COLUMBIA BUSINESS SCHOOL

CS@CU

# Data Science and Technology Entrepreneurship

## Legal aspects of a Startup (Guest Lecture)
## Technology Choices in a Startup

Week 13

Sameer Maskey

# Announcements

- Friday Open Office hours + Co-work with teams
  - Come and work together
  - Ask questions
  - Dates :
    - April 19 - 4:30 - 6:30 pm
    - April 26 - 4:30 - 6:30 pm
    - May 3 - 4:30 - 6:30 pm
- Location
  - CS conference room

# Announcements

- Assignment 4 is out
- Due May 5th Sunday @6pm
- 4 parts
  - Part 1 - Final Report
    - I have written down 16 points or so that should guide you on how to write the Final Report
    - (Important) – Please provide the final business model canvas.
    - Write a paragraph on individual role and contribution of each team member to the business.

# Announcements

- Assignment 4
  - Part II - Final Technical Report
    - System Architecture Diagram
    - Overall choices you made
  - Part III - MVP
    - Last iteration of your MVP
    - Please write about the changes you have made from the last MVP
  - Part IV - MVP Field Test
    - Similar to last homeworks - customer validation of your MVP with at least 5 customers

# Announcements

- Final Presentation
  - Date : May 7th (Tuesday)
    - Yes it's on **Tuesday**
    - First day of reading week
  - Location - Uris 142
  - Schedule
    - 12:45 - 1:00 pm - Registration
    - 1:00 - 1:15 - Overview of the class and projects
    - 1:15 - 1:30 - Mentors/Advisors Introductions
    - 1:30 - 2:30 - Student Presentation (4 min each)
    - 2:30 - 3:15 - Coffee and Sandwiches
    - 3:15 - 4:15 - 12 concurrent Student Demos (judges score the demos)
    - 4:15 - 4:30 - Winners announcement

# Announcements

- Final Pitch/Demo Guidelines
  - 4 min long
  - Judges will provide feedback in judging time
  - 12 concurrent demos
    - 2 laptops
      - 1 for slides
      - 1 for demo

# Guest Lectures

- Jane Jablons
  - Partner, Kelley & Drye

# Topics for Today

- Legal aspects of a Startup - Guest Lecture - Jane Jablons
- Technology Choices in Startups

# Technology Choices for a Startup

- We will particularly talk about web technology choices
- Programming Language?
- Frameworks?
- Databases?
- Hosting Provider?
- Linux vs Windows?
- many choices you will have to make ...

# Technology Baggage is Heavy

- If you make a wrong choice in the beginning it will cost a lot of
  - time
  - effort
  - money
- You can't really start changing programming language you use after 1 year of development

# Technology Choices

**Programming languages used in most popular website**

| Website | Popularity (unique visitors)[1] | Frontend (Client-side) | Backend (Server-side) | Database |
|---|---|---|---|---|
| Google.com[2] | 1,000,000,000 | JavaScript, Ajax | C, C++, Go[3], Java, Python, PHP | BigTable[4] |
| Facebook.com | 880,000,000 | JavaScript, Ajax | PHP, C++, Java, Python, Erlang | MySQL |
| YouTube.com | 800,000,000 | Flash, JavaScript | C, Python, Java | MySQL |
| Yahoo | 590,000,000 | JavaScript, Ajax | PHP | MySQL |
| Live.com | 490,000,000 | JavaScript, Ajax | ASP.NET | Microsoft SQL Server |
| MSN.com | 440,000,000 | | ASP.NET | Microsoft SQL Server |
| Wikipedia.org | 410,000,000 | | PHP | MySQL |
| Blogger | 340,000,000 | | Python | BigTable |
| Bing | 230,000,000 | JavaScript | ASP.NET | Microsoft SQL Server |
| Twitter.com | 160,000,000 | | C++, Java, RoR, Scala | |
| Wordpress.com | 130,000,000 | JavaScript | PHP | MySQL |
| Amazon.com | 110,000,000 | | Java, J2EE, C++, Perl | |
| eBay.com | 88,000,000 | | Java, WebSphere, Servlets | Oracle Database |
| Linkedin.com | 80,000,000 | | Java, Scala | |

# Impossible to be Expert in All Choices

- If someone comes up to you and says he/she knows all languages, frameworks and databases that person is probably someone you don't want to hire
- Being expert takes time!
- Real life experience in deploying live systems

# Use Open Source Software

- Don’t reinvent the wheel
- Use and re-use what has already been done
- Look at the terms carefully
  - All open source software are not the same
  - Gnu Public License Variation
  - Creative Common License
- Are there risks for using open source software in enterprise framework?

