JupterHub for OER

Ewan Klein, Francisco Vargas, Ben Butchart and Gavin Gray

27th November 2015

Abstract

JupyterHub provides a multi-user web-based interactive notebook for groups. One of the main applications for this is in teaching, making programming accessible for anyone with a browser. We describe our experiences deploying this in a real educational environment on internal hardware.

In order to help students with little or no programming experience to become comfortable with basic data analysis and visualisation tasks, we have been developing course content with IPython (Pérez and Granger, 2007) and JupyterHub (JupyterHub, n.d.). This approach has several attractive features:

- it offers a browser-based 'notebook' with support for rich text, executable code blocks and interactive data visualisation;
- it allows students to receive the same learning materials but to
- experiment with them on an individual basis within a standardised environment;
- the components are all open source and actively maintained; since the
- platform is made available as a service, students are not required to download and install the necessary software.

One of the most attractive aspects of Jupyter notebooks is the possibility of displaying inline data visualisation. We have chosen to address this area using two relatively new Python libraries, Bokeh and Folium.

Bokeh (Bokeh Development Team, 2014) is a browser-based interactive visualization library that builds on the success of Javascript-based tools such as D3.js (Bostock et al., 2011). It is designed to allow beautiful and versatile plots to be constructed on the basis of simple commands and aims to be scalable over big and streaming data sets. Using Bokeh, we have built sample notebooks which illustrate a variety of visualisations and plot types, including heatmaps, scatter plots, timeseries and binary trees.

Folium (Folium, n.d.) is a Python interface to leaflet.js, a widely-used interactive Javascript mapping library and allows us to create attractive map visualizations. So far, we have used Folium for plotting geocoded data elements as markers on a map; and choropleth maps and Voronoi tessellations to visualise location-related datasets.

We are using Docker (Merkel, 2014) to power our JupyterHub service, following Jessica Hamrick's JupyterHub deployment for Computational Models of Cognition. However, rather than adopting an elastic cloud framework, we decided to start with a single server, and to postpone scaling up issues to later.

Docker allowed us to rapidly prototype the server configuration, and to keep the configuration hierarchical. We have built two server configurations. The first is a standard, centralised JupyterHub instance. Each student has a persistent account with storage, and server instances will persist. Docker volumes are used to manage user home directories, allowing for seamless changes to apply to the entire operating system. These are also used to link to datasets that the students can access at a shared location inside the container. In addition, we have built a temporary notebook server which serves each new visitor a new notebook instance that is removed after the user terminates their session.

JuyterHub has an extendable authentication architecture that allows deployments to override default

authentication strategies with custom Authenticator modules. To support institutional access using a single sign-on service based on the CoSign (CoSign, n.d.) system, we are deploying JupyerHub behind an Apache proxy server, allowing us to use well-understood Apache CoSign modules. The Apache server provides a public-facing SSL port, which redirects requests to the JupyterHub server instance running on a private port on the same virtual machine.

Bokeh Development Team (2014) *Bokeh: Python library for interactive visualization*. Available from: http://www.bokeh.pydata.org.

Bostock M, Ogievetsky V and Heer J (2011) D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, Piscataway, NJ, USA: IEEE Educational Activities Department, 17(12), 2301–2309, Available from: http://dx.doi.org/10.1109/TVCG.2011.185.

CoSign (n.d.) http://weblogin.org/.

Folium (n.d.) https://github.com/python-visualization/folium.

JupyterHub (n.d.) https://github.com/jupyter/jupyterhub.

Merkel D (2014) Docker: Lightweight linux containers for consistent development and deployment. *Linux J.*, Houston, TX: Belltown Media, 2014(239), Available from: http://dl.acm.org/citation.cfm?id=2600239.2600241.

Pérez F and Granger BE (2007) IPython: A system for interactive scientific computing. *Computing in Science and Engineering*, IEEE Computer Society, 9(3), 21–29, Available from: http://ipython.org.