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

<table>
<thead>
<tr>
<th>Website</th>
<th>Popularity (unique visitors)</th>
<th>Frontend (Client-side)</th>
<th>Backend (Server-side)</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google.com</td>
<td>1,000,000,000</td>
<td>JavaScript, Ajax</td>
<td>C, C++, Go, Java, Python, PHP</td>
<td>BigTable</td>
</tr>
<tr>
<td>Facebook.com</td>
<td>880,000,000</td>
<td>JavaScript, Ajax</td>
<td>PHP, C++, Java, Python, Erlang</td>
<td>MySQL</td>
</tr>
<tr>
<td>YouTube.com</td>
<td>800,000,000</td>
<td>Flash, JavaScript</td>
<td>C, Python, Java</td>
<td>MySQL</td>
</tr>
<tr>
<td>Yahoo</td>
<td>590,000,000</td>
<td>JavaScript, Ajax</td>
<td>PHP</td>
<td>MySQL</td>
</tr>
<tr>
<td>Live.com</td>
<td>490,000,000</td>
<td>JavaScript, Ajax</td>
<td>ASP.NET</td>
<td>Microsoft SQL Server</td>
</tr>
<tr>
<td>MSN.com</td>
<td>440,000,000</td>
<td></td>
<td>ASP.NET</td>
<td>Microsoft SQL Server</td>
</tr>
<tr>
<td>Wikipedia.org</td>
<td>410,000,000</td>
<td></td>
<td>PHP</td>
<td>MySQL</td>
</tr>
<tr>
<td>Blogger</td>
<td>340,000,000</td>
<td></td>
<td>Python</td>
<td>BigTable</td>
</tr>
<tr>
<td>Bing</td>
<td>230,000,000</td>
<td>JavaScript</td>
<td>ASP.NET</td>
<td>Microsoft SQL Server</td>
</tr>
<tr>
<td>Twitter.com</td>
<td>160,000,000</td>
<td></td>
<td>C++, Java, RoR, Scala</td>
<td></td>
</tr>
<tr>
<td>Wordpress.com</td>
<td>130,000,000</td>
<td>JavaScript</td>
<td>PHP</td>
<td>MySQL</td>
</tr>
<tr>
<td>Amazon.com</td>
<td>110,000,000</td>
<td></td>
<td>Java, J2EE, C++, Perl</td>
<td></td>
</tr>
<tr>
<td>eBay.com</td>
<td>88,000,000</td>
<td>Java, WebSphere, Servlets</td>
<td>Oracle Database</td>
<td></td>
</tr>
<tr>
<td>Linkedin.com</td>
<td>80,000,000</td>
<td>Java, Scala</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
3 inevitabilities we design for:

1. Things break, unexpectedly
2. What we're building changes
3. We don't get to start over
Usual Dev to Production Pipeline

- Local Dev Env 1
- Local Dev Env 2
- Local Dev Env 3
- Local Dev Env N
- Stage Env 1
- Stage Env 2
- Prod Env 1

...
Scalability?

Local Dev Env 1
Local Dev Env 2
Local Dev Env 3
...  
Local Dev Env N

Stage Env 1
Stage Env 2

Prod Env 1

How to make it scalable?
# Amazon Elastic Beanstalk Useful

## Amazon Web Services

### Compute & Networking
- Direct Connect
- EC2
- Elastic MapReduce
- Route 53
- VPC

### Storage & Content Delivery
- CloudFront
- Glacier
- S3
- Storage Gateway

### Database
- DynamoDB
- ElastiCache
- RDS
- Redshift

### Deployment & Management
- CloudFormation
- CloudWatch
- Data Pipeline
- Elastic Beanstalk
- IAM
- OpsWorks

### App Services
- CloudSearch
- Elastic Transcoder
- SES
- SNS
- SQS
- SWF
Load Testing

Load test: SRM Blog Test 100 SBU 5 mins

|-----------------------|--------------|---------------------------------------------------|------------------------------------------------------------------|

Choose from which load zone to display data.

<table>
<thead>
<tr>
<th>SBUs active</th>
<th>Connections active</th>
<th>Bandwidth</th>
<th>Data received</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0 bit/s</td>
<td>806.57 MiB</td>
<td>32828 (0 req/s)</td>
</tr>
</tbody>
</table>

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.


Map data ©2013 Google - Terms of Use Report a map error
Elastic Beanstalk

### Elastic Beanstalk Application Details

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

<table>
<thead>
<tr>
<th>Event Time</th>
<th>Event Type</th>
<th>Event Details</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-04-24 13:35 EDT</td>
<td>INFO</td>
<td>Removed instance '1-ceef10af' from your environment. (Reason: Instance is in 'terminated' state)</td>
<td>smblogger-prod</td>
</tr>
<tr>
<td>2013-04-24 13:29 EDT</td>
<td>INFO</td>
<td>Adding instance '1-b74544d7' to your environment.</td>
<td>smblogger-prod</td>
</tr>
<tr>
<td>2013-04-24 13:28 EDT</td>
<td>INFO</td>
<td>Added EC2 instance '1-b74544d7' to Auto Scaling Group 'awseb-e-vhmdcz24q-stack-AWSEBAutoScalingGroup-1FLFG44H0X5GF'.</td>
<td>smblogger-prod</td>
</tr>
<tr>
<td>2013-04-24 13:22 EDT</td>
<td>INFO</td>
<td>Adding instance '1-ceef10af' to your environment.</td>
<td>smblogger-prod</td>
</tr>
<tr>
<td>2013-04-24 13:21 EDT</td>
<td>INFO</td>
<td>Added EC2 instance '1-ceef10af' to Auto Scaling Group 'awseb-e-vhmdcz24q-stack-AWSEBAutoScalingGroup-1FLFG44H0X5GF'.</td>
<td>smblogger-prod</td>
</tr>
<tr>
<td>2013-04-09 15:29 EDT</td>
<td>INFO</td>
<td>Removed instance '1-1672ee77' from your environment. (Reason: Instance is in 'terminated' state)</td>
<td>smblogger-prod</td>
</tr>
<tr>
<td>2013-04-09 15:23 EDT</td>
<td>INFO</td>
<td>Adding instance '1-1672ee77' to your environment.</td>
<td>smblogger-prod</td>
</tr>
<tr>
<td>2013-04-09 15:22 EDT</td>
<td>INFO</td>
<td>Added EC2 instance '1-1672ee77' to Auto Scaling Group 'awseb-e-vhmdcz24q-stack-AWSEBAutoScalingGroup-1FLFG44H0X5GF'.</td>
<td>smblogger-prod</td>
</tr>
<tr>
<td>2013-04-08 02:30 EDT</td>
<td>ERROR</td>
<td>Stack deletion failed: The following resource(s) failed to delete: [AWSEBSecurityGroup].</td>
<td>smblogger-stage</td>
</tr>
<tr>
<td>2013-04-08 02:30 EDT</td>
<td>INFO</td>
<td>Deleted load balancer named: awseb-e-d-AWSEBLoa-1QN62HK8X2PO4</td>
<td>smblogger-stage</td>
</tr>
</tbody>
</table>