



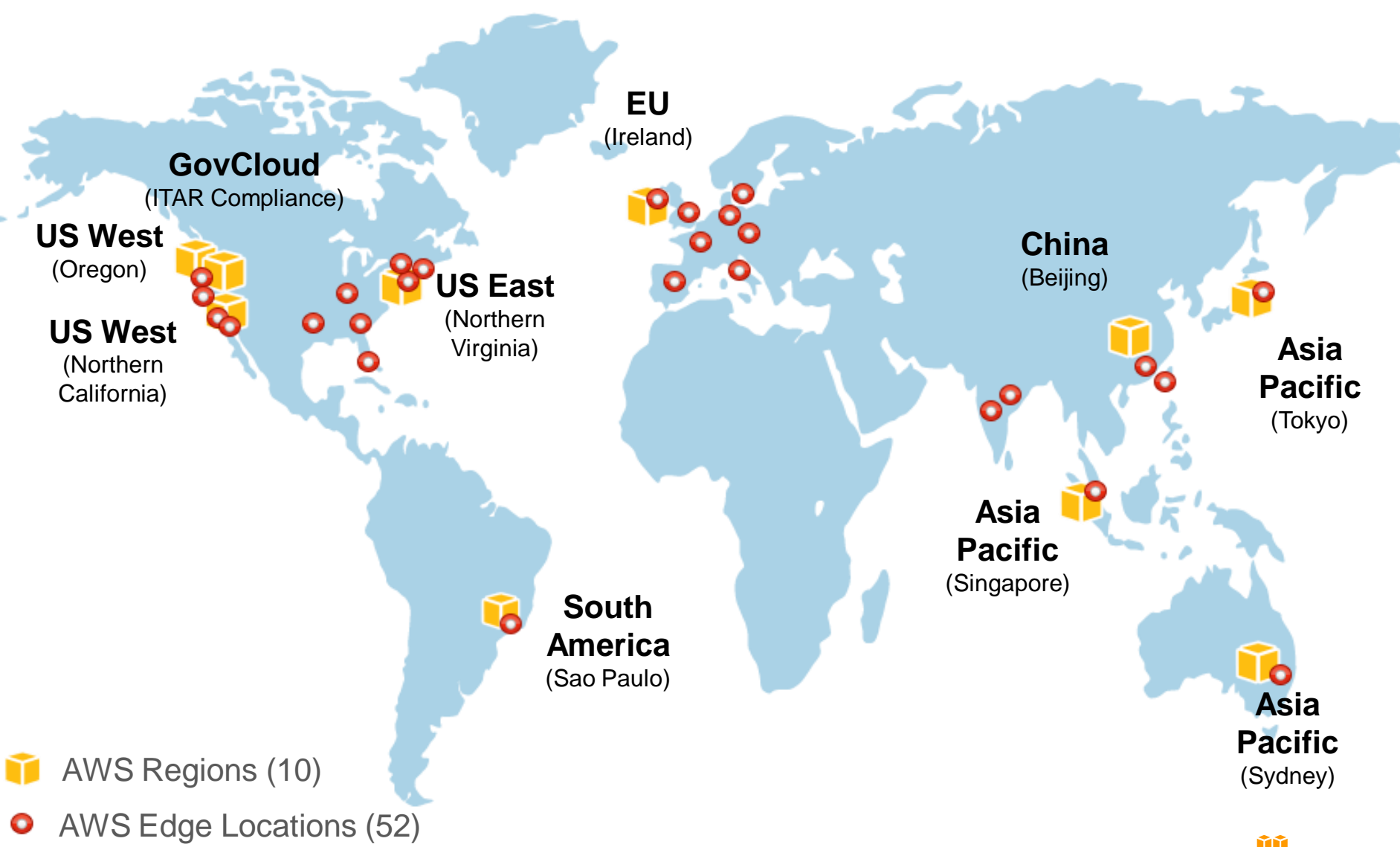
## **AWS Cloud for HPC and Big Data**

David Pellerin, Business Development Principal  
IDC HPC User Forum

September 16, 2014



# AWS Regions



# Customer Applications

## Development & Deployment

### Monitoring



CloudWatch

### Deployment & Management



BeanStalk



OpsWork



Cloud  
Formation



CloudTrail

### Identity & Access



IAM



Federation

## Human Interaction



Support

## Interaction



Web Console



Command Line

API

## Libraries, SDK's



## Application Services

### Applications



SES



SNS



SQS



Elastic  
Transcoder



CloudSearch



SWF



AppStream

## Foundation Services

### Databases



RDS



Dynamo



ElastiCache



RedShift

### Analytics



EMR



DataPipeline



Kinesis

### Content Delivery



CloudFront

### Compute



EC2



WorkSpaces

### Storage



S3



EBS



Glacier



Storage  
Gateway

### Networking



VPC



Direct  
Connect



ELB



Route53

Regions

Availability Zones

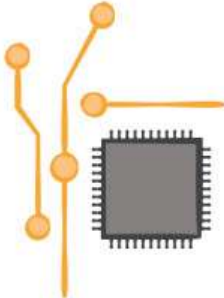


Edge Locations

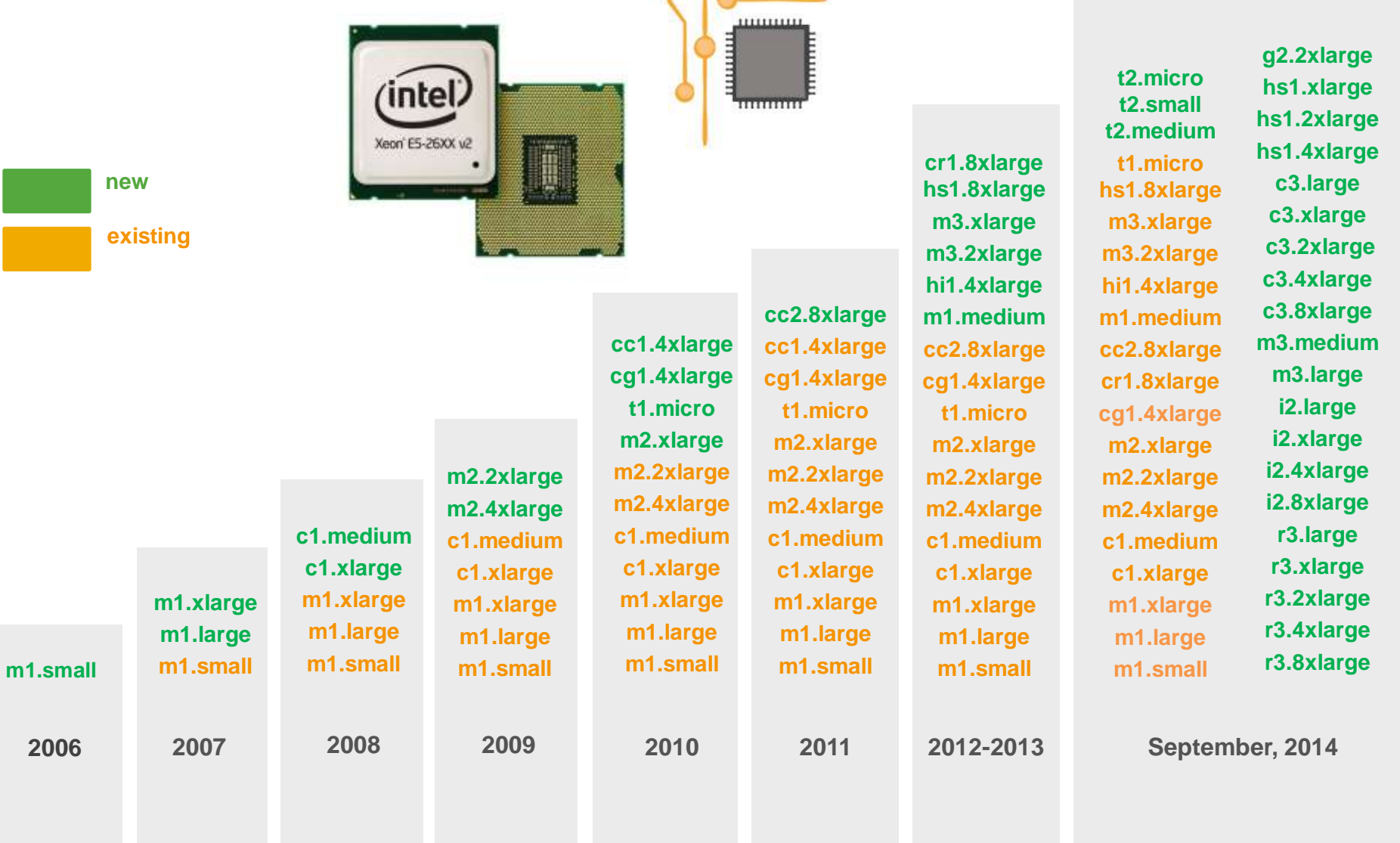
AWS Global Infrastructure

# EC2 Instance Type History

Increasing customer choice...



■ new  
■ existing

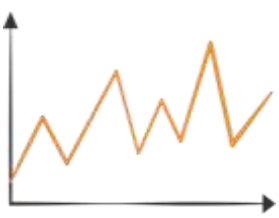


# Multiple Purchase Models

## On-Demand

Pay for compute capacity by the hour with no long-term commitments

