Towards Reproducible Research on CyberGISX with Lmod and Easybuild

Alexander Michels, Anand Padmanabhan, Zhiyu Li, and Shaowen Wang*

CyberGIS Center for Advanced Digital and Spatial Studies
University of Illinois at Urbana-Champaign
Urbana, IL, USA
{michels9, apadmana, zhiyul, shaowen}@illinois.edu
*Corresponding Author

Gateways 2021
October 19-21, 2021

CyberGISX

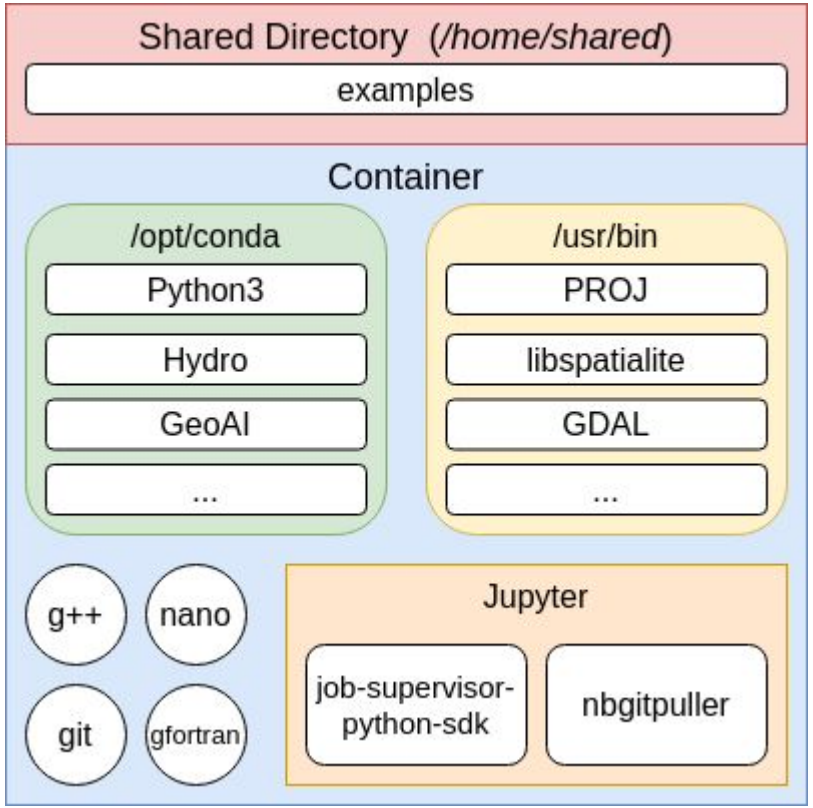
CyberGISX is a Jupyter-based science gateway targeted at the GeoSciences.

https://cybergisxhub.cigi.illinois.edu/

Our Previous Solution

Everything in one container

7.66GB compressed
22.4GB uncompressed



Updating our Compute Environment

Any kind of update or upgrade to our compute environment:

- Led to a rebuild/reinstall of some proportion of the software
- Old versions of software are no longer available in our compute environment.

Binder

Binder could alleviate some of the problem, but:

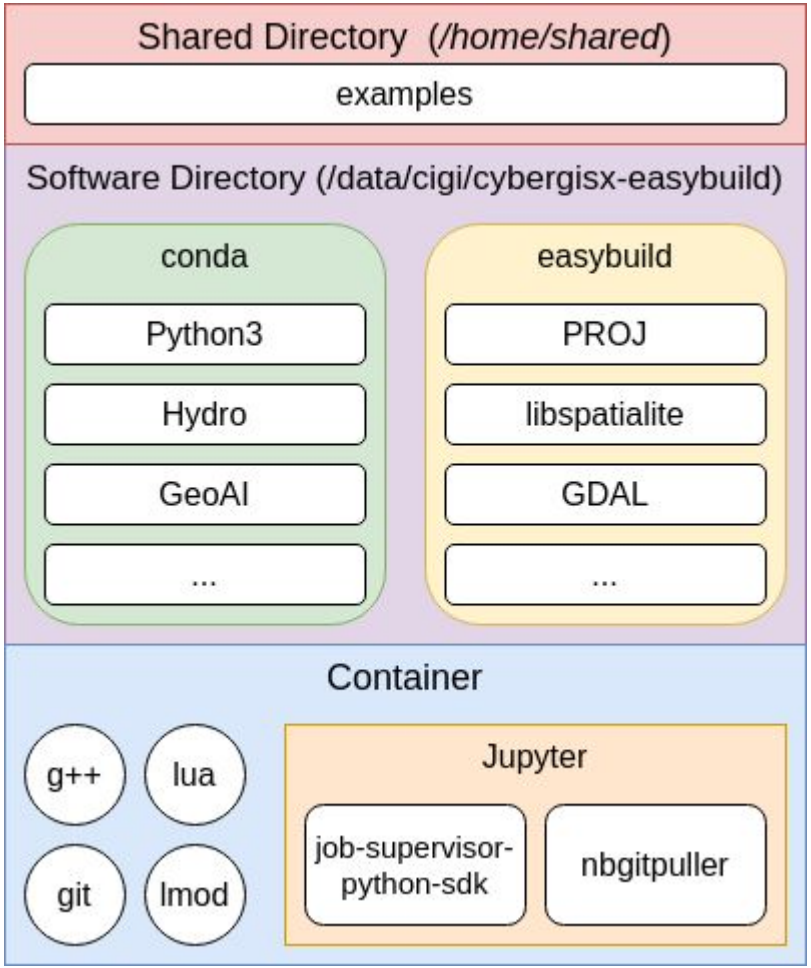
a. software is built at the time it is needed
b. the onus of specifying and managing software installations is passed to notebook developers

