

Wearable Signal Processing Using Docker Notebook Containers on AWS

Faras Sadek, Yasha Iravantchi, Diana Zhang and Demba Ba

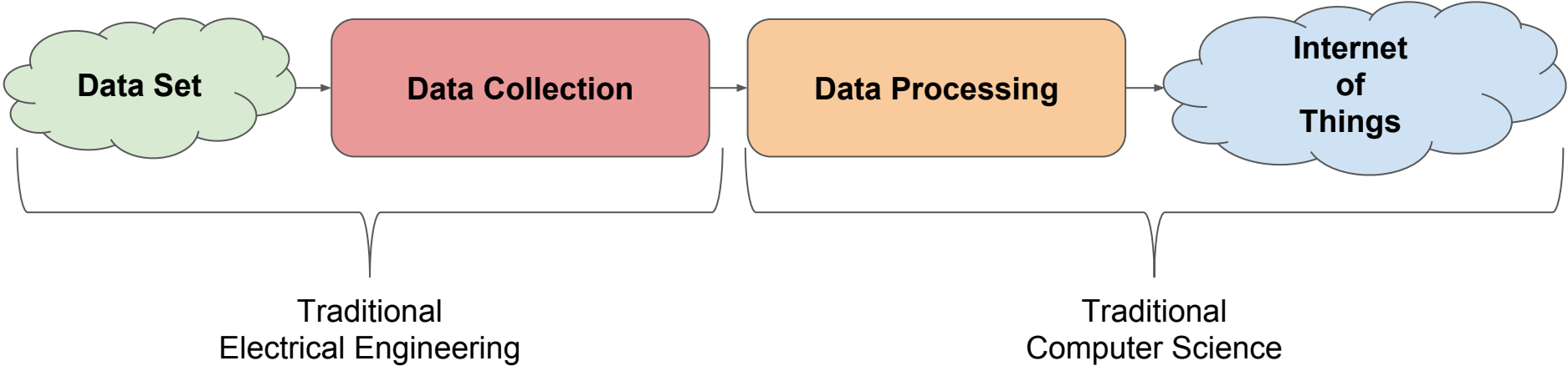
March 17, 2016



HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

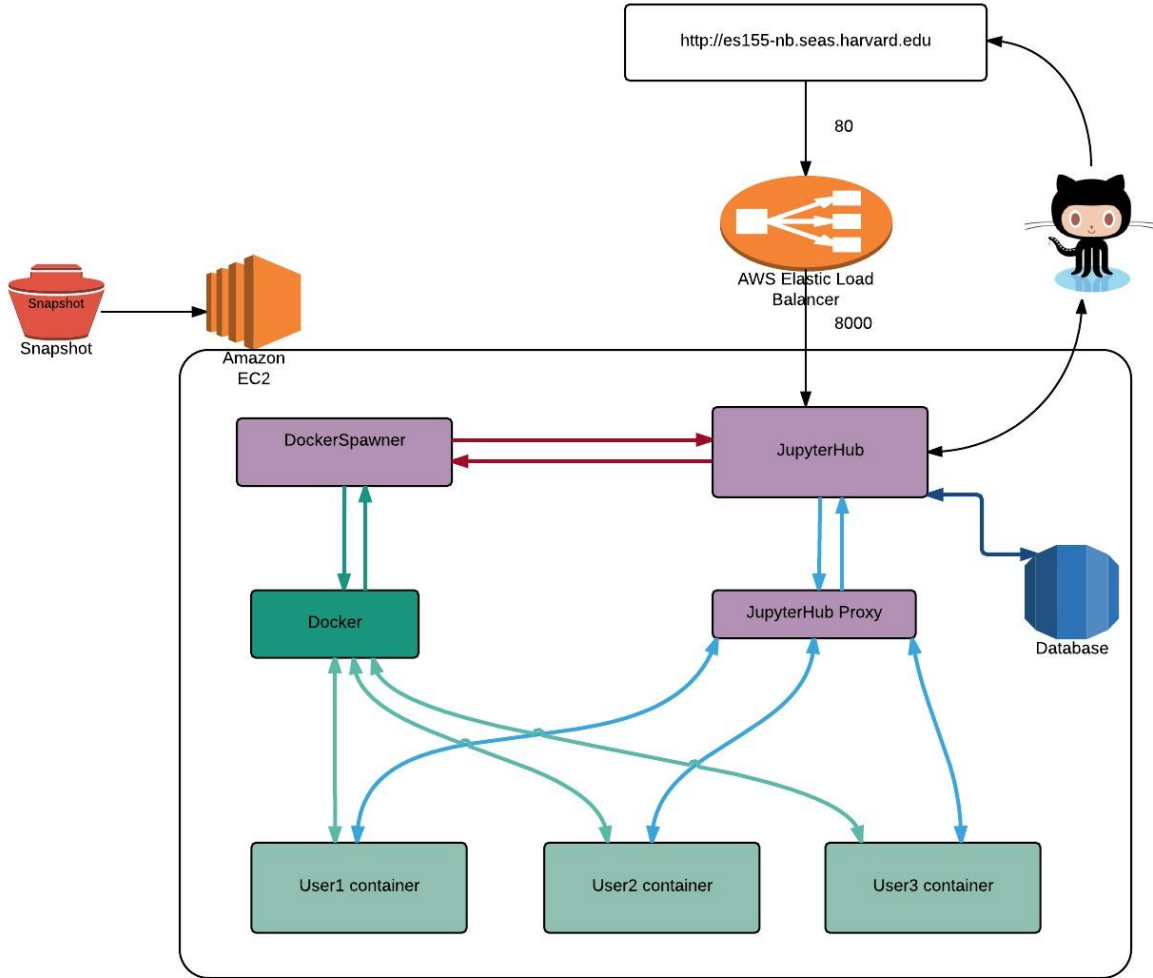