For spiky workloads, or to define needs



## Reserved

Make a low, one-time payment and receive a significant discount on the hourly charge

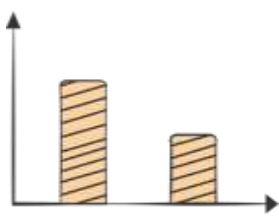
For committed utilization



## Spot

Bid for unused capacity, charged at a Spot Price which fluctuates based on supply and demand

For time-insensitive or transient workloads



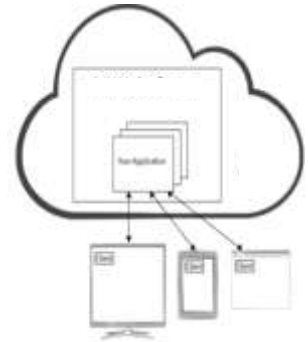
**AWS Spot is a game-changer for HPC**

# Motivators for HPC in the Cloud

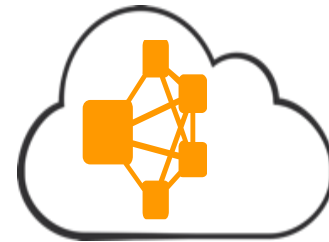
**Cloud for HPC Scalability**



**Cloud for Secure Global Collaboration**



**Cloud for Big Data**



## Zero to Cloud in 6+ Month

By 31 Oct 2013:

- ✓ Cloud eMail – Microsoft Office365
- ✓ Cloud eMail archiving/eDiscovery
- ✓ External SingleSignOn (off VPN)
- ✓ Cloud File/Collaboration – BOX
- ✓ Cloud CRM – Salesforce.com
  - ✓ Integrated to save files in BOX
- ✓ Cloud–High Performance Computing (HPC) on Amazon AWS
- ✓ Cloud – Big Data Platform on Amazon AWS



*“HGST is using AWS for a higher performance, lower cost, faster deployed solution vs buying a huge on-site cluster.”*

*- Steve Philpott, CIO*

## HGST application roadmap:

- ✓ Molecular dynamics
- ✓ CAD, CFD, EDA
- ✓ Collaboration tools for engineering
- ✓ Big data for manufacturing yield analysis, including Amazon Redshift