# Our Technology Choices

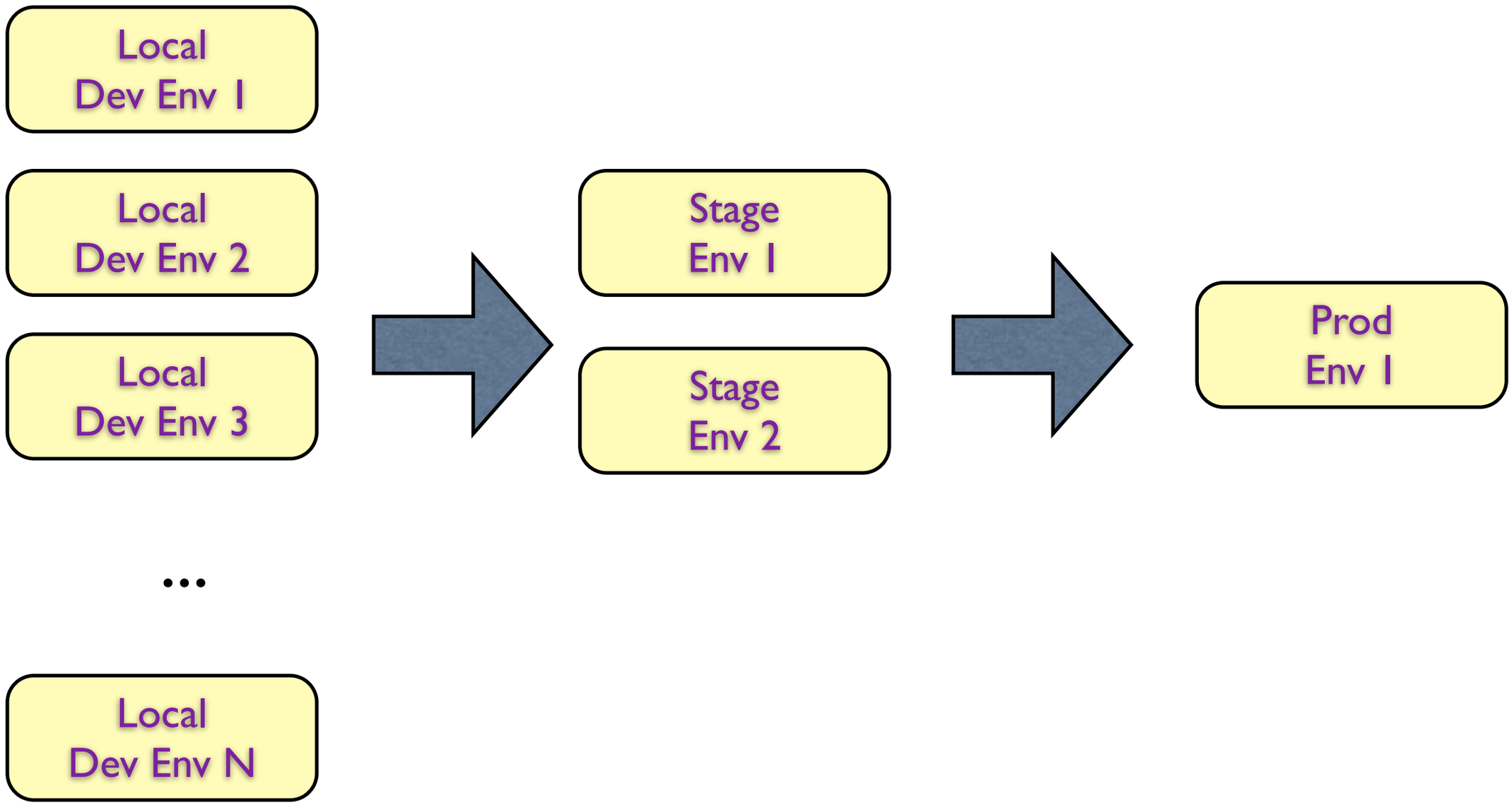
- Some of the technologies and languages we use
  - Java
  - C++
  - Zend Framework (php)
  - Python
  - Perl
  - XMPP
  - NumPy
  - MatLab
  - MapReduce
  - Amazon S3
  - Elastic Beanstalk
  - Route 53
  - Amazon RDS
  - EC2 servers
  - Caching system
  - Amazon VPC
  - Zoho Projects
  - Zoho Invoice
  - Zoho CRM
  - Google Apps

