Scientific cloud computing

(a brief overview)

Matias Carrasco Kind (NCSA)

June 20th, 2018 - Data Visualization and Exploration in the LSST Era @ NCSA

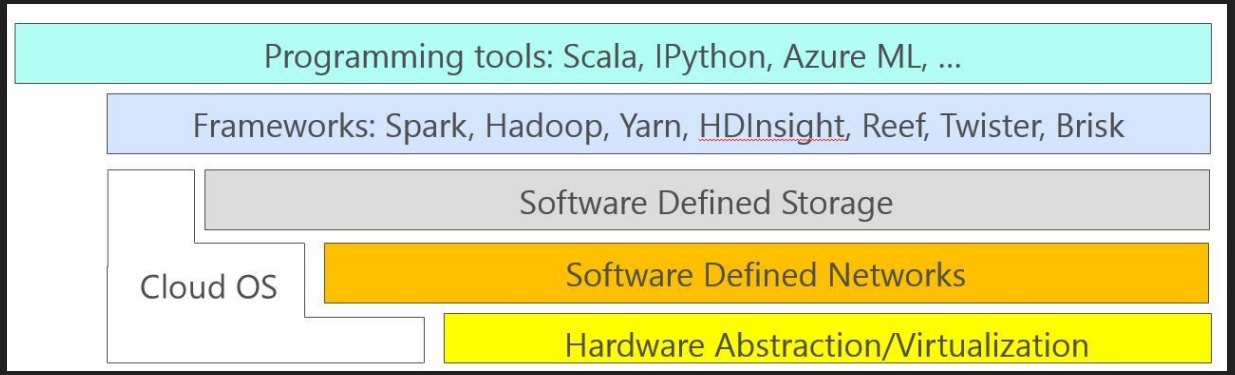
Outline

- This is not a new discussion
- Is not a in-depth introduction
- Machine Learning i the Cloud is a whole beast
- Engage discussion and share current vision
- Tutorial

Why we should be doing science on the cloud

- Remote dynamic data (!= Big data)
- Big data ⇒ Data Gravity
- Remote software/server
- Easy to deploy*
- Asynchronous
- Web applications / Shareable
- Serverless applications
- …. more

*Arguable



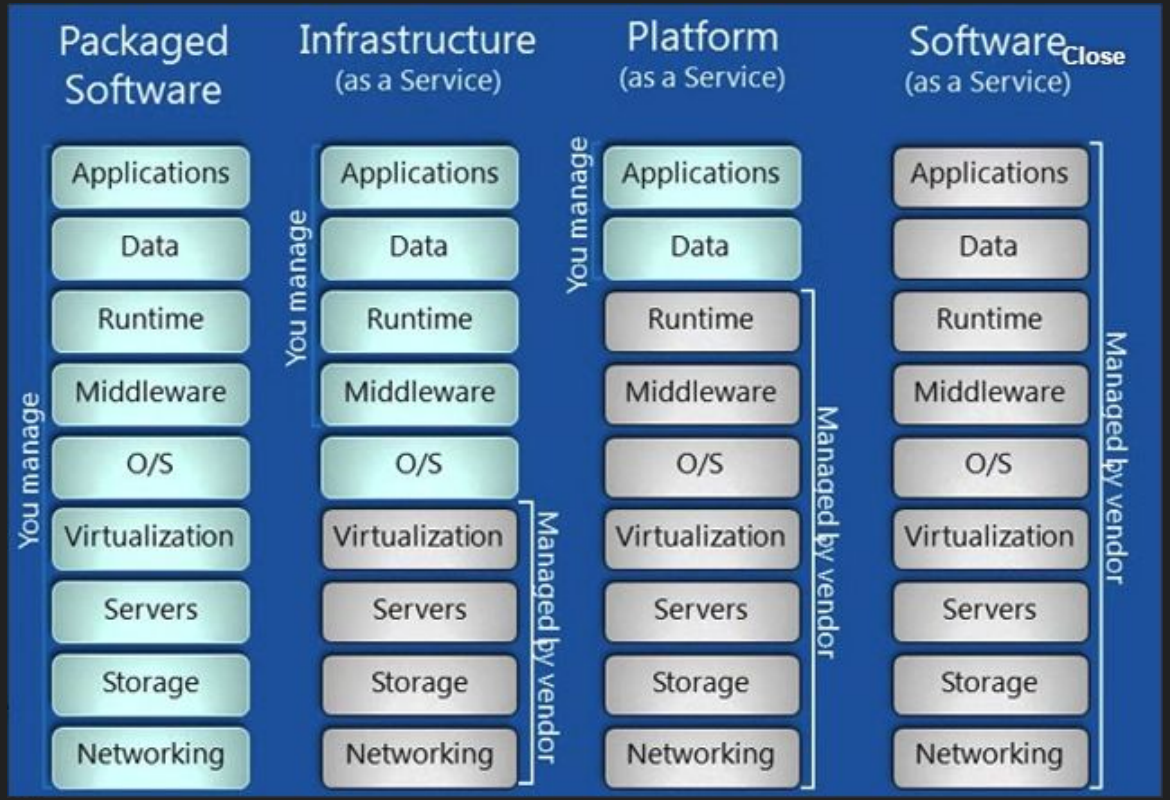
Why we shouldn’t be doing science on the cloud