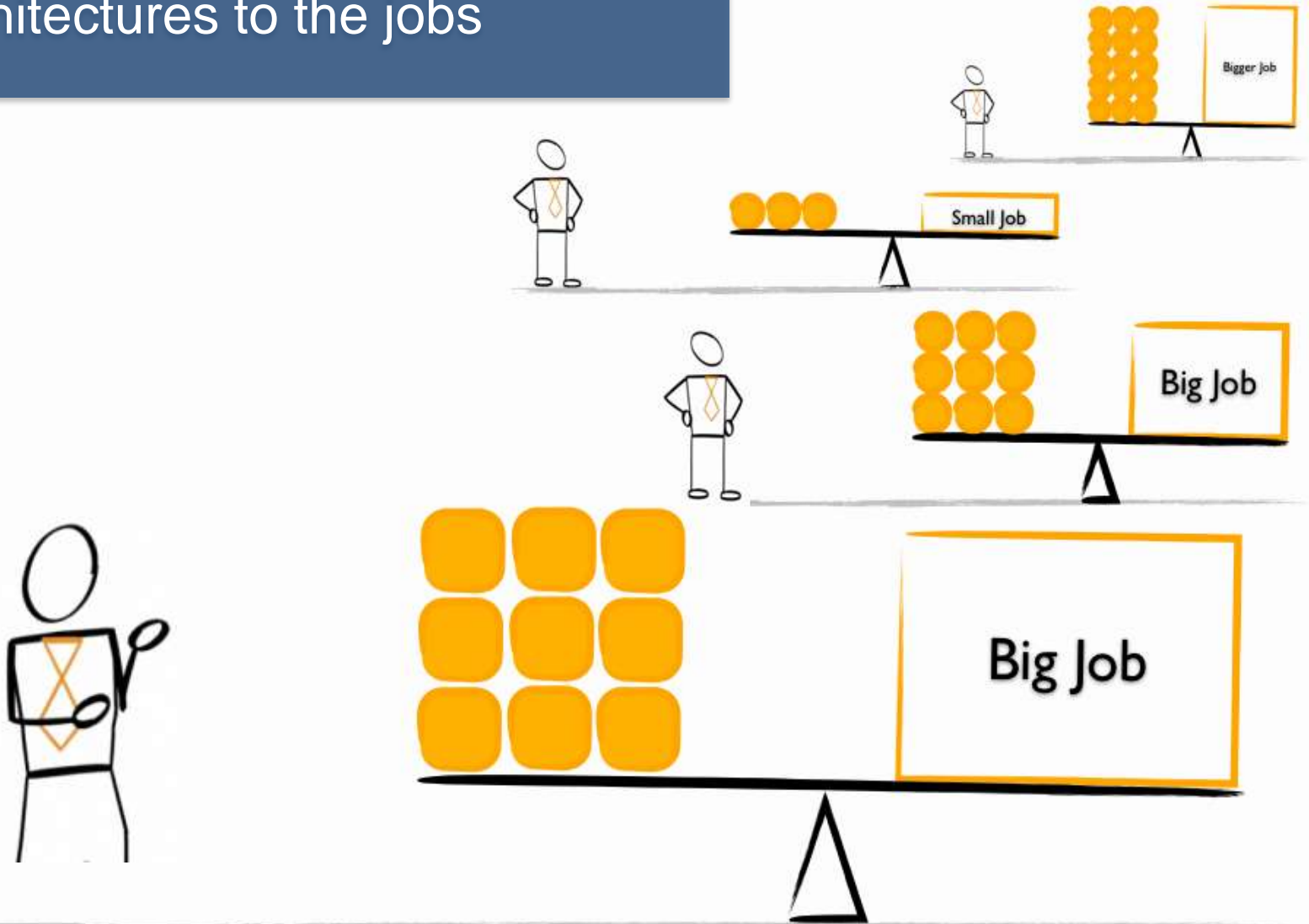
## HGST's Amazon HPC Platform



### Base HPC Platform

- Scalable to thousands of instances to support numerous simultaneous simulations

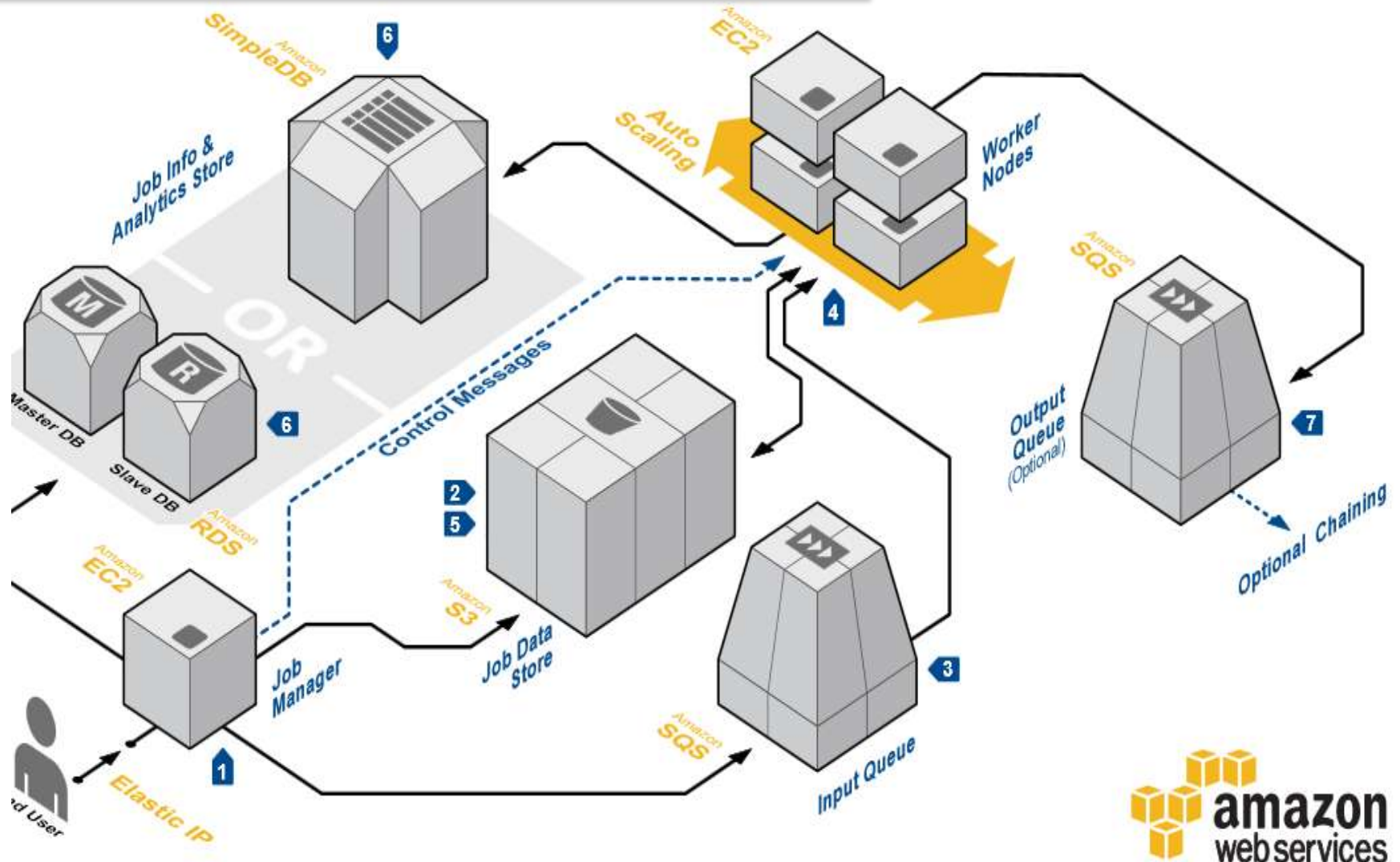
On AWS, deploy multiple clusters running at the same time and match the architectures to the jobs