https://mybinder.org/

Our New Approach

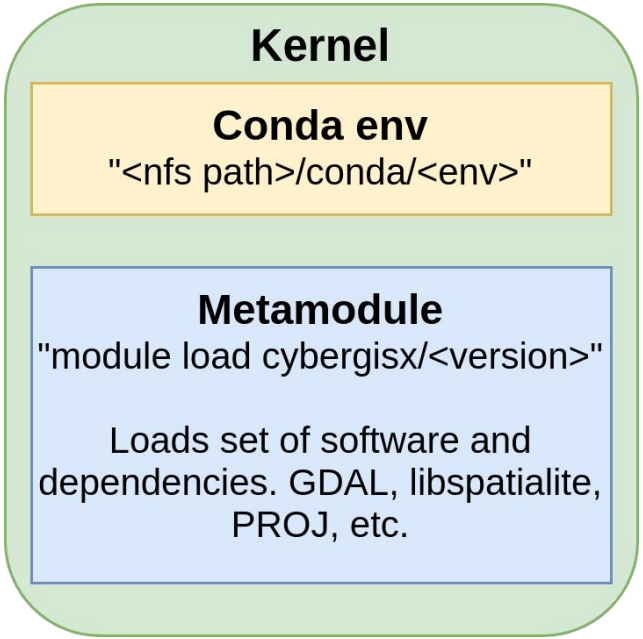
Compute environment software is moved outside of the container. To do this, we utilize:

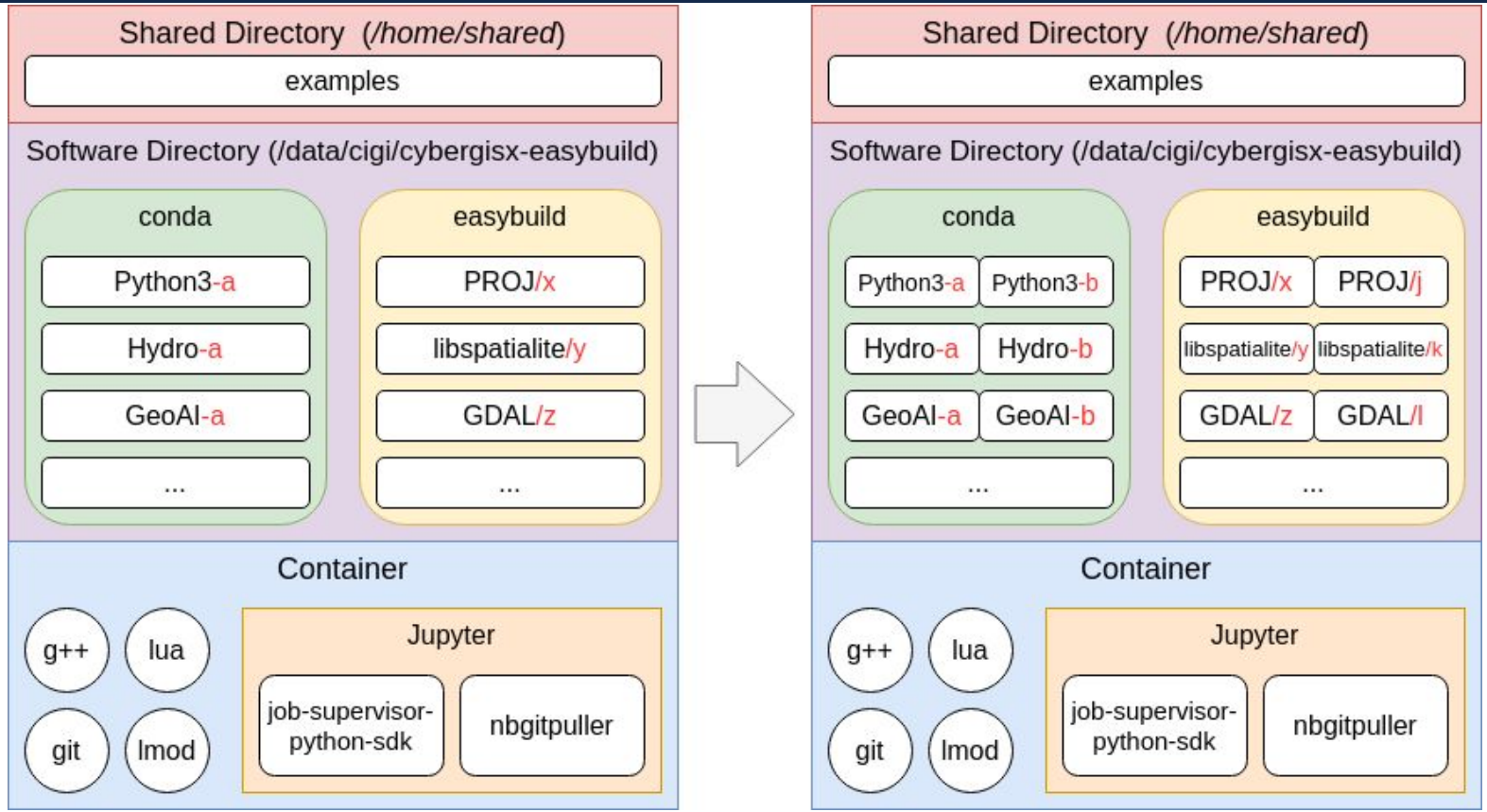
- **Easybuild** - to build software and provide use with modulefiles
- **Lmod** - for users to manage software
- **Conda** - for Python environments



