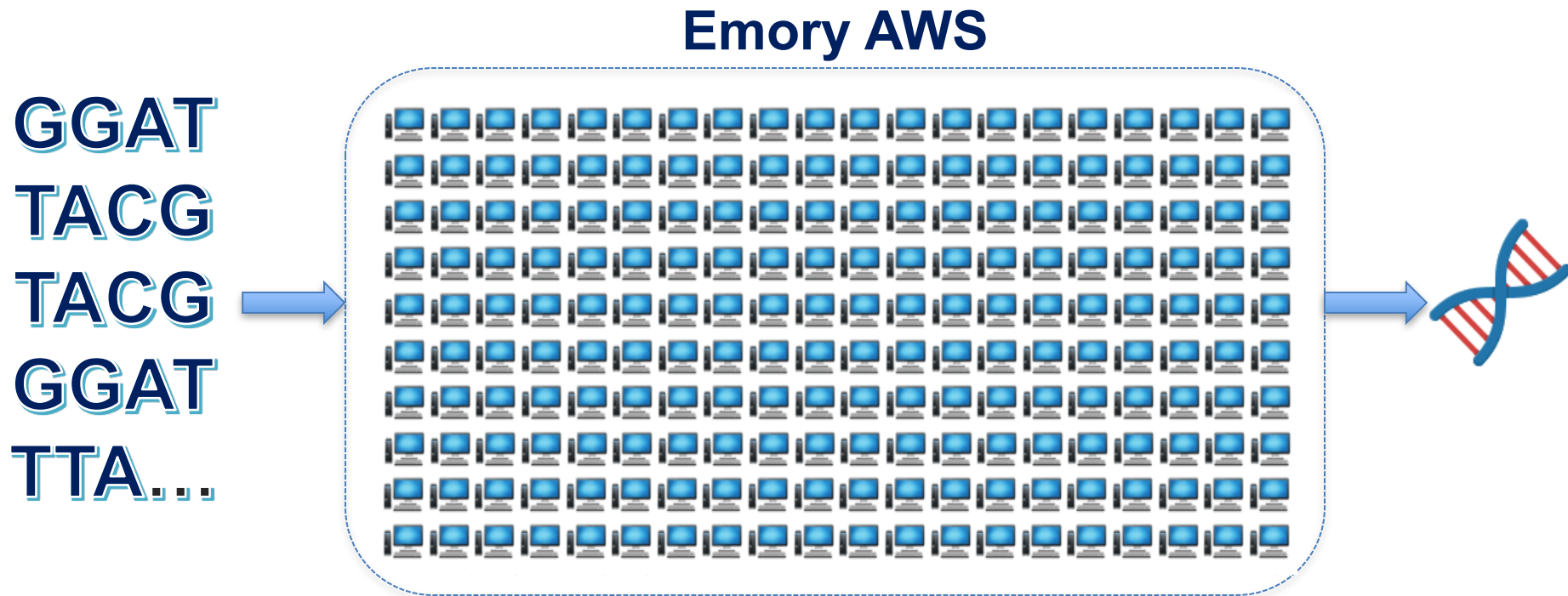- Install Oracle Sof...
- Create a...
- ...ery

Ready in 5 days to 5 weeks

Acknowledgement: Andi Plotsky

Emory Integrated Genomic Core (EIGC) Project

- Needed to analyze 1000+ human genomes (~1 PB total) fast!
- With AWS, 200 genomes were analyzed in parallel with 200 EC2s, i.e. "virtual computers", each with 32 vCPU and 244 GB memory.