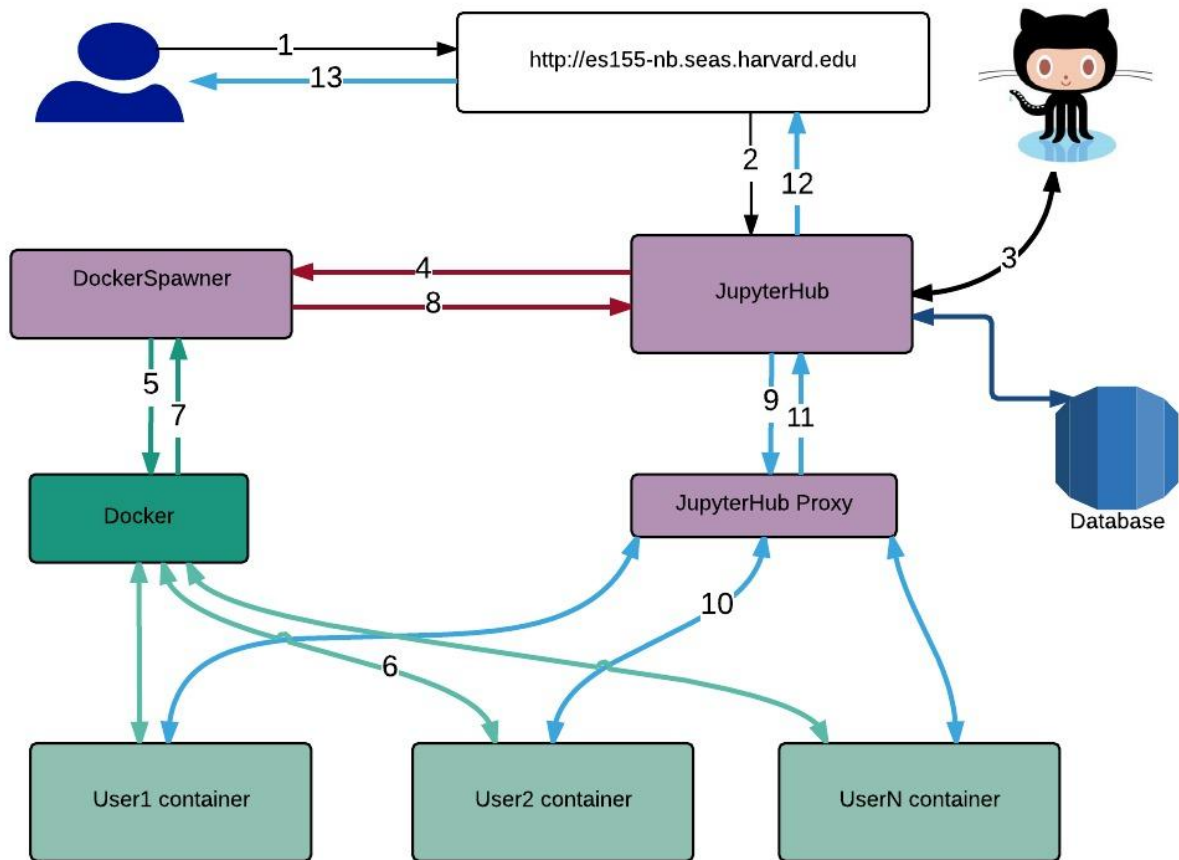
ES155: Bridging the gap between EE and CS education



ES155 bridges the gap between electrical engineering and computer science by using Jupyter and data-science “friendly” devices like the Empatica E4.