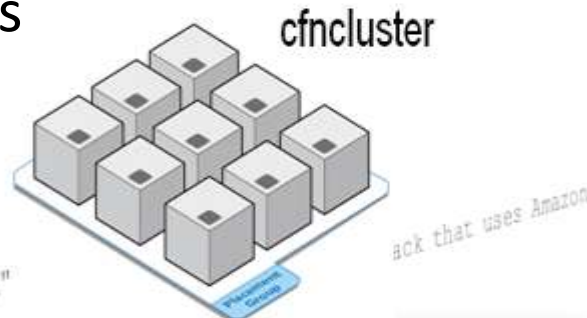
# Use automation to manage cluster sizing and monitor jobs and costs

*AWS Auto Scaling can work with existing HPC scheduling software*



# Many HPC Deployment Methods

- Traditional HPC schedulers and cluster managers
- “Born in the cloud” tools
  - MIT StarCluster
  - Cycle Computing CycleServer
- AWS-provided tools and APIs
  - Cloudformation, Auto Scaling
  - cfncluster (github.com/awslabs/cfncluster)
- For many use-cases...



```
"AWSTemplateFormatVersion": "2010-09-09",  
"Description": "This template creates a cluster.",  
"Parameters": {  
  "InstanceType": {  
    "Type": "String",  
    "Description": "The EC2 instance type to use.",  
    "AllowedValues": ["m1.xlarge", "m2.xlarge", "m3.xlarge", "m3.xlarge"],  
    "ConstraintDescription": "Must be one of the allowed values."  
  }  
}
```

```
aws@ip-10-10-10-10: ~$ cfncluster --help  
usage: cfncluster [-h] [-config CONFIG_FILE] [-region REGION] [--noexit] [status_stop_create_list_update_instances_scheduler_delete] ...  
cfncluster is the a tool to launch and manage cluster.  
  
positional arguments:  
  status_stop_create_list_update_instances_scheduler_delete  
  status      creates a cluster  
  create      updates a running cluster  
  update      stops a cluster  
  stop        deletes a cluster  
  delete      pulls the current status of the Cluster  
  list        displays a list of stacks associated with cfncluster  
  instances   displays a list of all instances in a cluster  
  scheduler   sets to scheduler instance  
  
optional arguments:  
  -h, --help      show this help message and exit  
  --config CONFIG_FILE, -c CONFIG_FILE  
                  specify a distinctive config file  
  --region REGION, -r REGION  
                  specify a specific region to connect to  
  --noexit, -n    do not wait for stack events, after executing stack command  
aws@ip-10-10-10-10: ~$
```

# Touch-Sensor Modeling on AWS

for TRUETOUCH® Touchscreen Controllers



A. Gourevitch

*Cypress Semiconductor Corp., San Jose, CA, USA*

We report an implementation of parallel computing on Amazon Web Services™ (AWS) for touch-sensor modeling. COMSOL Multiphysics® was used to simulate an electromagnetic field distribution in a capacitive sensor assembly. Multiple COMSOL jobs were deployed on separate AWS instances and were executed in parallel. The simulation results indicate that implementation of parallel computing for COMSOL simulations can significantly reduce the computational time required for optimization of capacitive touch sensor patterns.

Files Available for Download

[Abstract](#)

[Paper](#)



*Courtesy of Cypress Semiconductor*

# Reservoir Simulation on AWS



Search 

+61 (08) 9446 2099

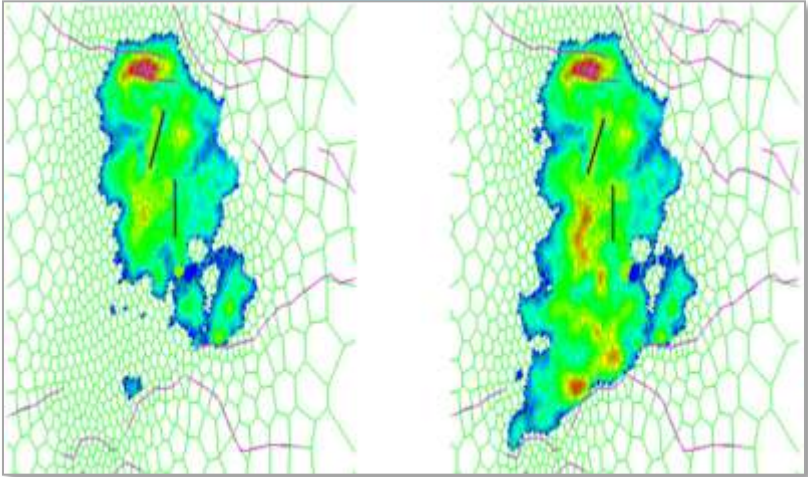
[Home](#) | [About Us](#) | [ResAssure](#) | [GasAssure](#) | [News](#) | [Resources](#) | [Investors](#) | [Contact](#)