- Because there is no a real reason for it
- HPC is not there yet, large latencies
  ... but HPC is adopting cloud technologies
- Full control on data and application
- Security concerns
- Faster development*
- Billing (if a commercial provider)
- more ...

*Arguable

What kind of science?

- HTC vs HPC
- Interactive
- Small projects
- Visualizations
- Short term projects*



*Arguable

Which Clouds?

Amazon Web Services (AWS) – 40%

Microsoft Azure – about 50% of AWS

Google Cloud – 3rd place

IBM Bluemix – growing fast

Salesforce, DigitalOcean, Rackspace, 1&1, UpCloud, CityCloud, CloudSigma, CloudWatt, Aruba, CloudFerro, Orange, OVH, T-Systems

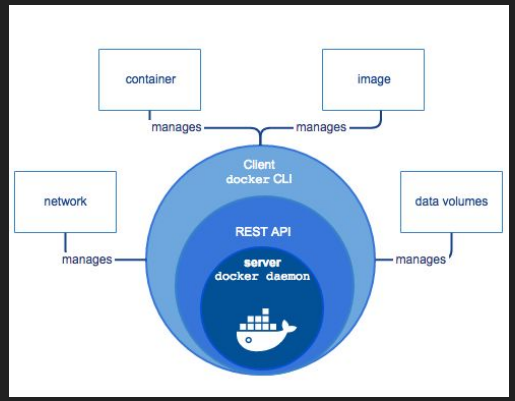
Cloud for Research: Aristotle, Bionimbus, Jetstream, Chameleon, RedCloud

Chameleon

Jetstream

Containerization to the rescue*

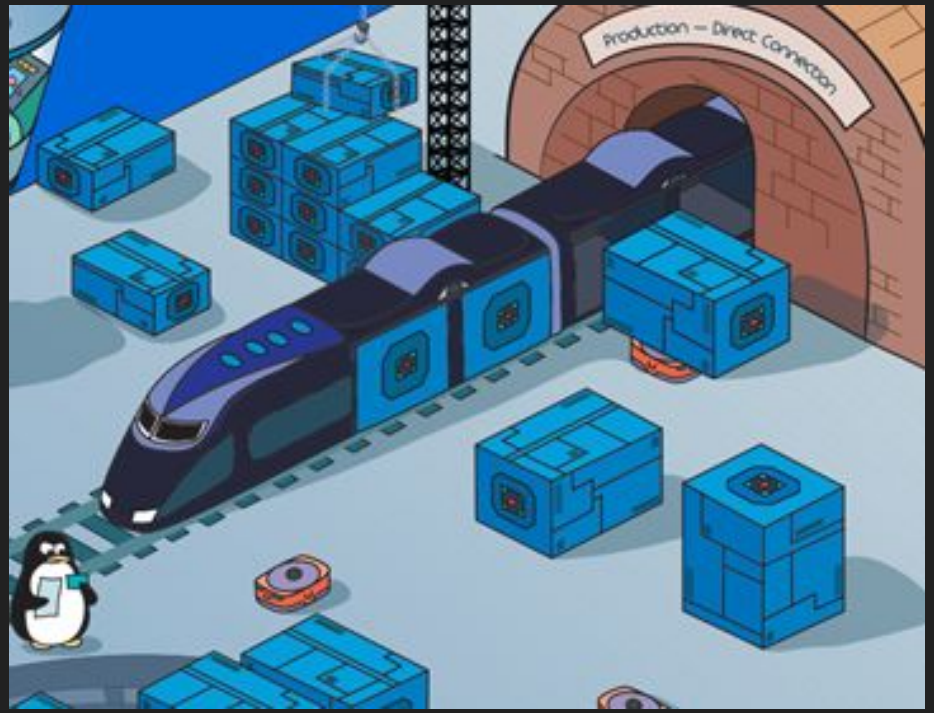
- It's been around for over 10 years, but popular since 2014 thanks to Docker
- Many other alternatives (rkt, kata, shifter, singularity, etc...)
- Lightweight, stand-alone, executable package of a piece of software that includes everything to run it
- Not just applications
- Software designed storage
- Software designed network