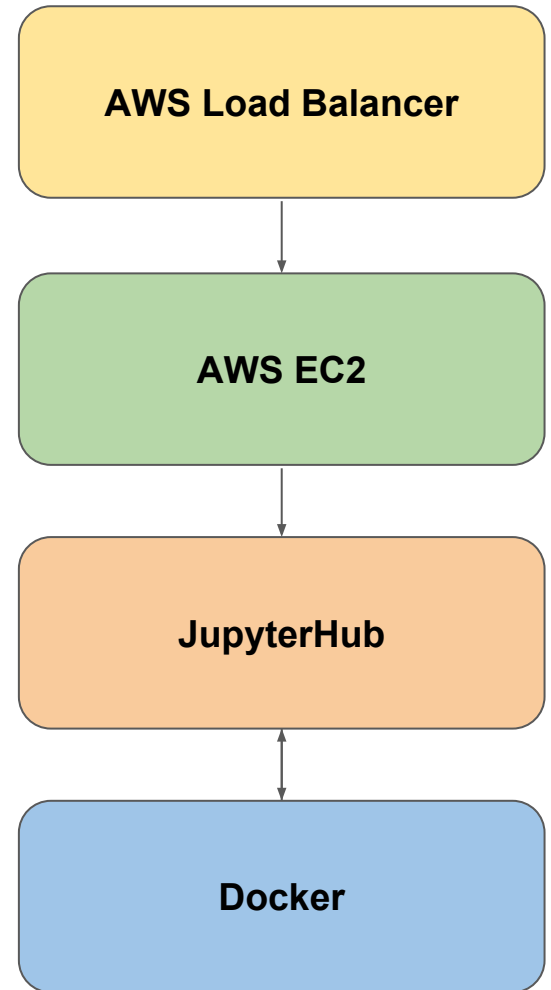
The Backend/Infrastructure





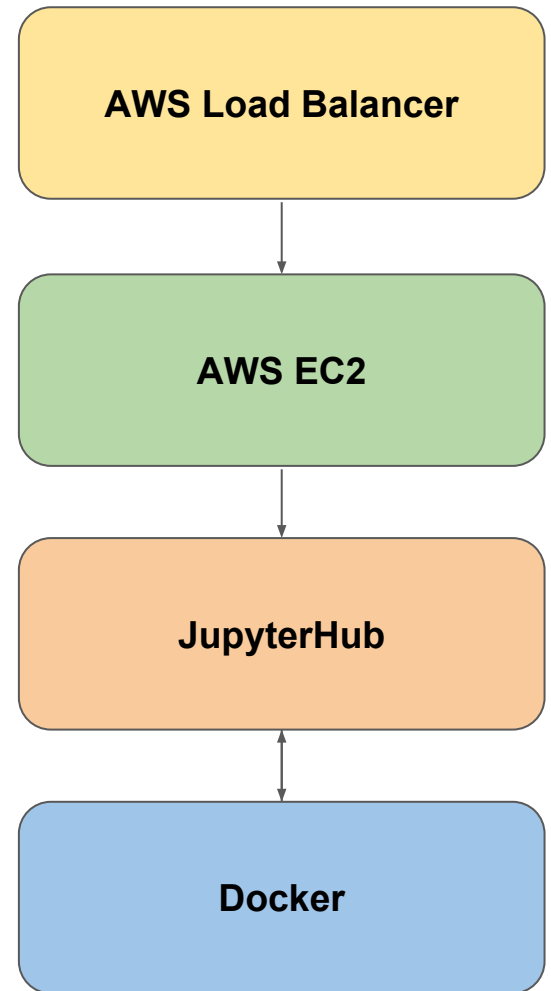
Our Implementation of JupyterHub

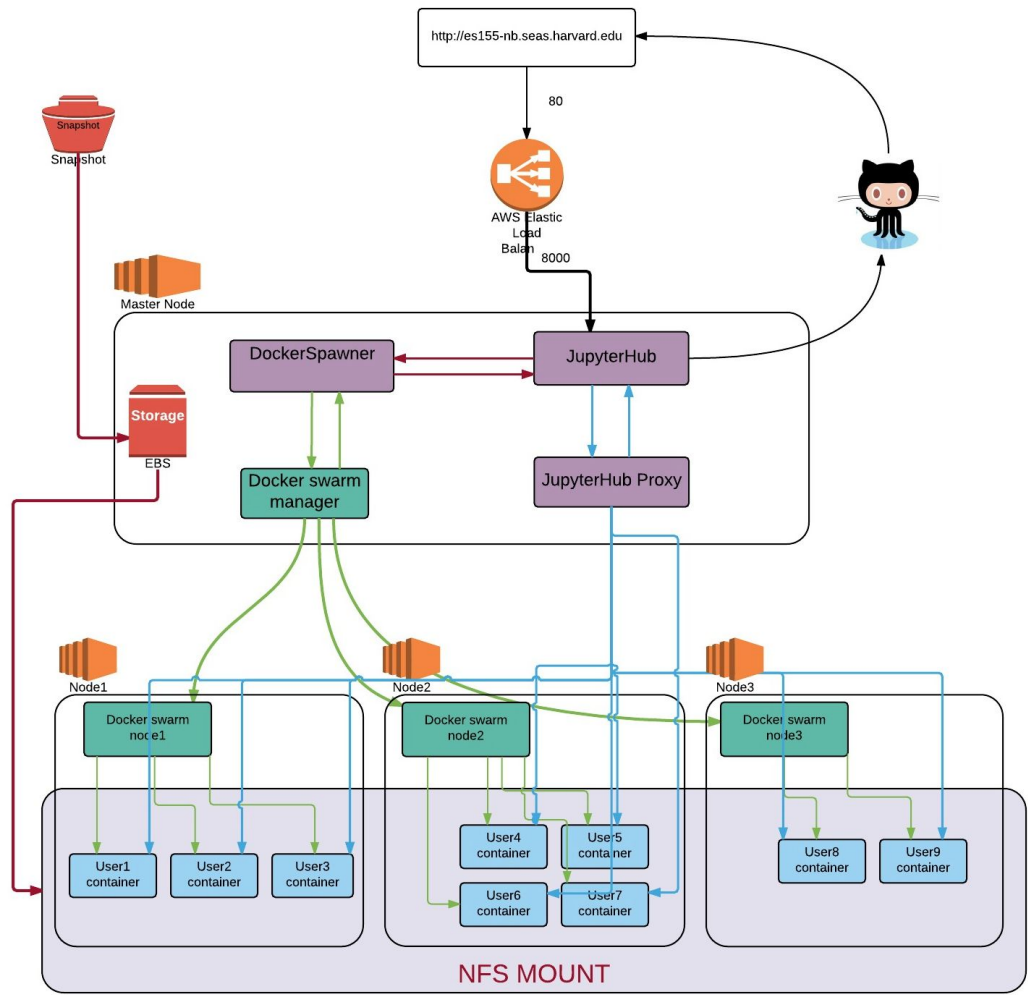
- Load balancer picks request on port 80 and redirects to EC2 instance on port 8000, where JupyterHub listens.
- GitHub as an auth schema for student to login to to the server
- If username in database, Jupyter will spawn a Docker container if one does not exist for that user