[Technical Overview](#) | [Unique Features](#) | [Million Realisations in 24 Hours](#) | [Features and Benefits](#) | [Consulting](#) | [FAQs](#) | [Register Your Interest](#) | [Benchmark Studies](#)

## Technical Overview

### What is ResAssure?

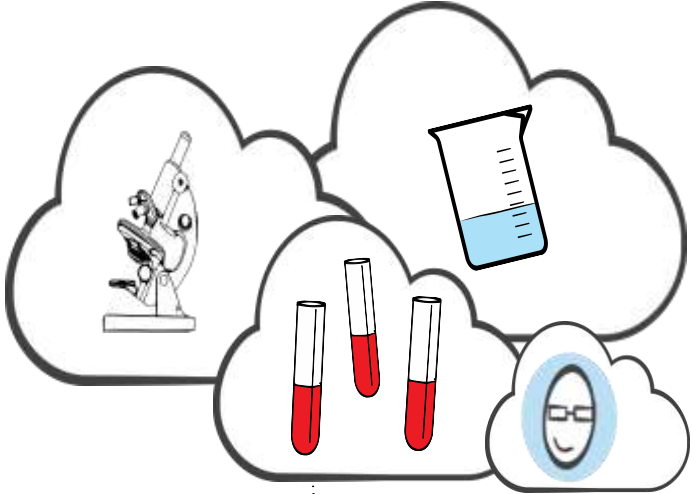
ResAssure is a Stochastic Simulation software solution and it's innovation in reservoir simulation currently solves fully-implicit, dynamic three-phase fluid flow equations for every geological realisation. ResAssure marks a significant milestone in the history of reservoir simulation, leading technological advancements in the oil and gas industry. It has been developed to work with standard reservoir simulation package datasets such as Eclipse, CMG and VIP.






# Bristol-Myers Squibb Clinical Trials on AWS

	On-Premises	Cloud
# of Simulations	2000	2000
# of Servers	2	256
<b>Total Run Time (hr)</b>	<b>60</b>	<b>1.2</b>



Clinical trial simulations took 98% less time  
More efficient and iterative simulations results in fewer human trials  
64% savings on clinical trial costs

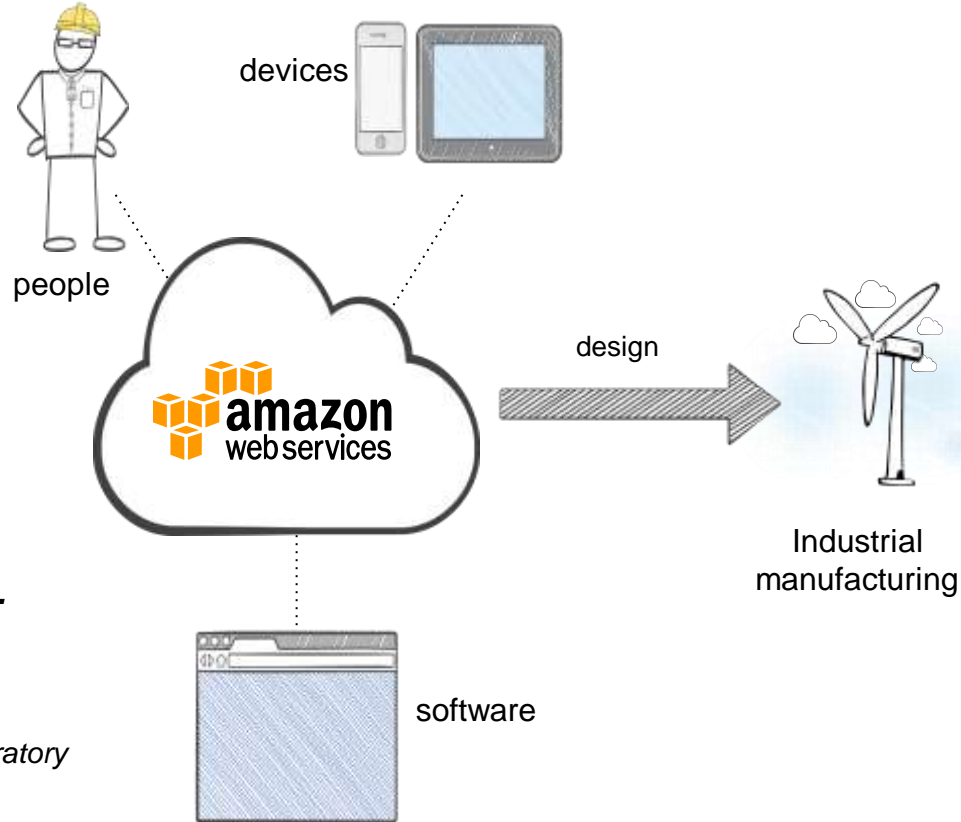
**“We’re using fewer subjects in these trials, and needing fewer blood samples.”**

