Unlimited Scientific Libraries and Applications in Kubernetes, Instantly!

SV Data Science, Machine Learning, MLOps & Kubeflow Meetup

2022-05-05

Guillaume Moutier

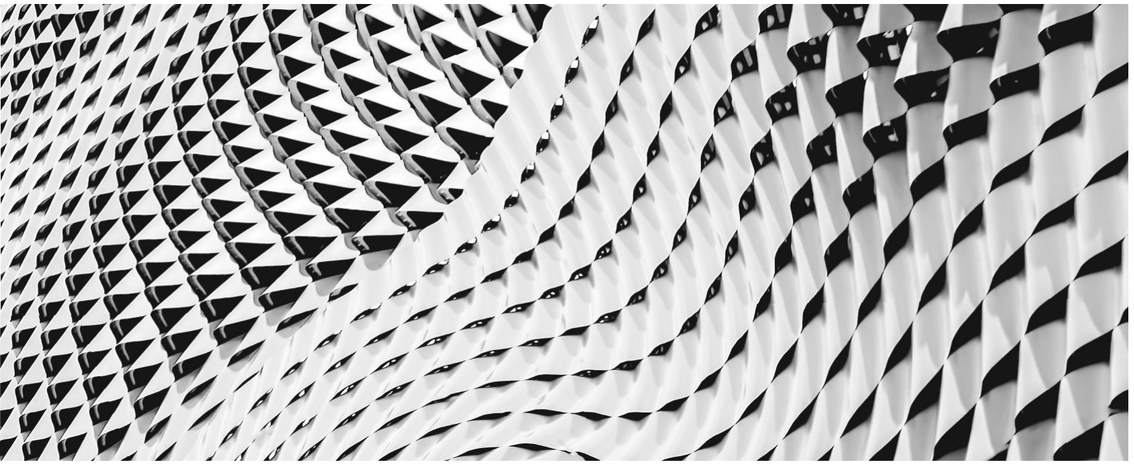
Sr Principal Data Engineering Architect

Red Hat

What we'll discuss today

- Introduction and Context
- The problem
- The solution
- Demo
- What's next?

Introduction and Context



"Don't adventures ever have an end? I suppose not. Someone else always has to carry on the story."
Bilbo Baggins

Red Hat

Who am I?

