

David Bond  
Mentored by Mary Thomas

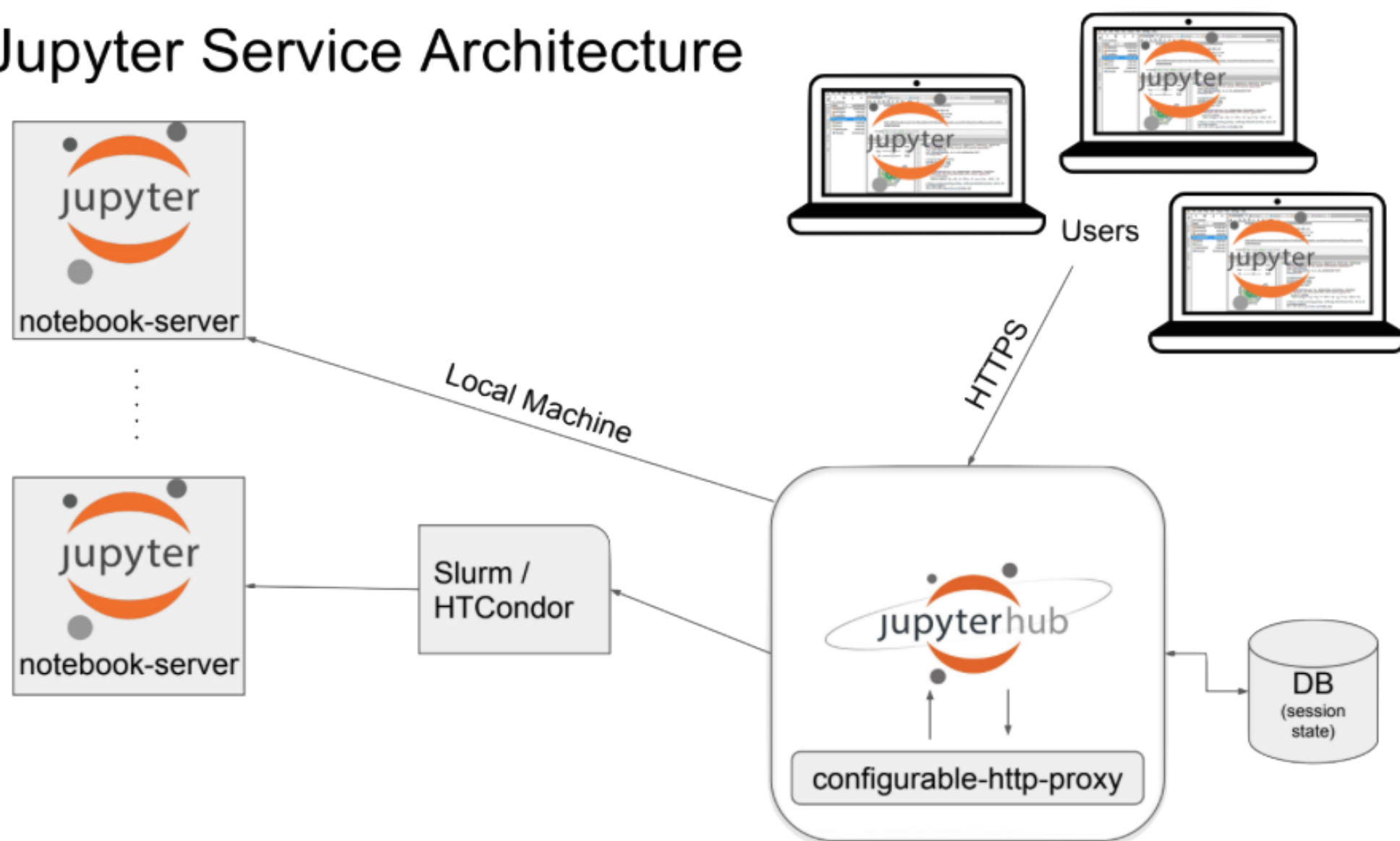
## Abstract

The goal of this project is to ensure that the example Jupyter Notebooks on the Expanse supercomputer run properly. Expanse is complex and constantly evolving, so the programs made for it have to be maintained over time. To ensure the notebooks run properly on Expanse, every notebook has to be individually tested. Any notebook that no longer runs properly has to be altered. These notebooks are used by users of Expanse to ensure everything is running properly or to learn about new concepts. As a result of the project, we now know what notebooks run, what notebooks requires changes, and what notebooks should be deprecated.