 Russell Towell  
Senior Solutions Specialist




# Collaboration and Design in the Cloud

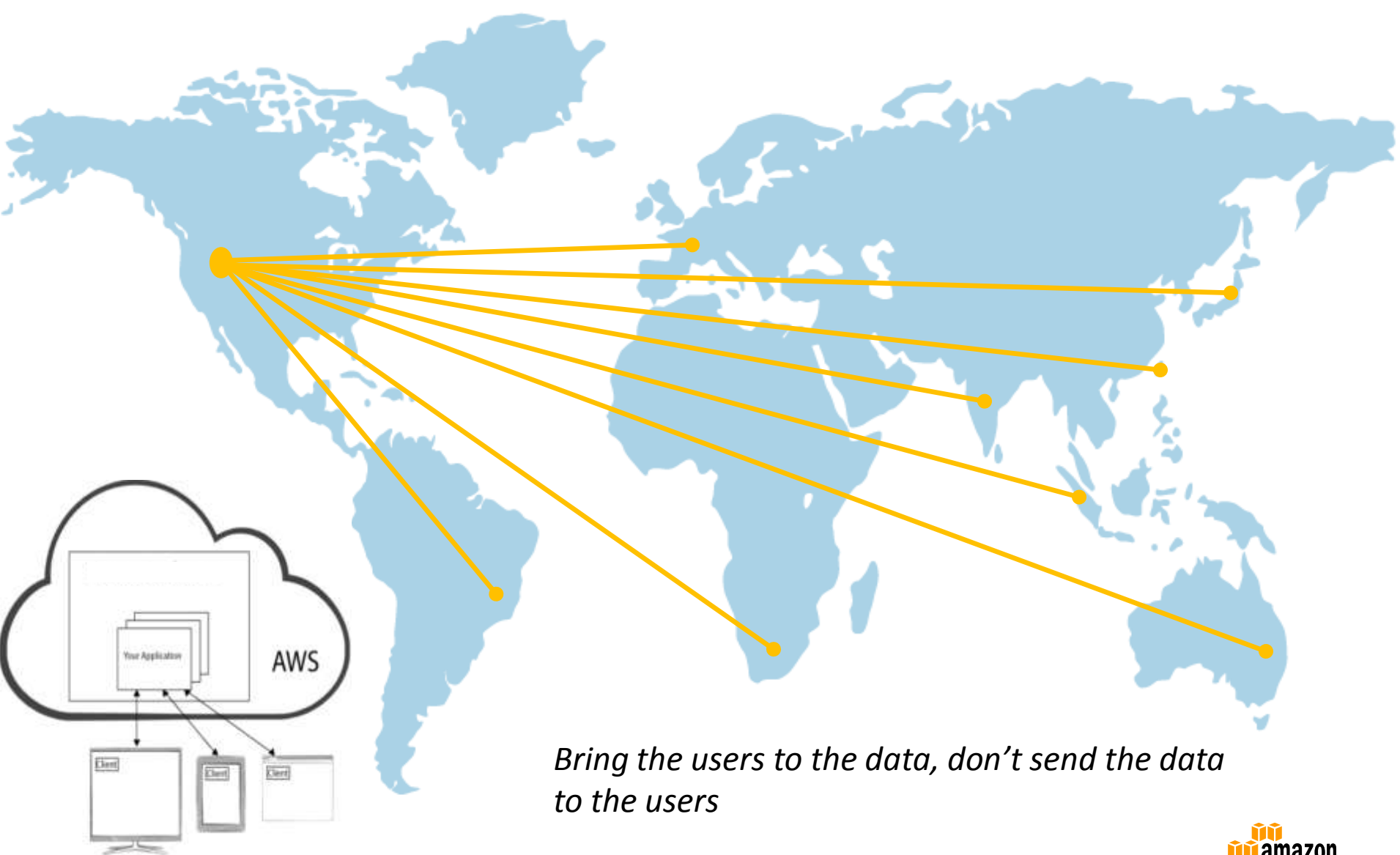
Cross-functional collaboration app  
Helps design around manufacturing  
Allows users to define how they work  
Users can spin-up their own environments



**“ This could change the way manufacturing is architected. ”**

 *Joe Salvo  
Manager, Business Integration Technologies Laboratory*

# Enabling Global Collaboration



*Bring the users to the data, don't send the data to the users*

# Thin Client Remote Collaboration

*Calgary Scientific PureWeb™*

CALGARY SCIENTIFIC



[www.calgaryscientific.com/resolutionmd/web/](http://www.calgaryscientific.com/resolutionmd/web/)

[demos.getpureweb.com/](http://demos.getpureweb.com/)