*Arguable

Container organization and orchestration

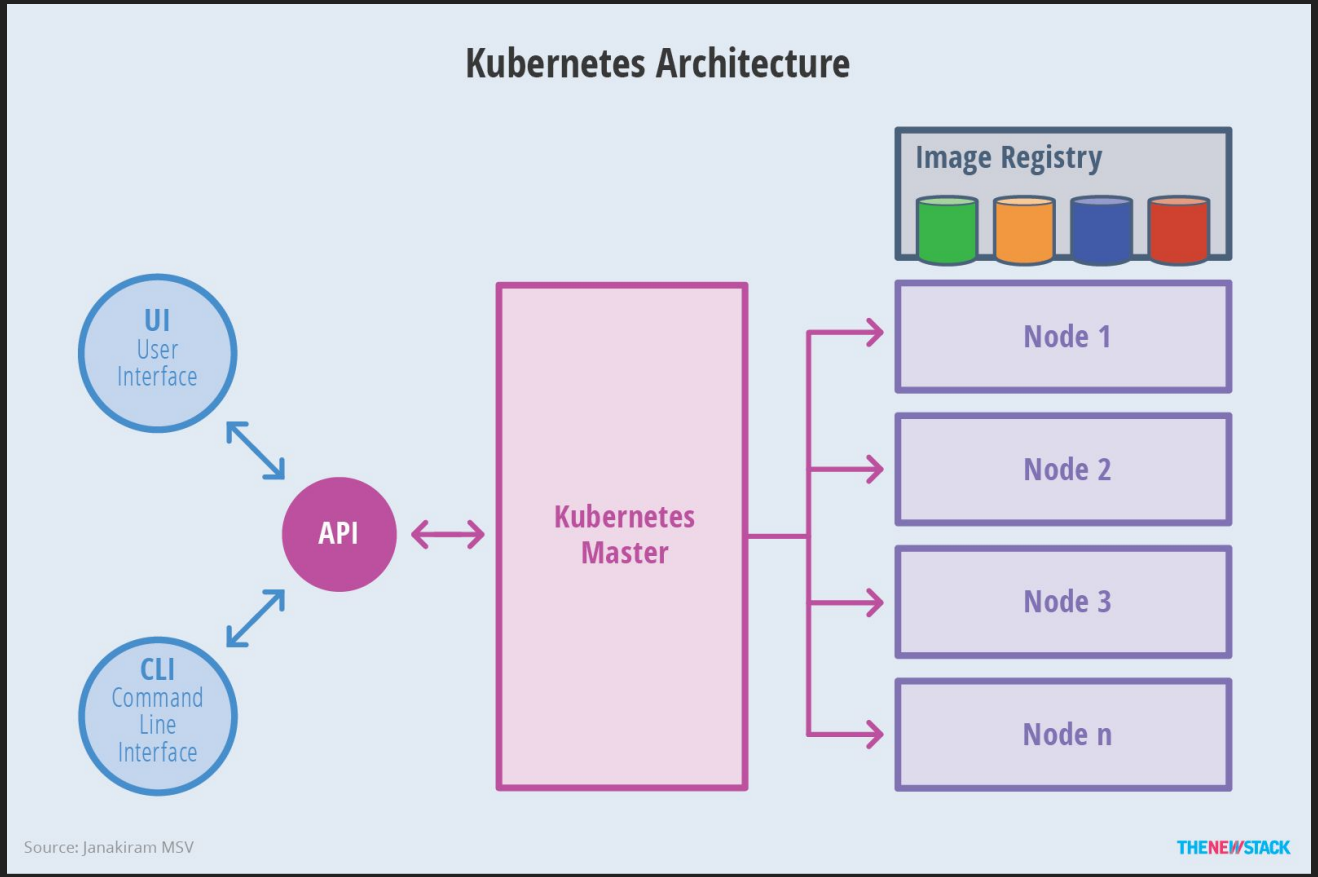
- We can create a container with an application inside, now what?
- Need to consider:
  - Resource needs
  - Fault tolerant
  - Load balancing
  - Storage management
  - Lifecycle
  - Service Discovery
  - Scalability



