

Scaling Reproducible Research with Jupyter

Carol Willing
Project Jupyter
San Diego, USA
0000-0002-9817-8485

Abstract—Jupyter Notebooks have taken the scientific and open data world by storm the past five years. Being able to tell a computational narrative that combines prose, code, media, and rich visualizations have increased a researcher’s ability to collaborate with others, share research in a reproducible way, and educate others in their scientific discipline and beyond.

A suite of tools, processes that scale, and modern ways to communicate openly about scientific research have grown rapidly within Project Jupyter’s open source community. Beyond the Jupyter Notebook, open source projects, including JupyterLab, JupyterHub, Binder, and nteract’s Papermill, offer new pipelines and services to allow open research to scale and impact others on a global scale.

Index Terms—Jupyter Notebook, JupyterHub, Binder, Papermill, Reproducible Research

I. INTRODUCTION

Scientific research in many disciplines relies heavily on computation. Through its tools, processes, and collaborative communication, Project Jupyter enables computational narratives which open the door to more effective cross-discipline work and reproducible research.

II. TOOLS

A. Jupyter Notebook

The Jupyter Notebook is a document composed of a sequence of code cells and markdown cells. This structure allows authors to interleave formatted narrative and mathematics with blocks of code and their rich outputs, rendering the notebook document a powerful tool for communicating insights and results.

B. JupyterLab

JupyterLab is a next-generation web-based user interface for Project Jupyter. JupyterLab enables you to work with documents and activities such as Jupyter notebooks, text editors, terminals, and custom components in a flexible, integrated, and extensible manner. Documents and activities integrate with each other, enabling new workflows for interactive computing.

III. PROCESSES

The ability to serve data at scale is critical for higher education and scientific research. Modern tools that can run in the cloud or on-premise at a research center are needed to develop processes to support reproducible research.

A. JupyterHub

First released in March 2015, JupyterHub has grown in scope and serves research centers, universities, and scientists throughout the world. JupyterHub is the best way to serve the Jupyter notebook for multiple users.

B. Papermill

Papermill, an nteract project, is a library for parameterizing, executing, and analyzing Jupyter notebooks. With it, you can spawn multiple notebooks with different parameter sets and execute them concurrently. Papermill can also help collect and summarize metrics from a collection of notebooks.

IV. COLLABORATIVE COMMUNICATION

The beauty of Jupyter has been the creative ways that it has enabled communication among researchers, students, and citizen scientists.

About two years ago, the Binder project evolved into the community led project. The deployment at mybinder.org was upgraded to use BinderHub, a scalable open-source web application that runs on Kubernetes and provides free, sharable, interactive computing environments to people all around the world.

People use the public deployment at mybinder.org around 100,000 times each week, and there are around 8,000 unique repositories compatible with Binder.

V. CONCLUSION

Research will benefit for years to come from the Jupyter ecosystem: its tools, processes, and collaborative communication. Jupyter empowers its users to share open science, knowledge, and discoveries.

ACKNOWLEDGMENT

Thanks to the Project Jupyter team and its contributors. I wish to recognize the work of Min Ragan-Kelley who led the JupyterHub and Binder projects. Project Jupyter grant funders are also generously thanked for their vision and support.

REFERENCES

- [1] M. Ragan-Kelley, C. Willing, and J. Grout “Jupyter: Tools for the Life Cycle of a Computational Idea,” Siam News, March 1, 2018.
- [2] Project Jupyter, JupyterLab documentation, web, November 2019.
- [3] Project Jupyter, “Announcing JupyterHub 1.0,” Project Jupyter blog, May 2019.
- [4] Project Jupyter, JupyterHub documentation, web, November 2019.
- [5] Project Jupyter, “The International Binder Federation,” Project Jupyter blog, July 2019.