Former head of Architecture and CTO at

Laval University in Québec City

Now a Data Engineering Architect with the

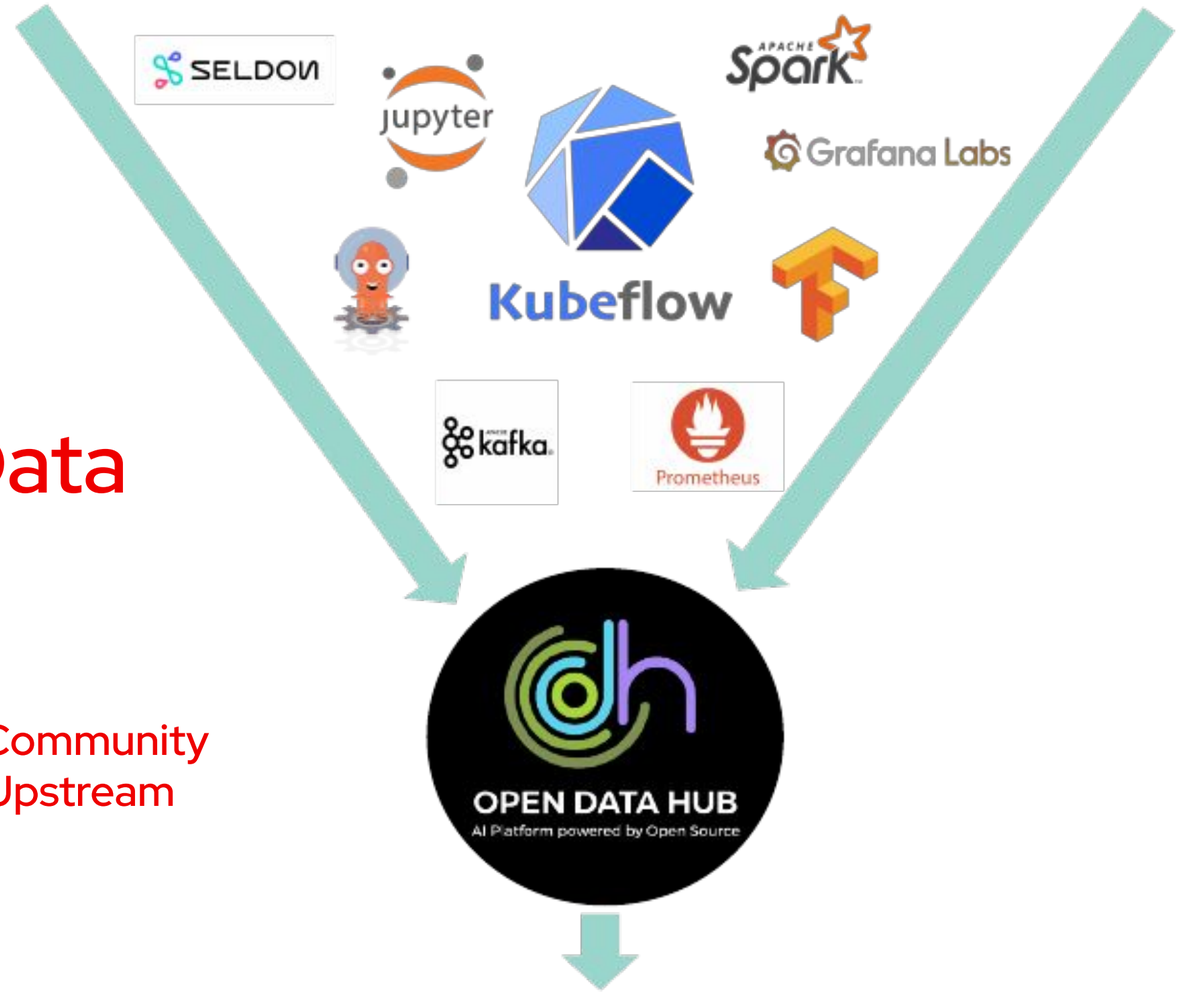
Red Hat OpenShift Data Science team

Red Hat
OpenShift
Data Science

Unlimited Scientific Libraries and Applications in Kubernetes, Instantly!

Open Data Hub

Open Source Community
Downstream/Upstream



Red Hat Application Services

Red Hat OpenShift

Red Hat OpenShift Data Foundation

Goals

- Provide an end-to-end AI/ML platform on OpenShift
- One stop easy operator deployment for the platform on OCP
- Provide Tools for each stage in the AI/ML platform and for all AI/ML user personas optimized for OpenShift
- Provide monitoring tools for model and services used by DevOps
- Provide development tools for Data Scientists
- Provide ETL tools used by Data Engineers
- AI/ML pipelines and long processing tasks.

Red Hat

Red Hat OpenShift Data Science

Addressing AI/ML experimentation and integration use cases on a managed platform

Cloud Service

Available on Red Hat OpenShift Dedicated (AWS) and Red Hat OpenShift Service on AWS

Increased capabilities/collaboration

Combines Red Hat components, open source software, and ISV certified software available on Red Hat Marketplace

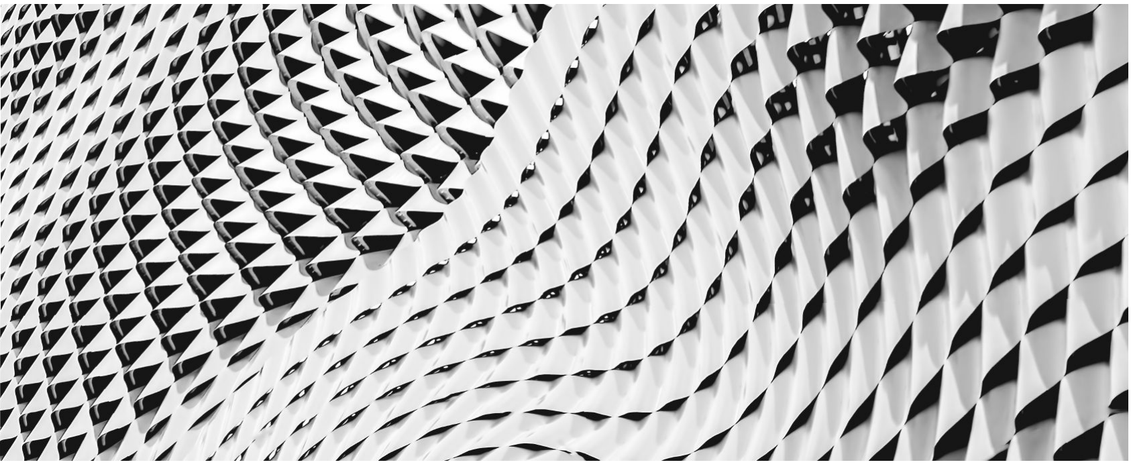
Core data science workflow

Provides data scientists and intelligent application developers the ability to build, train, and deploy ML models

Rapid experimentation use cases

Model outputs are hosted on the Red Hat OpenShift managed service or exported for integration into an intelligent application

The problem



"I have a bad feeling about this..."