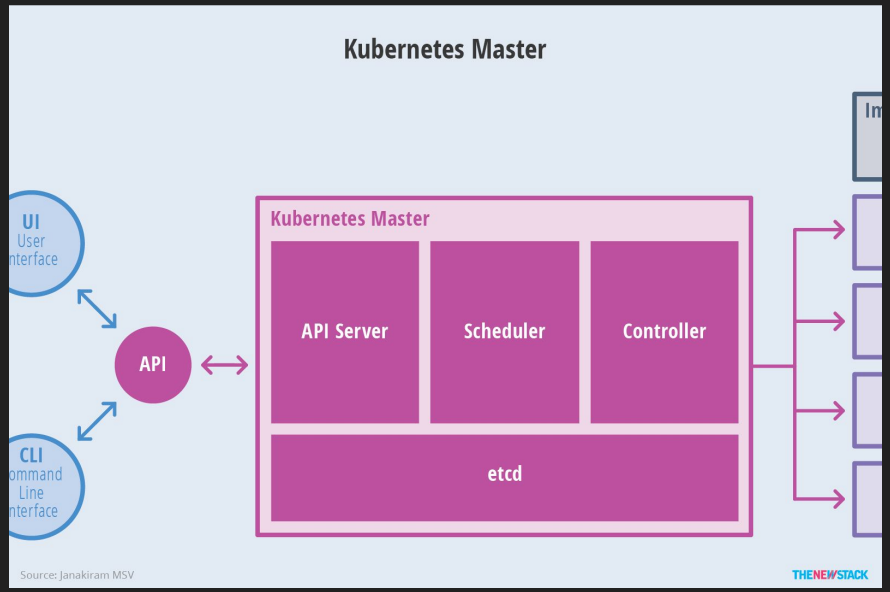
The Kubernetes Factor

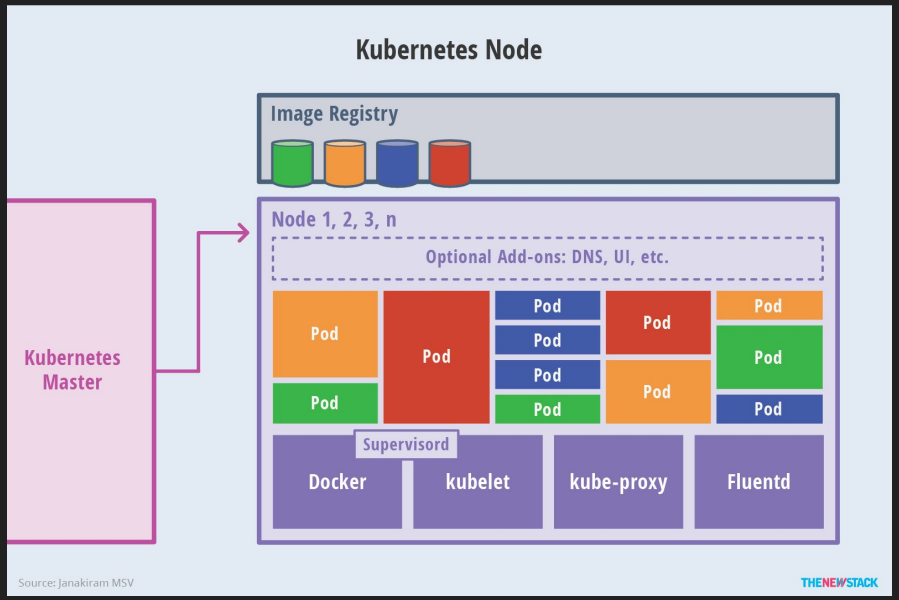
- It solves all previous issues and more (not the only one but most popular)
- Open source container management and orchestration platform
- Developed by Google, made open sourced
- One of top 5 most commented open source repositories and #2 in number of pull request
- Standard within all cloud platforms
- Flexible and extensible, customize schedulers
- Is changing the cloud computing paradigm