Emory AWS

**GGAT
TACG
TACG
GGAT
TTA…**

Emory Integrated Genomic Core (EIGC) Project

On AWS

~~Took 5 days~~

On Tardis HPC Cluster

~~Take 50 days~~

On EIGC Compute Cluster

~~Take 500 days~~

On your local computer

~~Forget about it~~

Acknowledgement: Rich Johnston

EMORY
LIBRARIES &
INFORMATION
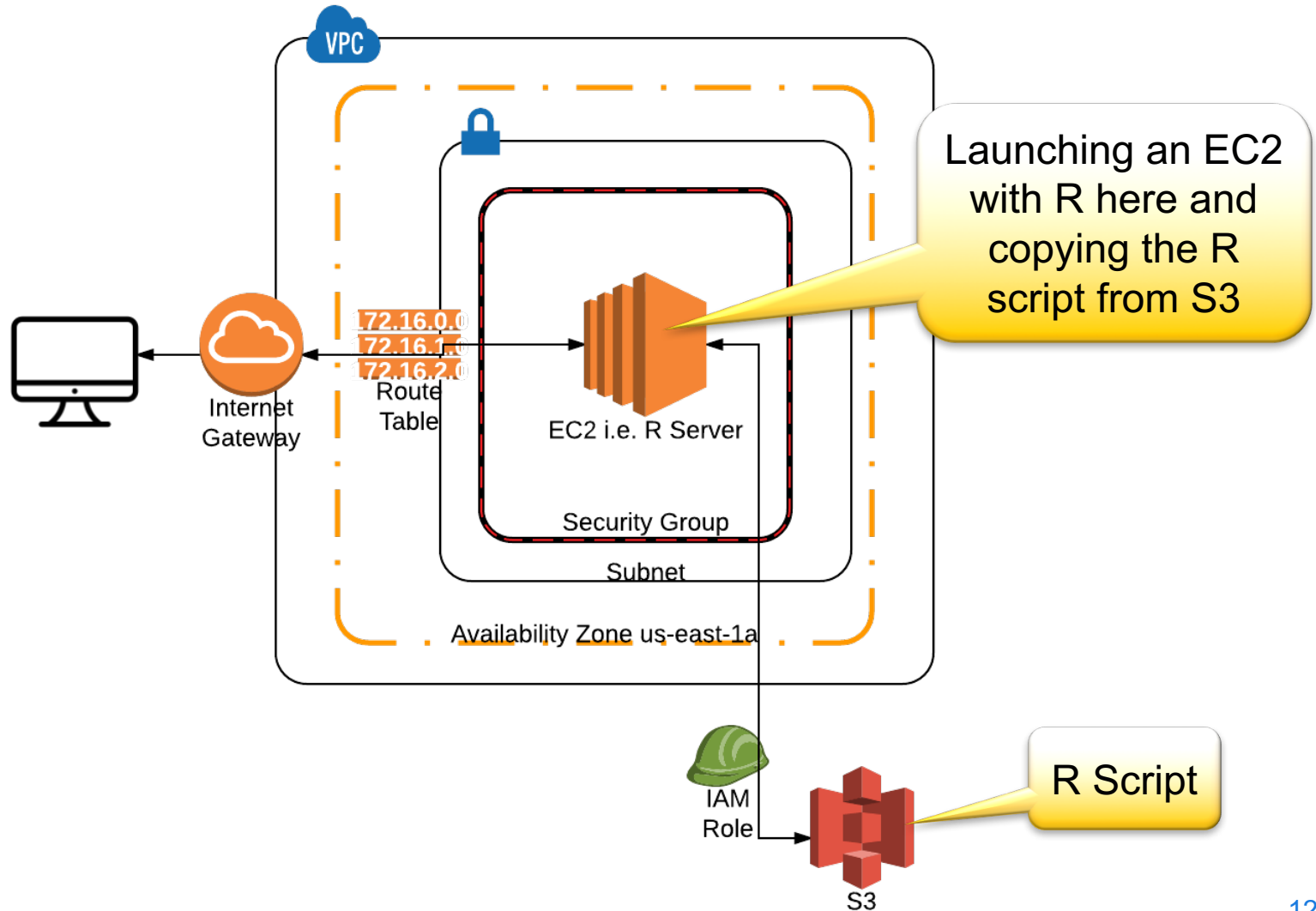TECHNOLOGY

Demo

EMORY
LIBRARIES &
INFORMATION
TECHNOLOGY

AWS Demo

Analyze 2.8 million records from the Centers for [Medicare & Medicaid Services](#) in an on-demand AWS m4.2xlarge (32G RAM / 8 vCPU) EC2 built with LITS's "R Server Amazon Machine Image" launched by a Cloudformation script.

Credits: https://www.r-bloggers.com/bigglm-on-your-big-data-set-in-open-source-r-it-just-works-similar-as-in-sas/

EMORY
LIBRARIES &
INFORMATION
TECHNOLOGY

AWS Demo – R Sever & the AWS Environment

Service Offering

EMORY
LIBRARIES &
INFORMATION
TECHNOLOGY

How LITS Can Help:

- Setting up your AWS account
- Providing an Amazon Machine Image with commonly use scientific software packages
- Training
- Consultation by certified AWS Solution Architects
- Professional design, Implementation and Management under our service center model

14

EMORY
LIBRARIES &
INFORMATION
TECHNOLOGY

Questions?

hpc.help@emory.edu

https://wiki.service.emory.edu/display/public/Training

Blair Boyd

- Email: blair.boyd@emory.edu
- Phone: 404.778.4916

Circe Tsui

- Email: circe.tsui@emory.edu
- Phone: 404.727.7035