- If the user has a container, JupyterHub forwards the user to that container through the proxy



Security, Scalability, and Cost

- JupyterHub server located behind Amazon AWS
- AWS Instance located in private VPC
- SSH access restricted to Harvard Admin VPN only
- EC2 instance type: m4.xlarge for 13 students with 1GB per student
- Average cost per student: \$13 a month
- Horizontal scalability can be achieved using Docker Swarm

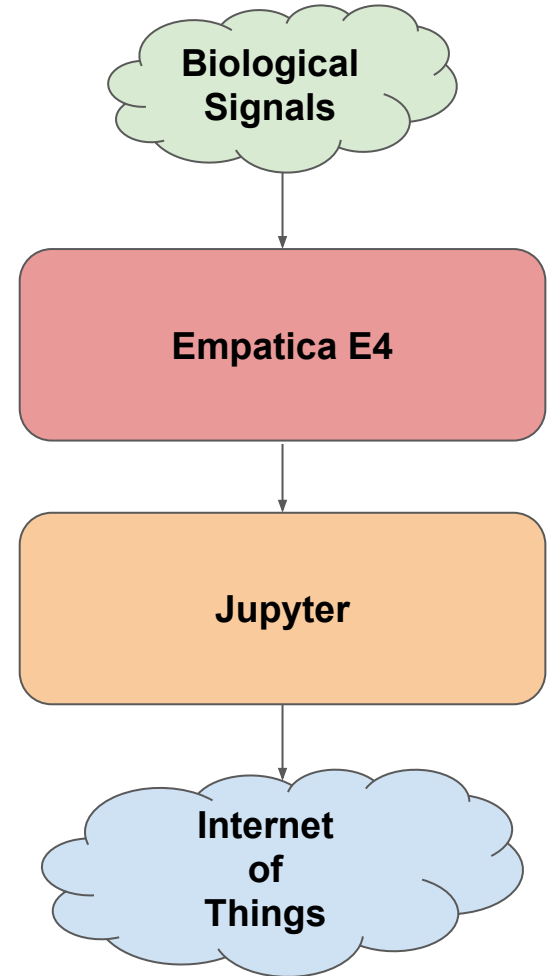




“Labs in the Wild”