The Kubernetes Factor



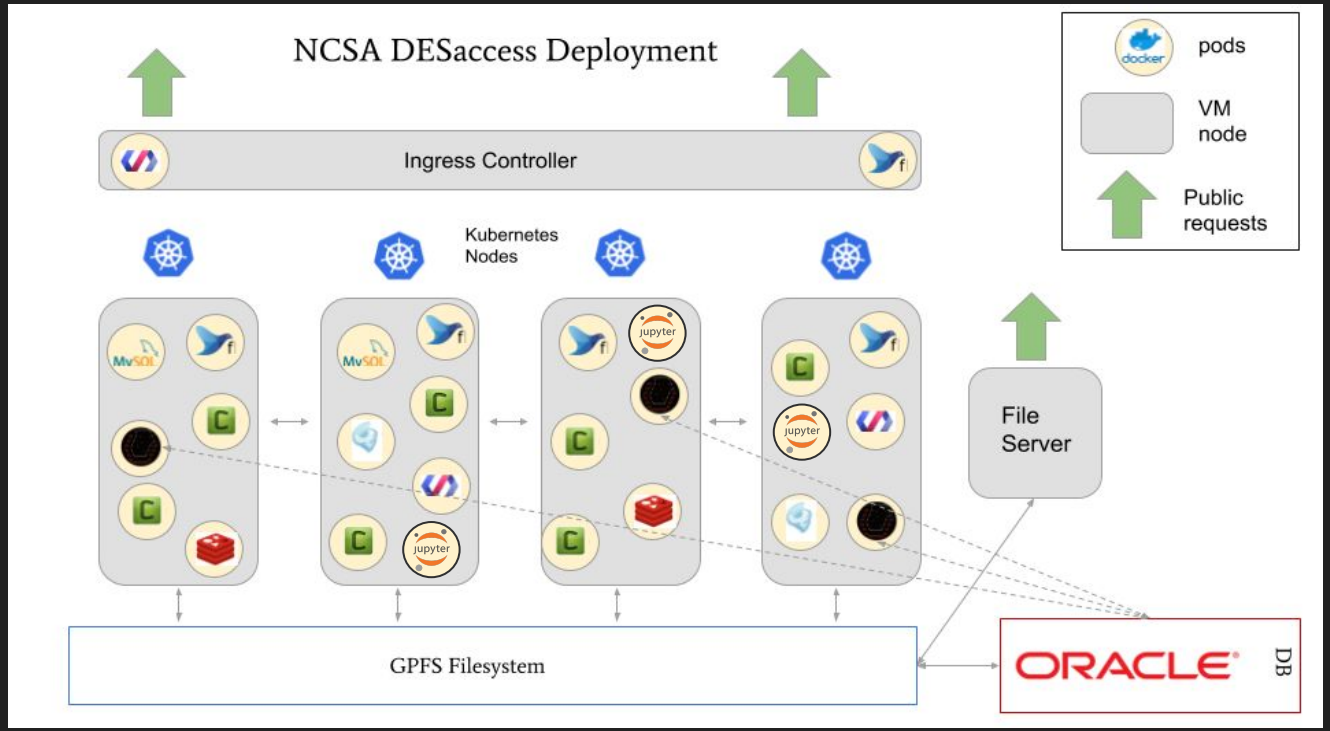
The Kubernetes Factor





Astronomical examples

- It's been adopted across data centers
- NCSA
- NASA
- CERN
- LSST



Final Remarks

- Doing science on the cloud is happening in many scientific fields including Astronomy
- HPC is adopting container technologies to leverage the benefits of both worlds
- Kubernetes provide means to have ‘the cloud’ outside the commercial world
- This is changing the way we do astronomy

Thanks!

Matias Carrasco Kind
mcarras2@illinois.edu
matias-ck.com

Pre-reqs for tutorial

Install Docker

Create DockerHub account

Talk:

go.ncsa.illinois.edu/mck

To get the repo:

https://github.com/mgckind/container_demo