# Dev to Production

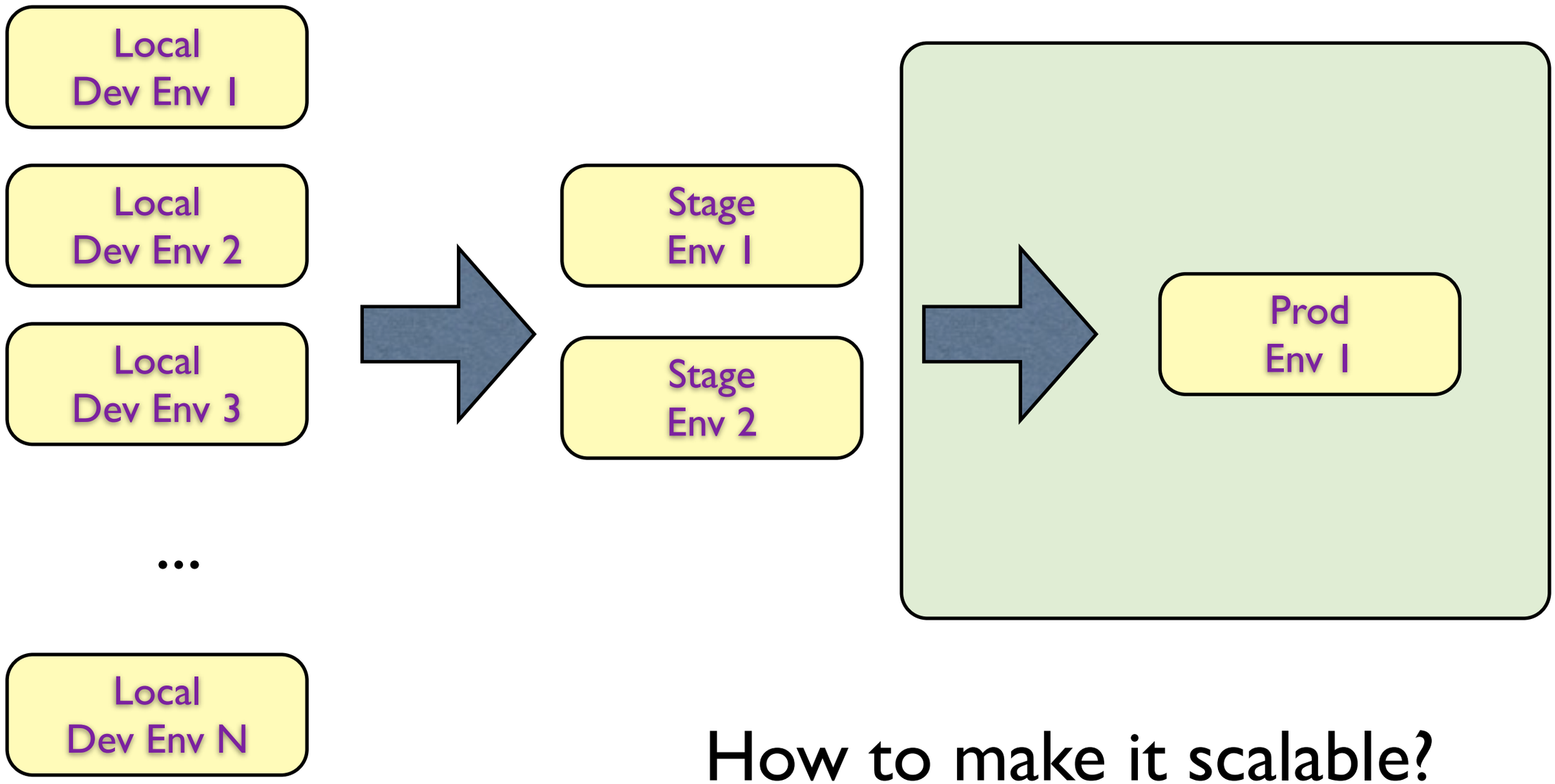
- Think in terms of environment
- Environment that is replicable, stable and easy to test
- Environment may contain
  - Code
  - Data
  - Models
  - Configuration parameters
  - Scaling parameters

# Etsy CTO Makes a Good Point

# Usual Dev to Production Pipeline



# Scalability?




How to make it scalable?