ES155: Labs in the Wild

- ES155 presents full vertical integration of data processing
 - Biological Signal Dataset (e.g. Blood Volume Pulse)
 - Data Collection with the Empatica E4
 - Data Processing with Jupyter
 - External action through the Internet of Things (e.g. Tweet when heart rate is high)
- Labs in the Wild allow students to collect their own data, process it themselves, and perform their own functions and analysis.



Empatica E4: “Mercedes Benz of wearables”

- Continuous, real-time data acquisition in daily life.
- Acceleration, blood volume pulse, electrodermal activity, and temperature.
- Offline or stream data directly to cloud via a smartphone (iPhone/Android)
- APIs for real-time data access, or post processing

- This combination of sensors allows students to perform interesting analyses: gait recognition, sleep pattern recognition, “lie detector” tests

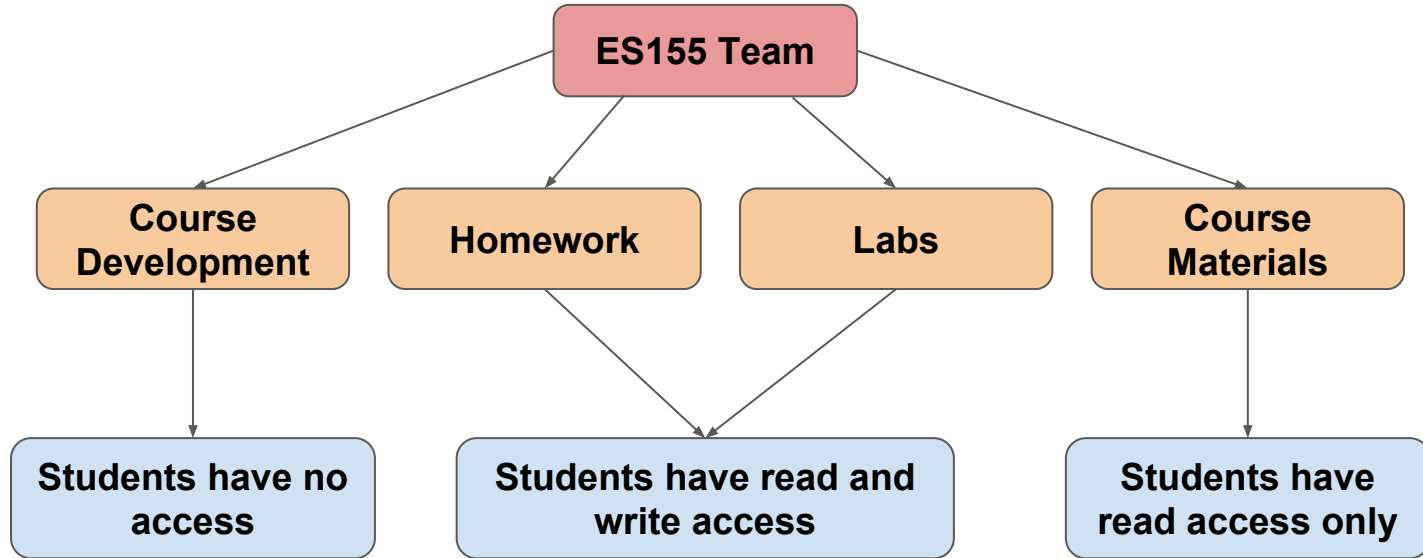


Course Management

Version control

- Why Git?
 - Version control
 - Students can retrieve their work in case of server/internet issues
 - Collaboration on projects
 - We want the students to get into a habit of using Git to communicate their work with each other
- Why Bitbucket?
 - Bitbucket provides university students with unlimited private repositories.

Bitbucket Teams



Live Demo and Example Lab

Concluding remarks

- Jupyter-powered platform that enables the concept of a “Labs in the Wild”
- In the future, facility with data manipulation is going to be part of literacy
- Potential impact on education
 - Data-centered teaching: e.g. government, journalism
 - Cloud-based platform for teaching in the developing world
- Collaborative data science
 - A github for data science?

Acknowledgements

- ES 155 Team
- Active Learning Labs team at Harvard SEAS
- Harvard SEAS Academic Computing
- JupyterDays organizer and Jupyter Project team

Thank you!!