## Research & Development Process

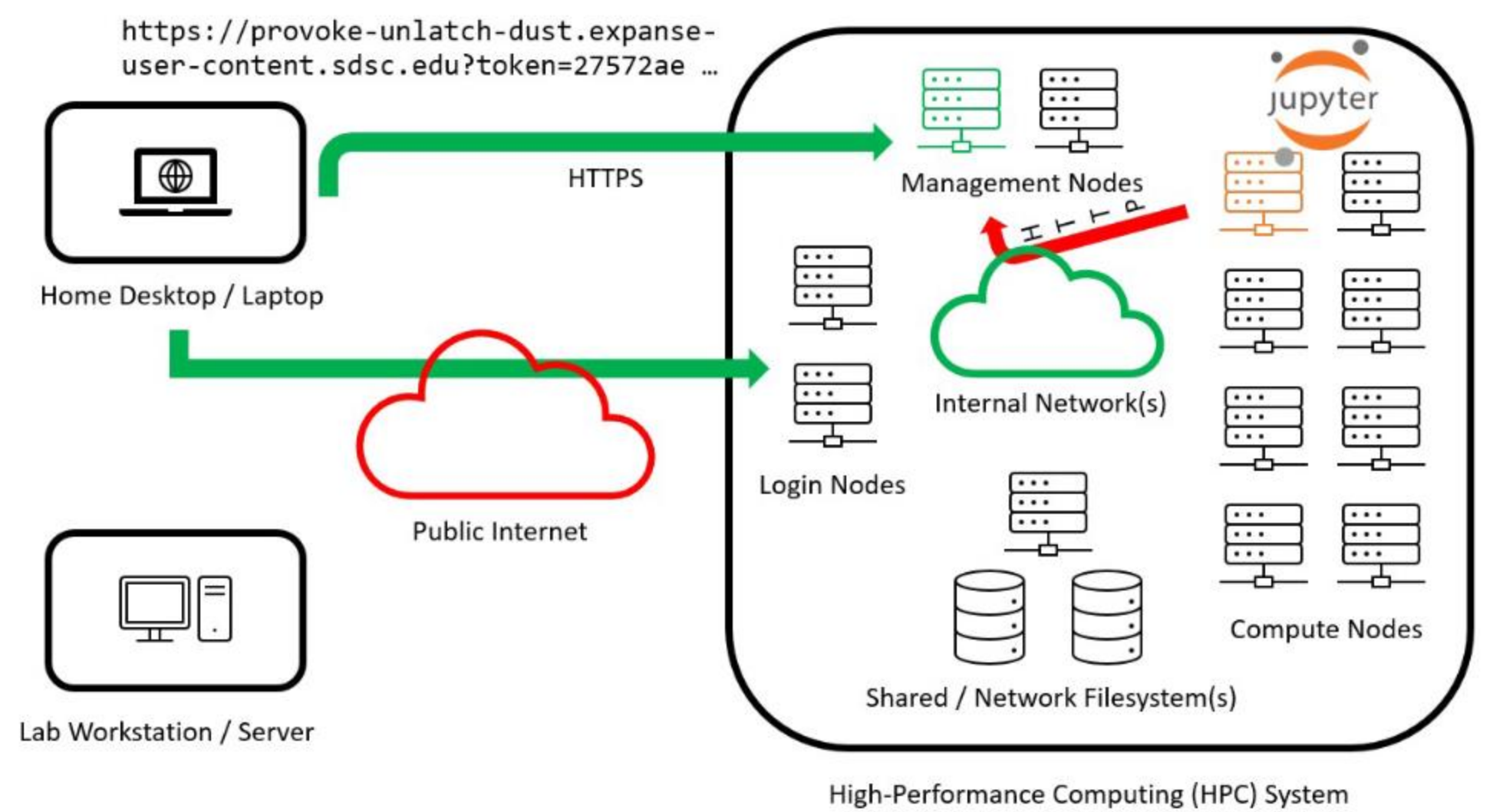
In order to test the notebooks, I had to learn how to run and edit them securely. I began by learning how to programs on Expanse. This required me to learn how to create batch jobs. I then learned how to run notebooks on Expanse securely using Galileo and Jupyter Lab. I had to create a GitHub repository to back up my progress while keeping my changes independent of the production notebooks. Everything was set up to test the notebooks. First, I picked a notebook. It was run, and I took notes. If the notebook could not be run, I attempted to make changes to fix the issues. This was repeated until all the notebooks were tested.