# Amazon Elastic Beanstalk Useful

## Amazon Web Services

### Compute & Networking

- **Direct Connect** — Dedicated Network Connection to AWS
- **EC2** — Virtual Servers in the Cloud
- **Elastic MapReduce** — Managed Hadoop Framework
- **Route 53** — Scalable Domain Name System
- **VPC** — Isolated Cloud Resources

### Storage & Content Delivery

- **CloudFront** — Global Content Delivery Network
- **Glacier** — Archive Storage in the Cloud
- **S3** — Scalable Storage in the Cloud
- **Storage Gateway** — Integrates On-Premises IT Environments with Cloud Storage

### Database

- **DynamoDB** — Predictable and Scalable NoSQL Data Store
- **ElastiCache** — In-Memory Cache
- **RDS** — Managed Relational Database Service
- **Redshift** NEW — Managed Petabyte-Scale Data Warehouse Service

### Deployment & Management

- **CloudFormation** — Templated AWS Resource Creation
- **CloudWatch** — Resource and Application Monitoring
- **Data Pipeline** — Orchestration for Data-Driven Workflows
- **Elastic Beanstalk** — AWS Application Container
- **IAM** — Secure AWS Access Control
- **OpsWorks** NEW — DevOps Application Management Service

### App Services

- **CloudSearch** — Managed Search Service
- **Elastic Transcoder** NEW — Easy-to-use Scalable Media Transcoding
- **SES** — Email Sending Service
- **SNS** — Push Notification Service
- **SQS** — Message Queue Service
- **SWF** — Workflow Service for Coordinating Application Components

# Load Testing

## Load test: SRM Blog Test 100 SBU 5 mins

**Target URL:** http://www.sameermaskey.com/ **Test result public URL:** http://loadimpact.com/load-test/www.sameermaskey.com-7

**Load zone data source**

Aggregated (World)

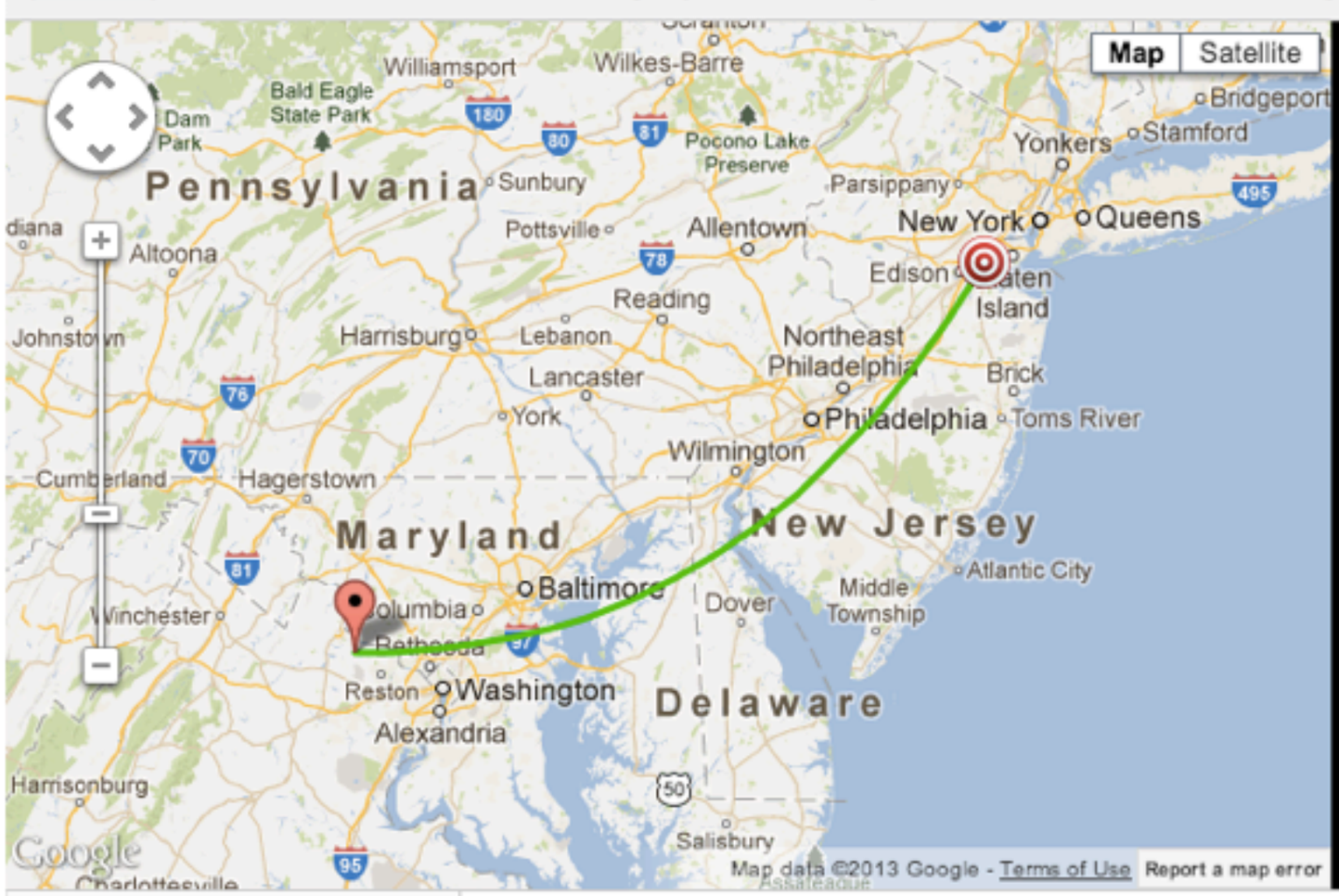
Choose from which load zone to display data.