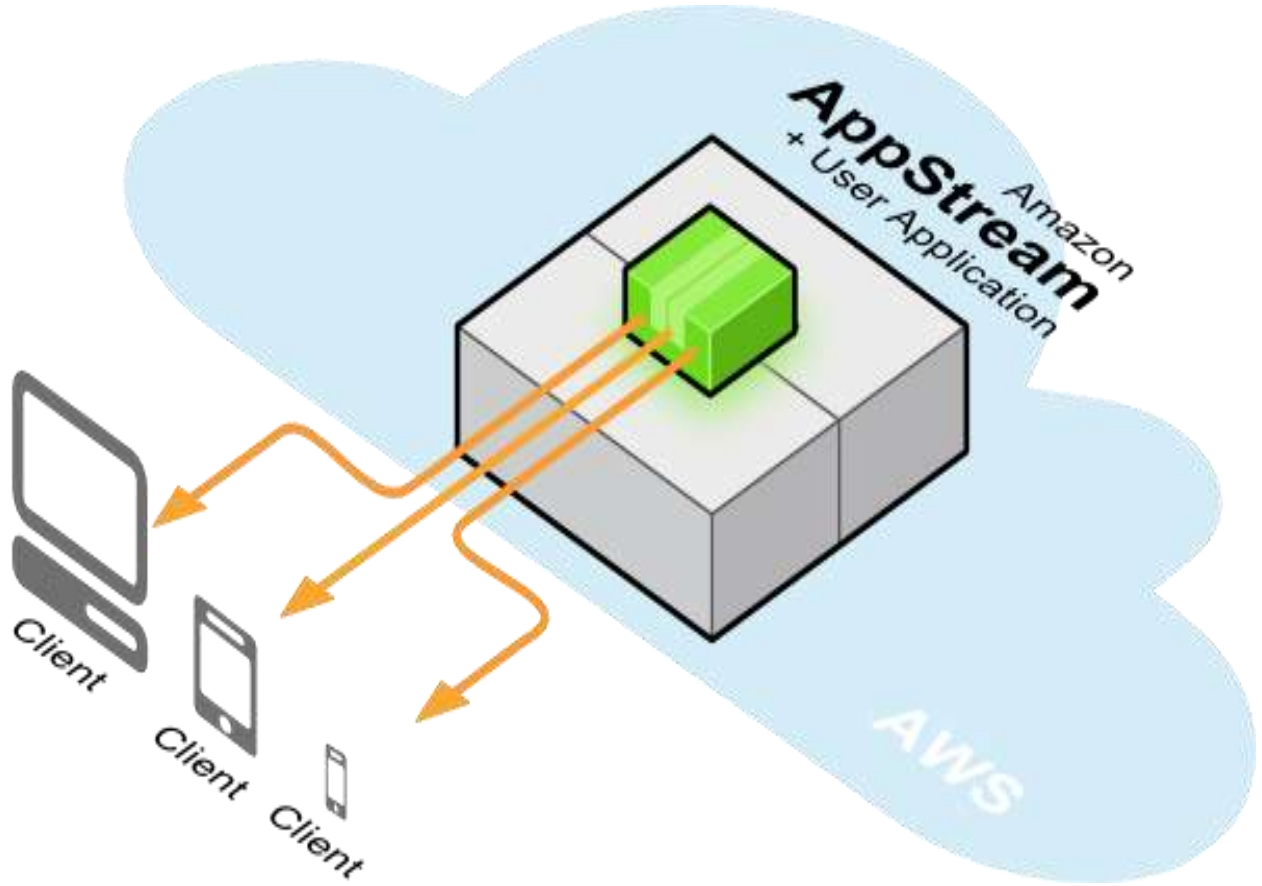




# Amazon AppStream



- Application Streaming
- Remote visualization
- Thin client 3D applications



Big Data  
Plus Cloud  
Equals Awesome

# AWS Has Always Been About Big Data

## WEB LOG ANALYSIS

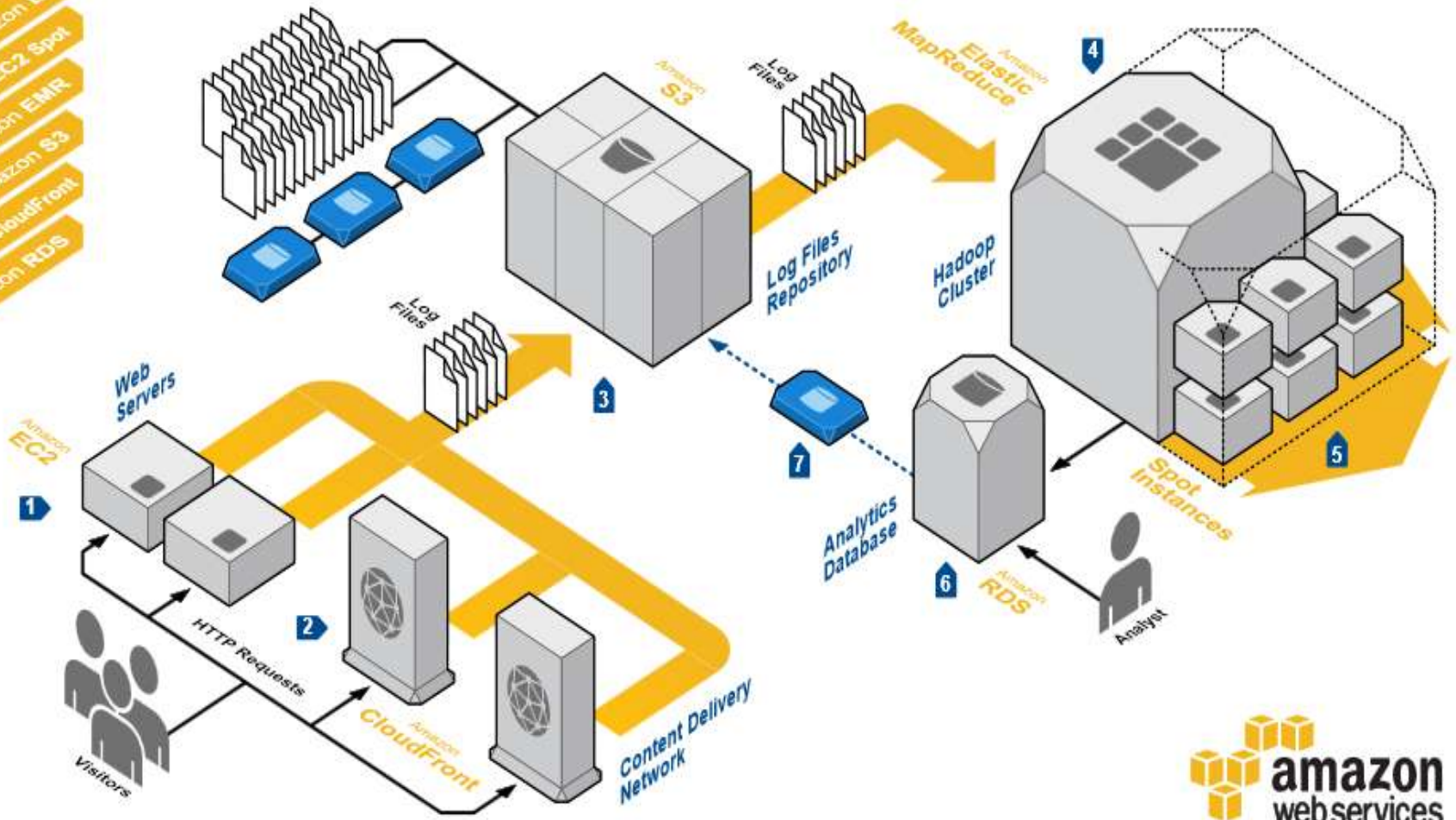
Amazon Web Services provides services and infrastructure to build reliable, fault-tolerant, and highly available web applications in the cloud. In production environments, these applications can generate huge amounts of log information.

This data can be an important source of knowledge for any company that is operating web applications. Analyzing logs can reveal information such as traffic patterns, user behavior, marketing profiles, etc.

However, as the web application grows and the number of visitors increases, storing and analyzing web logs becomes increasingly challenging.

This diagram shows how to use Amazon Web Services to build a scalable and reliable large-scale log analytics platform. The core component of this architecture is Amazon Elastic MapReduce, a web service that enables analysts to process large amounts of data easily and cost-effectively using a Hadoop hosted framework.