Your favorite Star Wars character

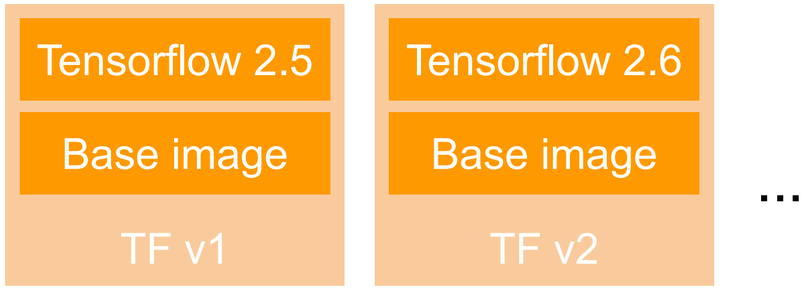
Red Hat

The on-demand notebook example

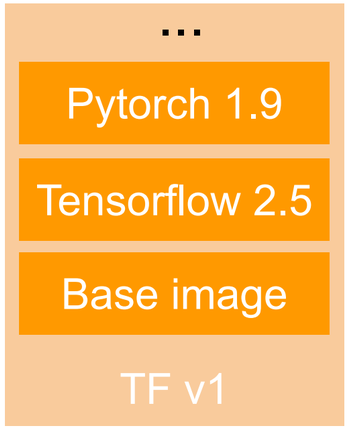
The “base” container image includes:

- Python at a specific version
- Some useful libraries and applications
- Jupyter with pre-defined extensions

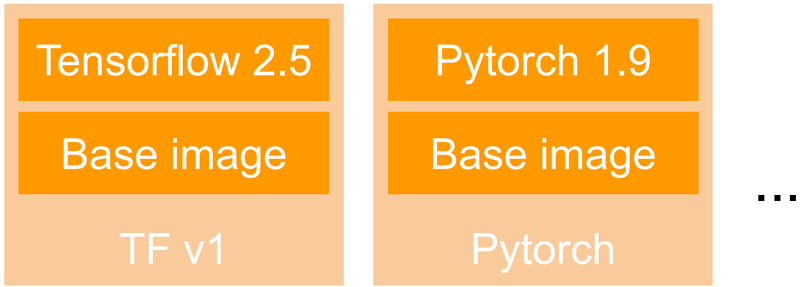
Users want more, what to do?



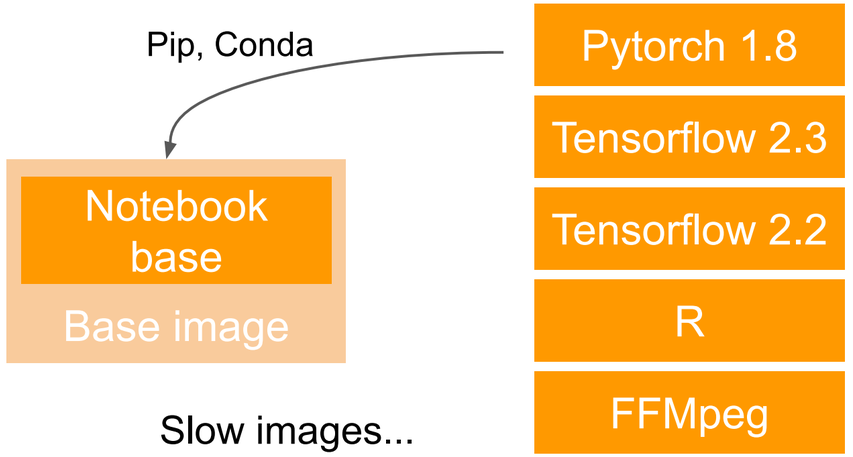
Multiple versions => Multiple images



Overweight images...



