

Review: “Interactive Supercomputing with Jupyter: Lessons Learned from the National Energy Research Scientific Computing Center”

John Shalf¹

¹Affiliation not available

January 22, 2021

This is a review of the article *“Interactive Supercomputing with Jupyter: Lessons Learned from the National Energy Research Scientific Computing Center,”* for the Leadership Computing department of *Computing in Science and Engineering*. The article was invited to complement a Special Issue on “Jupyter in Computational Science.”

Comments

In the abstract, the last sentence capitalizes “Jupyter Project” but then later says “engage with the Project”. I think the second project should be lower case or you could have a capital “Jupyter”, but project should not be capitalized.

Given this is the “leadership computing” column, could you also put in a connection between the scalability of leadership computing to enable the interactive computing with Jupyter? It is talking about this indirectly, but it would be nice to make that more explicit and up-front for the “leadership computing” column.

In section 1, it ends with the following sentence “Efforts to support Jupyter at NERSC began in the context of a tectonic shift in user workload driven by the explosion of data from...” This is a bit of a mouthful of a sentence. Can you break it into more bitesized pieces?

Maybe put second paragraph that introduces NERSC into the “how it started and how it’s going” section (seems to interrupt the flow between paragraph 1 and paragraph 3).

It would be nice to mention that the inventor of Jupyter notebooks and JupyterHub is in-house. Was there any recognition by him about these emerging use-cases that led him to develop JupyterHub as a new deployment mechanism (JupyterHub didn’t just appear... I think there was likely some motivation that led to its development to serve this emerging community??) So it would be nice if the story of how you came to deploy JupyterHub seemed more purposeful and driven by emerging community needs. I think there is a background story there from Fernando’s perspective that would show how the community needs drove to the Hub’s creation. (its outside of NERSC, but it is inside of UCB/LBL). It would show how research and deployment at NERSC work together.

In the Phase 3 discussion, the JupyterLab comes out of the blue. Is there some experience at NERSC or other data centers that fed into the design of JupyterLab? Should JupyterLab be the centerpiece of your Phase3 deployment? It seems to deserve its own phase (phase 4 or be the centerpiece of phase 3).

There is a lot of good discussion of the technical details of the deployment of Jupyter, which is interesting. It would be nice if you had more descriptions of the many different use cases and scientific activity that has been enabled by this infrastructure (opportunity for more pretty pictures and scientific impact). I would recommend even making your “lessons learned” section more compact and use the space to describe cool

science projects that benefit from this infrastructure and talk more about the scientific impact of this new paradigm of HPC access.