- AWS Reference Architectures
- Amazon EC2
- Amazon EC2 Spot
- Amazon EMR
- Amazon S3
- Amazon CloudFront
- Amazon RDS



# Big Data for Financial Market Analysis

## AWS Case Study: FINRA



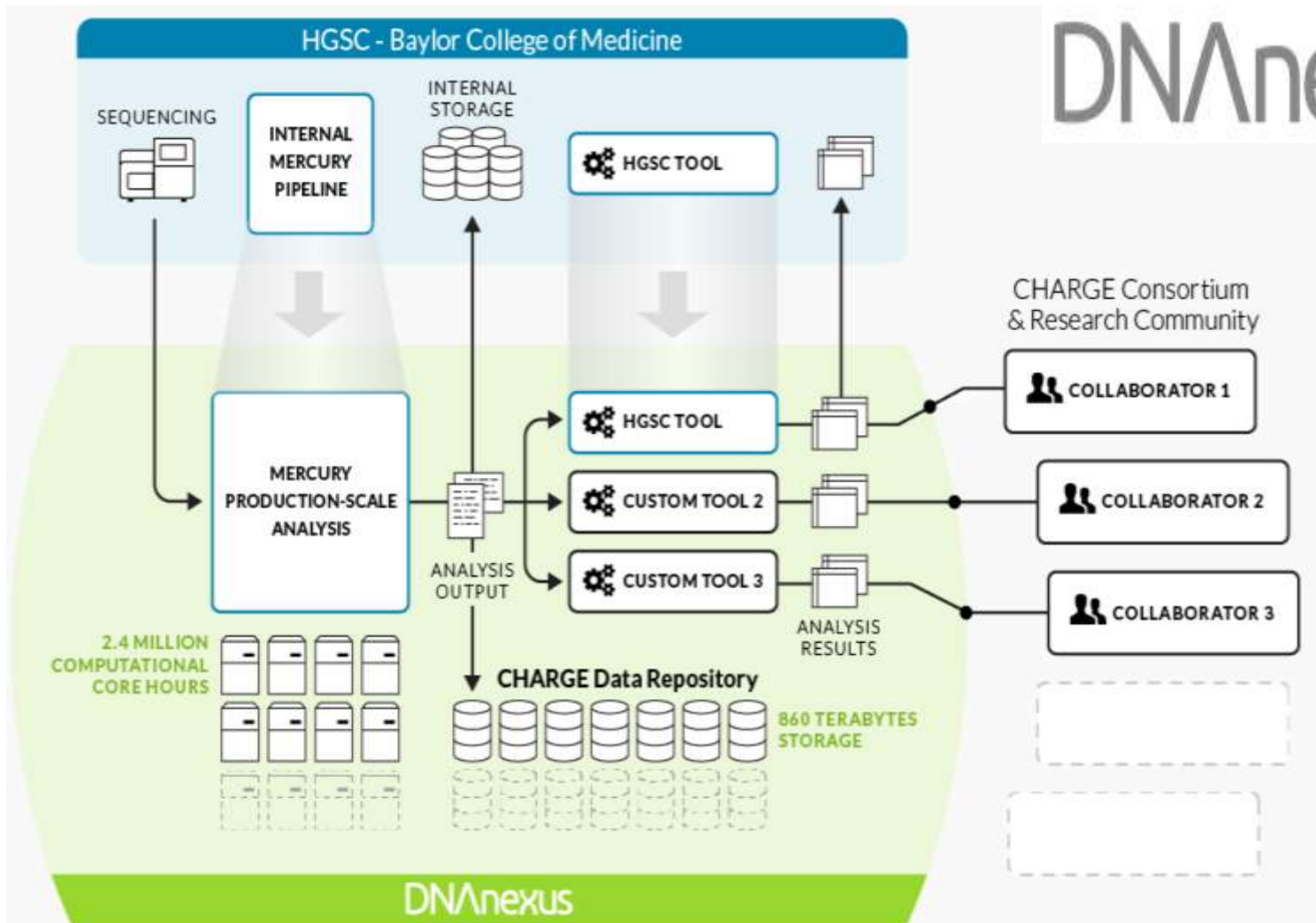
### About FINRA

The [Financial Industry Regulatory Authority \(FINRA\)](#), one of the largest independent securities regulators in the U.S., was established to help watch and regulate financial trading practices. To respond to rapidly changing market dynamics, FINRA is moving its platform to Amazon Web Services (AWS) to analyze and store approximately 30 billion market events every day.

FINRA selected AWS because it offered the right services while fulfilling the company's security requirements. By using dynamic clusters (Hadoop, Hive, and HBase), and services such as [Amazon Elastic MapReduce \(Amazon EMR\)](#) and [Amazon Simple Storage Service \(Amazon S3\)](#), FINRA was able to create a flexible platform that can adapt to changing market dynamics. By using the AWS Cloud, FINRA has been able to increase agility, speed and cost savings while allowing them to operate at scale. The company estimates it will save \$10 to \$20 million annually by using AWS.

# Big Data for Genome Analysis

Baylor College of Medicine, Amazon Web Services, and DNAnexus: cloud-based analysis of genomic data from over 14,000 patients





# Big Data is Everywhere!

**THE CLIMATE CORPORATION**

Products ▾ Testimonials Agents Sign In

## CLIMATE | BASIC

Better Data. Better Decisions.  
The most advanced way to get insight into your fields.

- Field Workability
- Weather
- Alerts
- Field Details
- Scouting & Notes
- Reports

Sign Up Learn More

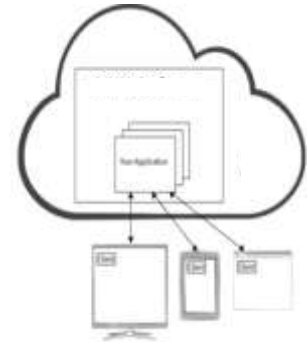
HOMESTEAD	
POTENTIAL YIELD (ESTIMATED)	155
YIELD (BU/AC)	145
POTENTIAL YIELD (ESTIMATED)	120
HARVEST DATE	Sept 27
GROWTH STAGE	R6
PRECIP LAST 24H	0.65"

# Summary

**Cloud for Scalability**



**Cloud for Global Collaboration**



**Cloud for Big Data**





Thank You