Kernels in the New Approach

The specifications for kernels are given by a JSON file and a script which runs before the kernel is started.





Updating in Our New Approach

User Experience

We have had this approach deployed for over a month on CyberGISX and its sister platform CyberGIS-Jupyter for Water (CJW), with minimal issues.

The average user doesn't know anything changed.

Accessing/managing "old" versions adds a bit of complexity. CyberGISX CLI designed to alleviate this, but more work needs to be done.

Future Work

- User Experience
- Scalability concerns
- Availability outside of the gateway

References

1. T. Kluyver, B. Ragan-Kelley, F. Perez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing, “Jupyter notebooks – a publishing format for reproducible computational workflows,” in Positioning and Power in Academic Publishing: Players, Agents and Agendas, F. Loizides and B. Schmidt, Eds. IOS Press, 2016, pp. 87 – 90.

2. Calyam, P., Wilkins-Diehr, N., Miller, M., Brookes, E. H., Arora, R., Chourasia, A., Jennewein, D. M., Nandigam, V., Drew LaMar, M., Cleveland, S. B., Newman, G., Wang, S., Zaslavsky, I., Cianfrocco, M. A., Ellett, K., Tarboton, D. G., Jeffery, K. G., Zhao, Z., González-Aranda, J., Perri, M. J., Tucker, G., Candela, L., Kiss, T., and Gesing, S. (2020) “Measuring Success for a Future Vision: Defining Impact in Science Gateways/Virtual Research Environments”. Concurrency and Computation: Practice and Experience, https://doi.org/10.1002/cpe.6099

3. D. Yin, Y. Liu, A. Padmanabhan, J. Terstriep, J. Rush, and S. Wang. (2017). A CyberGIS-Jupyter Framework for Geospatial Analytics at Scale. In Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact (p. 18). ACM.

4. Jupyter et al., "Binder 2.0 - Reproducible, Interactive, Sharable Environments for Science at Scale." Proceedings of the 17th Python in Science Conference. 2018. doi://10.25080/Majora-4af1f417-011

5. K. Hoste, J. Timmerman, A. Georges and S. De Weirdt, "EasyBuild: Building Software with Ease," 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, 2012, pp. 572-582, doi: 10.1109/SC.Companion.2012.81.

6. R. McLay, K. W. Schulz, W. L. Barth, and T. Minyard, "Best practices for the deployment and management of production HPC clusters," In State of the Practice Reports, SC11, pages 9:1–9:11, Nov. 2011. Doi.acm.org/10.1145/2063348.2063360.

7. Anaconda Software Distribution. Computer software. Vers. 2-2.4.0. Anaconda, Nov. 2016. Web. <https://anaconda.com>.

8. Zhiyu Li, Alexander Michels, Fangzheng Lu, Anand Padmanabhan, and Shaowen Wang (2021). CyberGIS-Jupyter for Water, HydroShare, http://www.hydroshare.org/resource/4cfd280e8eb747169b293aec2862d4f5

Questions?

Correspondence to:
Shaowen Wang
shaowen@illinois.edu