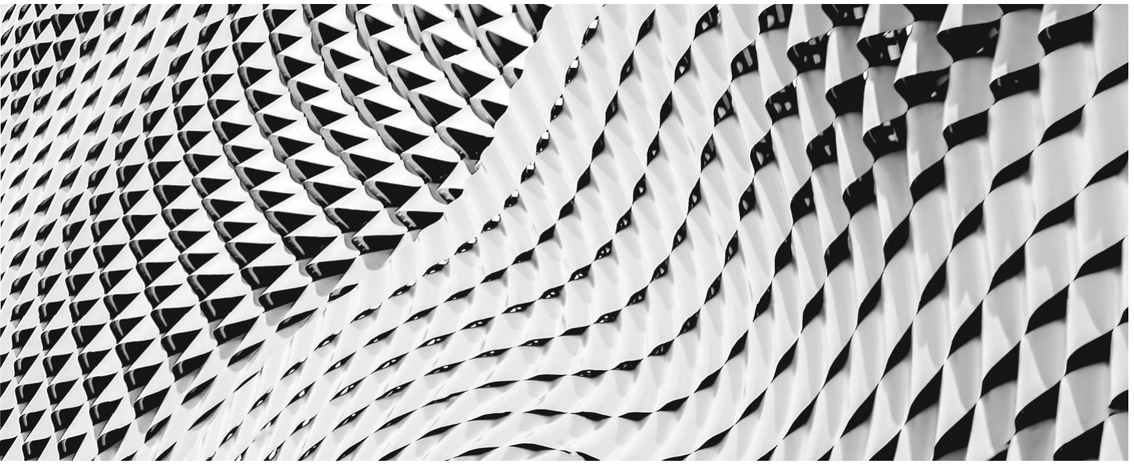
Multiple tools => Multiple images



Slow images...

They want more, what to do?

The solution

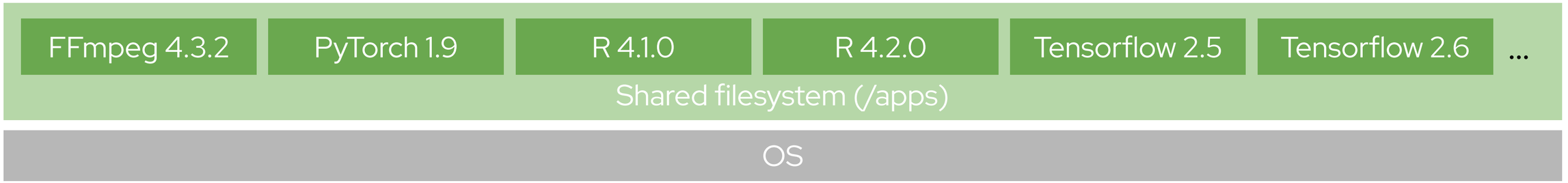


"You Did Good Son.
I'm Proud Of You."
Admiral Anderson

Red Hat

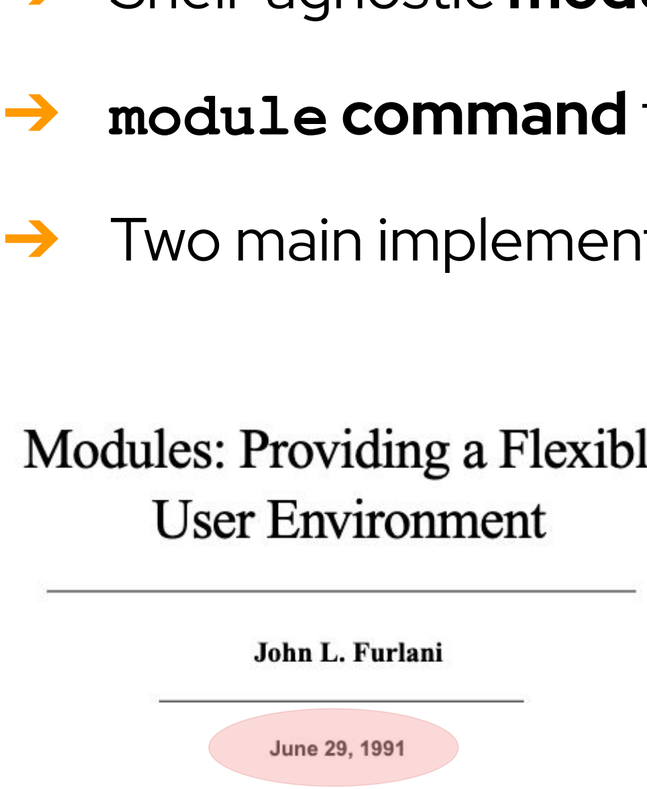
HPC systems usually provide a **central software stack**

- Software installed in a **non-standard location** on a shared filesystem (like `/apps`)
- Collection of **100's-1000's of installations** (diff. software, versions, compilers)
- **Separate directory** for each software installation (different applications/versions/...)
- Provided software is **optimized for system architecture**
- Software is **built from source** where possible (to ensure good performance)
- **Additional installations are added** on-demand, or as new versions are released



Easy access to central software stack using **environment modules**

- Traditional way to **let users** of HPC systems **manage their environment**
- Shell-agnostic **module files** specify what to change in shell environment
- **`module` command** to check for available modules, (un)load modules, …
- Two main implementations:


Modules: Providing a Flexible User Environment

John L. Furlani

June 29, 1991

http://modules.sourceforge.net/docs/Modules-Paper.pdf

ENVIRONMENT MODULES

- Evolution of original implementation in Tcl
- Module files written in Tcl
- Actively developed & maintained
- Less popular, but default in RHEL-based OSs
- http://modules.sourceforge.net