**Status**

Test finished

**Test configuration:** SRM Blog Test 100 SBU 5 mins
**User type:** SBU ?
**Started:** Wed, 24 Apr 2013 18:17:39
**Ended:** Wed, 24 Apr 2013 18:23:23

| SBUs active | Connections active | Bandwidth | Data received | Requests |
|---|---|---|---|---|
| 0 | 0 | 0 bit/s | 806.57 MiB | 32828 (0 req/s) |
| Total number of simulated clients (VU or SBU) active. | Number of open TCP connections to target system. | Current throughput to target system. | Total number of bytes received during test. | Current number of requests per second. |




```
[13:17] Initializing test...
[13:17] Target system located in Woodbridge,
        United States
[13:17] Booting load generators...
[13:17] Waiting for load generators...100%
[13:17] * Ashburn, US online
[13:17] Deploying test config...100%
[13:17] * Ashburn, US ready
[13:17] Running...100%
[13:23] Waiting for load generators...100%
[13:23] * Ashburn, US finished
[13:23] Cleaning up...100%
[13:23] * Ashburn, US offline
[13:23] Test finished
```

# Elastic BeanStalk

smblog

Upload New Version

▼ Elastic Beanstalk Application Details

Overview | **Events** | Versions

Below are the most recent 1000 events for this application. Click here to learn how you can retrieve all events.

**Viewing:** All Event Types

| Event Time | Event Type | Event Details | Environment |
|---|---|---|---|
| 2013-04-24 13:35 EDT | INFO | Removed instance 'i-ceef10af' from your environment. (Reason: Instance is in 'terminated' state) | smblog-prod |
| 2013-04-24 13:29 EDT | INFO | Adding instance 'i-b74544d7' to your environment. | smblog-prod |
| 2013-04-24 13:28 EDT | INFO | Added EC2 instance 'i-b74544d7' to Auto Scaling Group 'awseb-e-vhxmdcz24q-stack-AWSEBAutoScalingGroup-1FLFG44H0X5GF'. | smblog-prod |
| 2013-04-24 13:22 EDT | INFO | Adding instance 'i-ceef10af' to your environment. | smblog-prod |
| 2013-04-24 13:21 EDT | INFO | Added EC2 instance 'i-ceef10af' to Auto Scaling Group 'awseb-e-vhxmdcz24q-stack-AWSEBAutoScalingGroup-1FLFG44H0X5GF'. | smblog-prod |
| 2013-04-09 15:29 EDT | INFO | Removed instance 'i-1672ee77' from your environment. (Reason: Instance is in 'terminated' state) | smblog-prod |
| 2013-04-09 15:23 EDT | INFO | Adding instance 'i-1672ee77' to your environment. | smblog-prod |
| 2013-04-09 15:22 EDT | INFO | Added EC2 instance 'i-1672ee77' to Auto Scaling Group 'awseb-e-vhxmdcz24q-stack-AWSEBAutoScalingGroup-1FLFG44H0X5GF'. | smblog-prod |
| 2013-04-08 02:30 EDT | ERROR | Stack deletion failed: The following resource(s) failed to delete: [AWSEBSecurityGroup]. | smblog-stage |
| 2013-04-08 02:30 EDT | INFO | Deleted load balancer named: awseb-e-d-AWSEBLoa-1QN62HK8X2PO4 | smblog-stage |

**smblog Environments**

**smblog-prod**
Successfully running version **version_1.1.**

▶ Environment Details