## Jupyter Service Architecture



Jupyter starts up a server, which is the development environment that lets users run notebooks, a program that combines code, output, and documentation into one file. The server can optionally use a proxy to make connecting to the server more secure. Users can then use a standard web browser to connect to the server.

## Galyleo Architecture



Jupyter Lab is, by default, not secure, because it uses HTTP instead of HTTPS. Galyleo is used to fix that vulnerability. A user first logs into a Login Node and run the Galyleo command. Galyleo starts Jupyter Lab on a compute node. Then, inside of the Expanse's internal network, a reverse proxy service is started in a management node. Then, users can connect to the management node using a link that is sent to them through the Login Node, which uses HTTPS. That link allows a user to connect to Jupyter Lab.

### Technologies Used & Learned

- Code Development
  - Python
  - GitHub Repositories
  - Conda Environments
  - Singularity Containers
- Using Supercomputer
  - SSH
  - Batch Scripting
  - CPU, GPU
  - Galileo
- Jupyter Services
  - Jupyter Notebooks
  - Using Kernels (Python, Bash)
  - Jupyter Lab

### Table of Notebooks Tested

Title	Status	Short Description	Issues and Changes
Data_Analysis	Done	Notebook that shows how to use Pandas, a Python data analysis library	
Clustering_Visualizations	Done	Notebook that shows basic clustering methods using Python Packages	
Hello_World	Done	Notebooks that display hello world and GPU/CPU information	
Matplotlib_Intro	Done	Notebook that shows how to create graphs and visualize data in Python using Matplotlib	
Notebook_Dev_Basics	Done	Notebooks that show how to use Markdown and LaTeX in Jupyter Notebooks	
NumPy_Intro	Done	Notebook that shows how to use NumPy, a Python package for scientific computing	
Python_Data_Analysis_Library	Done	Notebook that covers the essentials of Pandas, a Python package that data analysis, and CSV files	
CUDA_GPU_NVidia	Fixed	Notebooks that have problems for using CUDA and the GPU in Python and their solutions	Instructions in README relied on files outside of collection, files were added and instructions were changed
Image_Processing	Fixed	Notebook that shows how to display and alter images in Python using Pillow.	Two identical copies of files are in collection, one copy was deleted. Notebook required changes to display images in Jupyter Lab
Graphs&Networks	Changes needed	Notebook that shows how to create and show graphs and networks	Some packages are conflicting on Expanse
PythonBasics	Changes needed	Notebooks that covers the basics of Python and strings	Code that detects if a GPU is present needs to be rewritten
PythonSeries-master	Changes needed	Notebooks that covers decision trees, parallel programming, and regression in Machine learning	Some packages are not supported on Expanse. Useless files were deleted
Tensorflow	Changes needed	Notebooks that use Tensorflow for simple training and image classification	Instructions in README contains confusing information
deep_learning	Pending	Notebook that showcases how to run Tensorflow on Expanse	Requires edits to get notebook running on GPU. Notebook cannot be run all the way through
Boring_Python	Deprecated	A collection of notebooks that teach the user how to use Python. There are 9 notebooks, and each covers specific topics	Some packages are not supported on Expanse
Web_Scraping	Deprecated	Notebook that extracts data from websites	Relies on a website that no longer exists

## NumPy\_Intro

[illegible]

## Image\_Processing

[illegible]

## Conclusion

Over the course of my internship as SDSC, I had to learn more about High Performance Computing and editing Jupyter Notebooks securely. I had to overcome many unexpected challenges, like learning how to use the batch queue, using Conda environments, and figuring out how Galileo. In the future, the issues that I found and fixes I made will be reviewed others, who will implement the finished notebooks into Expanse. The main issue with maintaining the notebooks in the future will be making sure they work with the packages installed on Expanse, which are constantly updated.

## Acknowledgments

Ange Mason  
Marty Kandes

## References

1. Galyleo Architecture:  
[https://education.sdsc.edu/training/interactive/?id=202112\\_ExpandWebinar-M.Kandes](https://education.sdsc.edu/training/interactive/?id=202112_ExpandWebinar-M.Kandes)
2. Jupyter Architecture:  
[https://docs.google.com/presentation/d/15MXC8HcqEhfCMeLrKl\\_oGO\\_rtn81R0ifg\\_/edit?usp=sharing&oid=105889292217036541066&rtopof=true&sd=true](https://docs.google.com/presentation/d/15MXC8HcqEhfCMeLrKl_oGO_rtn81R0ifg_/edit?usp=sharing&oid=105889292217036541066&rtopof=true&sd=true)