Why and How to Leverage Amazon Cloud Services to Deploy JupyterHub at Scale?

Demba Ba, Harvard University

Assitant Professor of Electrical and Bio Engineering

New York, NY, August 24th 2017
The “What?”, The “How?”, and The “Why?”
Bridging the gap between EE and CS education

**What:** Design of two courses that bridge gap between electrical engineering and computer science using Jupyter and data-science “friendly” devices (e.g. wearables)/assignments.

**Why:** “So what you are saying is that you wish you had taken your class from yourself?”

**How:** Integration of theory and computation + **Design of seamless coding interface**.

**Why:** “So what you are saying is that you wish you had taken your class from yourself?”.
The “What?”

ES155 and ES201 at Harvard University
ES155: Labz ‘N Da Wild

- **Topics**: Signals/systems, probability, time-series and filtering

- **Goal**: To empower student to collect, process and make decisions using their own data.

- **Philosophy**: full vertical integration of data processing in all Problem Sets, Labs, and Final project
  - Biological Signal Dataset (e.g. Blood Volume Pulse)
  - Data Collection with the Empatica E4
  - Data Processing with Jupyter
  - External action through IOT
ES201: Rigorous Data Science

- **Topics**: Regression, classification, **optimization and numerics**, state-space models, kernels, graphical models

- **Goal**: Understand and apply sophisticated statistical techniques

- **Philosophy**: integration of theory and data processing
  - Problem sets
    - Theory
    - Real-word (as opposed to toy) data processing
  - Final project
    - Question formulation
    - Data acquisition
    - Data analysis with Jupyter
The “How?”

Jupyter Notebooks on AWS
<table>
<thead>
<tr>
<th><strong>Version 1.0: Fall 2016 (ES 155)</strong></th>
<th><strong>Version 2.0: Spring 2017 (ES 201)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group AWS Resources</strong></td>
<td><strong>Group AWS Resources (cluster manager)</strong></td>
</tr>
<tr>
<td>m4.large INS ($180/month)</td>
<td>t2.medium CM 4GB/2vCPUs $35/month</td>
</tr>
<tr>
<td>16GB/4vCPUs</td>
<td></td>
</tr>
<tr>
<td># of students per mx4.large</td>
<td># of students</td>
</tr>
<tr>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>Resources per student</td>
<td>Resources per student</td>
</tr>
<tr>
<td>Docker container (1GB RAM)</td>
<td>t2.medium INS 4GB/2vCPUs</td>
</tr>
<tr>
<td>Up/Down scaling</td>
<td>Avg usage per month per student</td>
</tr>
<tr>
<td>No</td>
<td>30 hours</td>
</tr>
<tr>
<td>Monthly cost per student</td>
<td>Up/Down scaling</td>
</tr>
<tr>
<td>$15</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Monthly cost per student</td>
</tr>
<tr>
<td></td>
<td>$3</td>
</tr>
</tbody>
</table>
Example ES155 Project

Wearable Sleeve to Predict Free-throw Shots
“Wearable Sleeve to Predict Free-Throw Shots” *(Ryan Halvorson)

FLORA 9 DoF Sensors
- Accelerometer, gyroscope, magnetometer

Adafruit Feather M0
- Bluefruit LE
  - Bluetooth enabled microcontroller

MyoWare EMG Sensors with Amplifier
- High signal to noise ratio

Elastic Arm Sleeve
- Tight fit to maintain sensor position, without getting in the way of shooting

Bluetooth and SD
- Fast data transmission rate allows for communication with data processing software for logging or streaming

Quantitative Output
- Shot Detection, Shot Counter, Shot Velocity, Release Angle, Free Throw Classification
Other project topics

- **ES155**
  - Student stress monitoring
  - Sleep-quality assessment
  - Assessment of table tennis Performance

- **ES201**
  - Fantasy football team recommender
  - Topic popularity and connections assessment on Twitter
  - Modeling wolverine density in Norway
Course Management
Bitbucket Teams

ES155 Team

Course Development
- Students have no access

Homework
- Students have read and write access

Labs

Course Materials
- Students have read access only
The “Why?”

Student: “So you would have loved to take your own course?”

Me: “Yes, absolutely!”
Concluding remarks

● In the future, facility with data manipulation is going to be part of literacy
  ○ Teaching data-science to underserved communities (collaboration with Boston Public Schools)

● Potential impact on education
  ○ Data-centered teaching: e.g. government, journalism
  ○ Cloud-based platform for teaching in the developing world

● Key lesson learned: given these tools, students just want to be left alone!
Acknowledgements

Thank you!!

Demba Ba

Yingzhuo (Diana) Zhang

Harvard SEAS Active Learning Labs

Harvard SEAS Computing

Faras Sadek

Jim Waldo

Yasha Iravantchi