Lmod

- Modern implementation in Lua
- Module files written in Lua or Tcl
- Actively developed & maintained
- Developed for hierarchical modules
- Most popular (>85% of systems)
- https://lmod.readthedocs.io

ssh example@supercomputer

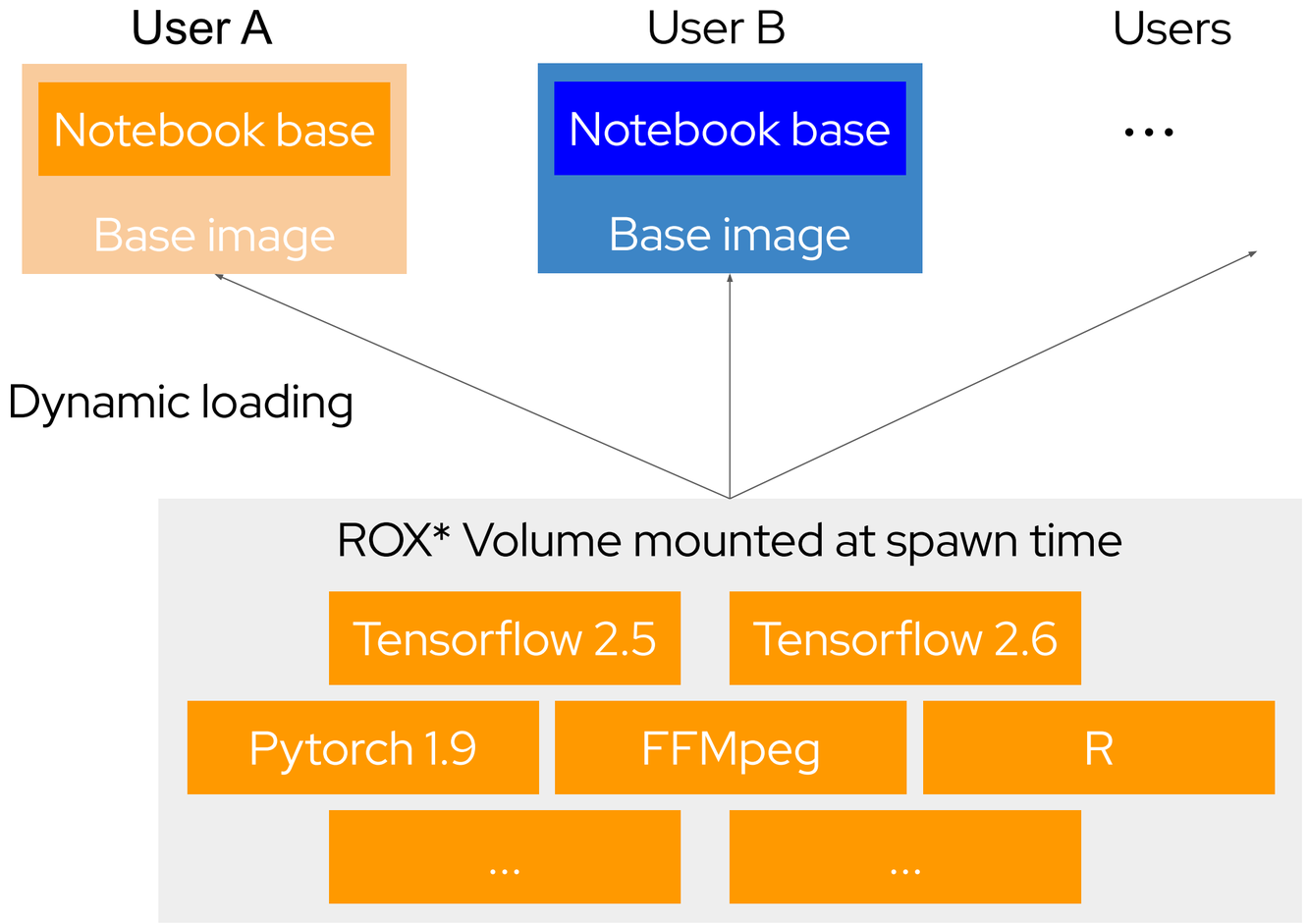
```
$ example --version
-bash: example: command not found

$ module use /apps/modules

$ module avail example
----- /apps/modules -----
example/1.2.3

$ module load example/1.2.3

$ example --version
1.2.3
```

Contents of `/apps/modules/example/1.2.3.lua`:

```
local root = "/apps/software/example/1.2.3"
prepend_path("PATH", pathJoin(root, "bin"))
prepend_path("LD_LIBRARY_PATH", pathJoin(root, "lib"))
prepend_path("LIBRARY_PATH", pathJoin(root, lib"))
prepend_path("PYTHONPATH", pathJoin(root, "lib/python3.9/site-packages"))
setenv("EXAMPLE_ROOT", root)
setenv("EXAMPLE_VERSION", "1.2.3")
setenv("EXAMPLE_ENABLE_DEBUG_OUTPUT", "1")
```

