Systems-Facing call, 2022-06-16 @ 1p ET/ 12p CT/ 11a MT/ 10a PT

Topic: **JupyterHub @ NCAR: A Case for a Notebook-Centric Analysis Platform Around the Jupyter Ecosystem**

Topic Overview:

Computational, visualization, and data analysis systems have evolved such that web-based access capabilities are readily available with the native web browser acting as a primary client interface. A low-barrier entry point for new researchers that has a customizable, yet familiar interface and powerful extensibility broadens the ability to create innovative tools. In addition, NCAR has been working to expand on geoscience ∑specific tools such as GeoCAT and wrf-python through academic and community groups like Project Pythia and Pangeo. NCAR has deployed JupyterHub, a multi-user web framework that provides the ability to access various HPC platforms at NCAR as a Notebook centric mode of computing. This presentation will focus on the use case, co-design, and configuration of the JupyterHub deployment at NCAR as well as the most desirable JupyterLab and JupyterHub features in the future.

Presenter:
Jared Baker - HPC Systems Engineer III, National Center for Atmospheric Research (NCAR)

Slides & Links:
https://drive.google.com/file/d/14_f7hgQzc-N9mG6KEYjB1ylr1ftgy0Us/view?usp=sharing

YouTube Link To Meeting Recording: https://youtu.be/fRO4Lg-tIXI

CaRCC YouTube Channel: https://www.youtube.com/carcc

CaRCC YouTube Channel Systems-Facing Playlist:
https://www.youtube.com/playlist?list=PLV-SC0CHLTwehApeMg_fJ5VbYndETh3yT

CaRCC Code of Conduct: https://carcc.org/about/carcc-code-of-conduct/

Join the CaRCC Slack Workspace:

https://join.slack.com/t/carcc/shared_invite/zt-9lhv4cfm-wy7RpbNI7V9qLJcvtzq8hA

Zoom Meeting Room:

https://utah.zoom.us/my/carcc?pwd=TjFuR3VVM2d5eE5zWnEvWWxDTFBCUT09

Meeting ID: 824 051 8198
Password: 31415926

One tap mobile
+13462487799,,8240518198#,,#,31415926# US (Houston)
+16699006833,,8240518198#,,#,31415926# US (San Jose)

Dial by your location
    +1 346 248 7799 US (Houston)
    +1 669 900 6833 US (San Jose)

Join by Skype for Business
https://utah.zoom.us/skype/8240518198

NOTE: **We will be recording today's meeting via Zoom**.
So if there are questions anyone wants to ask but do so anonymously please feel free to use the direct message feature in the chat to send Betsy Hillery or Brian Haymore your question and we will present the question.

Sign-In (Name / Affiliation / Email): (on call # 47)
1. Betsy Hillery / Purdue University / eahillery@purdue.edu
2. Sai Pinnepalli / LSU / sai@lsu.edu
3. Zach Weidner / Purdue University / zweidner@purdue.edu
4. Liam Forbes / University of Alaska Fairbanks / loforbes@alaska.edu
5. Jim Leous / Penn State Office of the Associate CIO for Research / leous@psu.edu
6. Brian Haymore / University of Utah / brian.haymore@utah.edu
7. Mark Keever / Vanderbilt University / mark.keever@vanderbilt.edu
8. Daniel Lucio / NC State University / dalucio@ncsu.edu
9. Aric Werner / NCAR / aricw@ucar.edu
10. John Blaas / NCAR / jblaas@ucar.edu
11. Rollin Thomas / NERSC (LBL) / rcthomas@lbl.gov
12. Mattie Niznik / CIRA/CSU / mniznik@colostate.edu
13. Jason Wells / Harvard SEAS / jrwells@seas.harvard.edu
14. Brad Spitzbart / University of Oklahoma / bspitzbart@ou.edu

15. Clark Gaylord / George Washington University / cgaylord@gwu.edu
16. Nick Smith / Purdue University / smithnp@purdue.edu
17. Chris Phillips / Purdue University / cphillips@purdue.edu
18. Rich Angeletti / 3ROX-PSC / reno@psc.edu
19. Dori Sajdak / UBuffalo CCR / djm29@buffalo.edu
20. Michael Tang / University of British Columbia / michael.tang@ubc.ca
21. Anita Orendt / University of Utah / anita.orendt@utah.edu
22. Hakizumwami Birali Runesha / University of Chicago / runesha@uchicago.edu
23. Mike Robbert / Colorado School of Mines / mrobbert@mines.edu
24. Bob Freeman / Harvard Business School / rfreeman@hbs.edu
25. Andy Ingham / Duke University / andy.ingham@duke.edu
26. Lev Gorenstein / Purdue University / lev@purdue.edu
27. Ric Anderson / The University of Arizona / ric@arizona.edu
28. David G. Cyrille / Stony Brook University / David.Cyrille@stonybrook.edu
29. Kelly L. Rowland / NERSC (LBNL) / kellyrowland@lbl.gov
30. Sam Weekly / Purdue University / sweekly@purdue.edu
31. Kirk M. Anne / Rochester Institute of Technology / kirk.m.anne@rit.edu

Questions:
- Do you have classrooms that use this and do they have unique environments? How has this been solved in your environment?
    - Kernels are built in a shared space and then generally available.
- Where does the hub actually run?
    - On one of the Support nodes - Cheyenne
- How long have you run this service?
    - 3 total people that support this in some fashion
    - Should require less than a full FTE to support

Notes:
- JupyterHUB@NCAR
- Jared Baker
- NCAR is a NSF funded project and discussion is personal opinion not a funding agency opinion
- A broad set of universities support NCAR.
- NCAR mission is to support Earth System Science
- More than 1500 users at over 500 universities and research institution is supported by NCAR
- Resources at NCAR
    - Cheyenne / PBS professional is the scheduling system
    - Casper - everything that doesn't fit into a super computing category.
        - High memory nodes, GPU (Data Analysis and vizulation cluster)

- - - Most of the JupyterHub users use this
      - ○ GLADE - Storage
      - ○ Campaign Store - Storage (Med to long term storage)
      - ○ Quasar - Storage for long term, curriated data (Data Archive)
      - ○ GLADE-U - Container Images, $HOME
      - ○ Stratus - Object Store, proof of concept to test various things, used for long term storage as well.
      - ○ DERECHO - New cluster coming up. It will have a heavy GPU system, moving away from infiniband and moving to Slingshot / Lustre based scratch will be used with this resource.
  - ● Why use JupyterHub -
    - ○ Old and complicated software that uses Python is a perfect fit as most of those applications are great.
    - ○ Community really likes a notebook approach to computing
    - ○ Low barrier to access for end users
    - ○ Maintains state between devices and sessions
    - ○ Extensible to the user community
  - ● Servers are all started via spawners.
  - ● Web interface and customizable
  - ● Proxy implementation -
    - ○ configurable - HTTP-Proxy
      - ■ NCAR leverages an NGINX installation to provide SSL
    - ○ Configurable - Traefik
  - ● The database is a state keeper.
    - ○ NCAR uses sqlite as the database on top of IBM Spectrun
    - ○ They have used PostgrSQL and may migrate there in the future
  - ● Login page for access is broken down into Production and Development
  - ● Metrics -
    - ○ Performance
    - ○ Management usage reports
    - ○ Software Consultnats (what kernel, etc are being used)

Next Month's Presentation:
- ● **July 2022: We will not hold a zoom meeting in July, but instead look to find an opportunity to do a social meetup at PEARC22 for those attending. More info to come.**

Suggested future topic for a systems facing call (Feel free to add one):

- ●