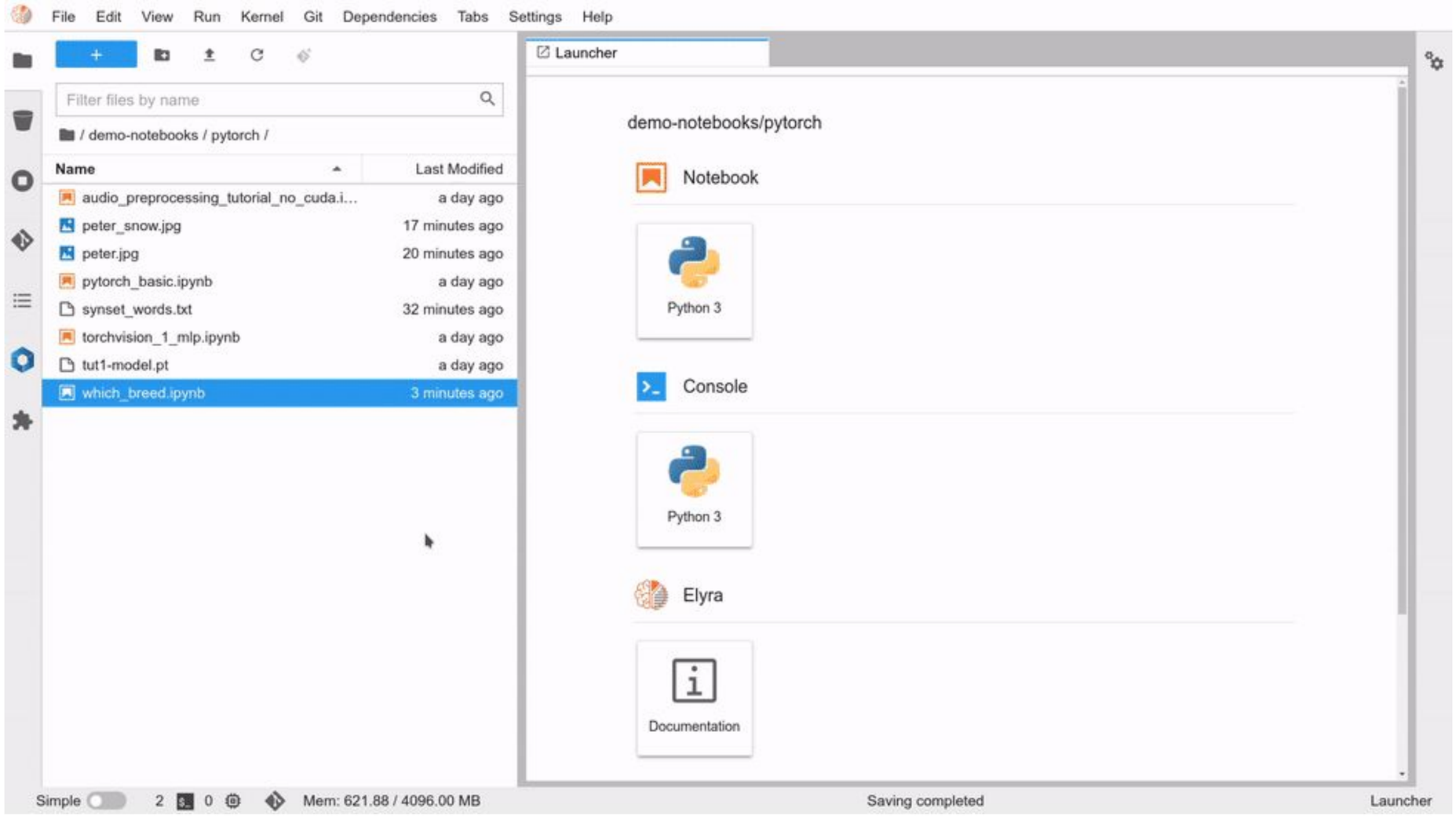
So let's use environment modules!

ODH Highlander



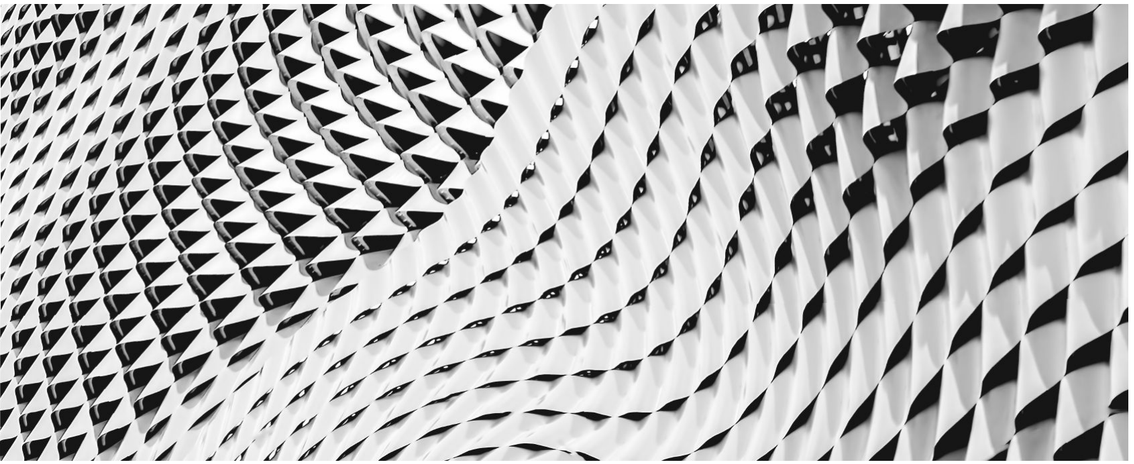
**ROX: Read-Only Many, a volume that can be mounted simultaneously into many containers*

And a miraculous Jupyter Extension!



Red Hat

Demo

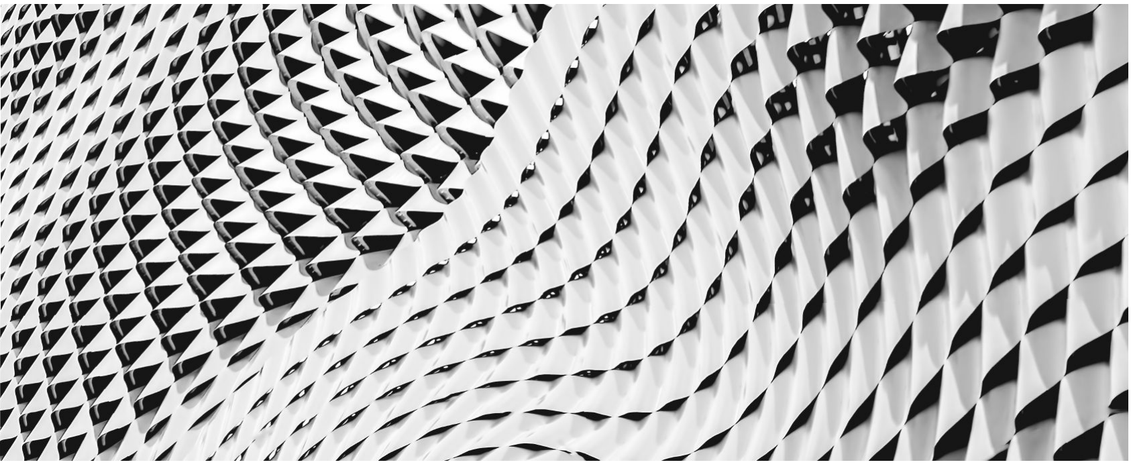


"Beam me up Scotty."

Captain James T. Kirk

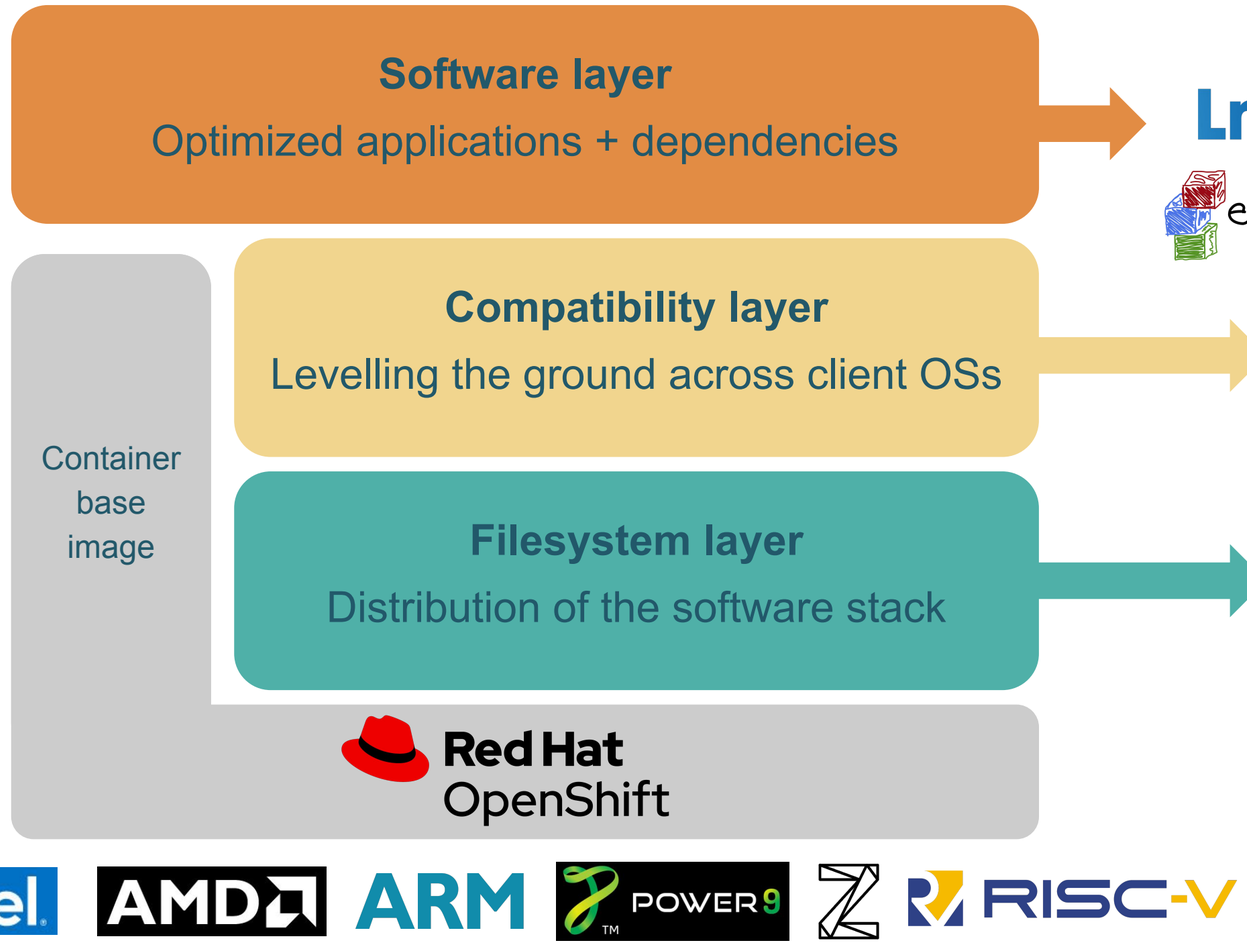
Red Hat

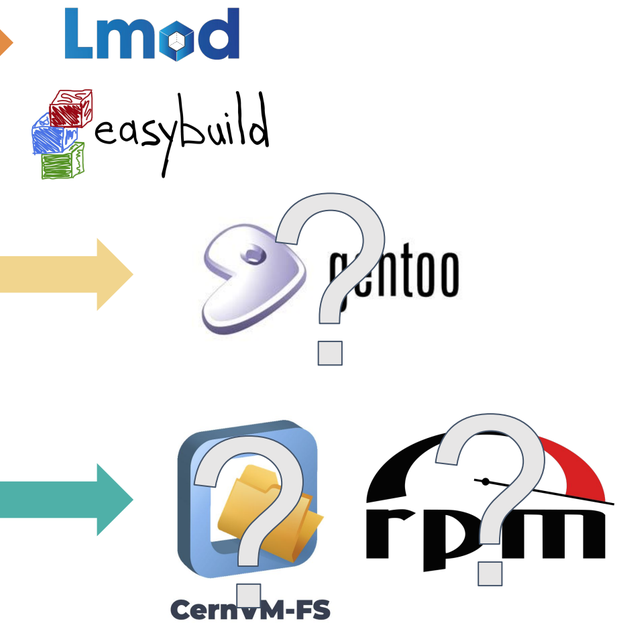
What’s Next?



“We're going to need a bigger boat.”
Chief Brody

Red Hat





ODH-Highlander
Repository of curated pre-compiled modules

References

- Open Data Hub: opendatahub.io
- OpenShift Data Science: red.ht/datascience
- ODH-Highlander: https://github.com/guimou/odh-highlander
  (soon at odh-highlander.github.io)
- Environment modules: modules.sourceforge.net
- Lmod environment modules tool: lmod.readthedocs.io
- EasyBuild: easybuild.io - dx.doi.org/10.1109/HUST.2014.8
- EESSI project: eessi-hpc.org - eessi.github.io/docs - dx.doi.org/10.1002/spe.3075

Thank you

Red Hat is the world's leading provider of enterprise open source software solutions. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500.

linkedin.com/company/red-hat

youtube.com/user/RedHatVideos

facebook.com/redhatinc

twitter.com/RedHat

Red Hat