
Deploying Jupyter Notebooks at scale on XSEDE resources for Science Gateways and workshops

A. Zonca, R. Sinkovits SDSC

Pre-print: <https://arxiv.org/abs/1805.04781>

Why?

- **Jupyter Notebooks for workshops:**
 - > 50 students
 - preconfigured software + data
 - just need their browsers
- **Interactive computing for Science Gateways**
 - companion to standard Science Gateway
 - in-situ post-processing and visualization
 - no need to download results

What?

- **JupyterHub (Jupyter Notebooks for multiple users)**
- **Lots of RAM and CPU to users -> distributed**
- **Users run inside pre-packaged Docker/Singularity container**

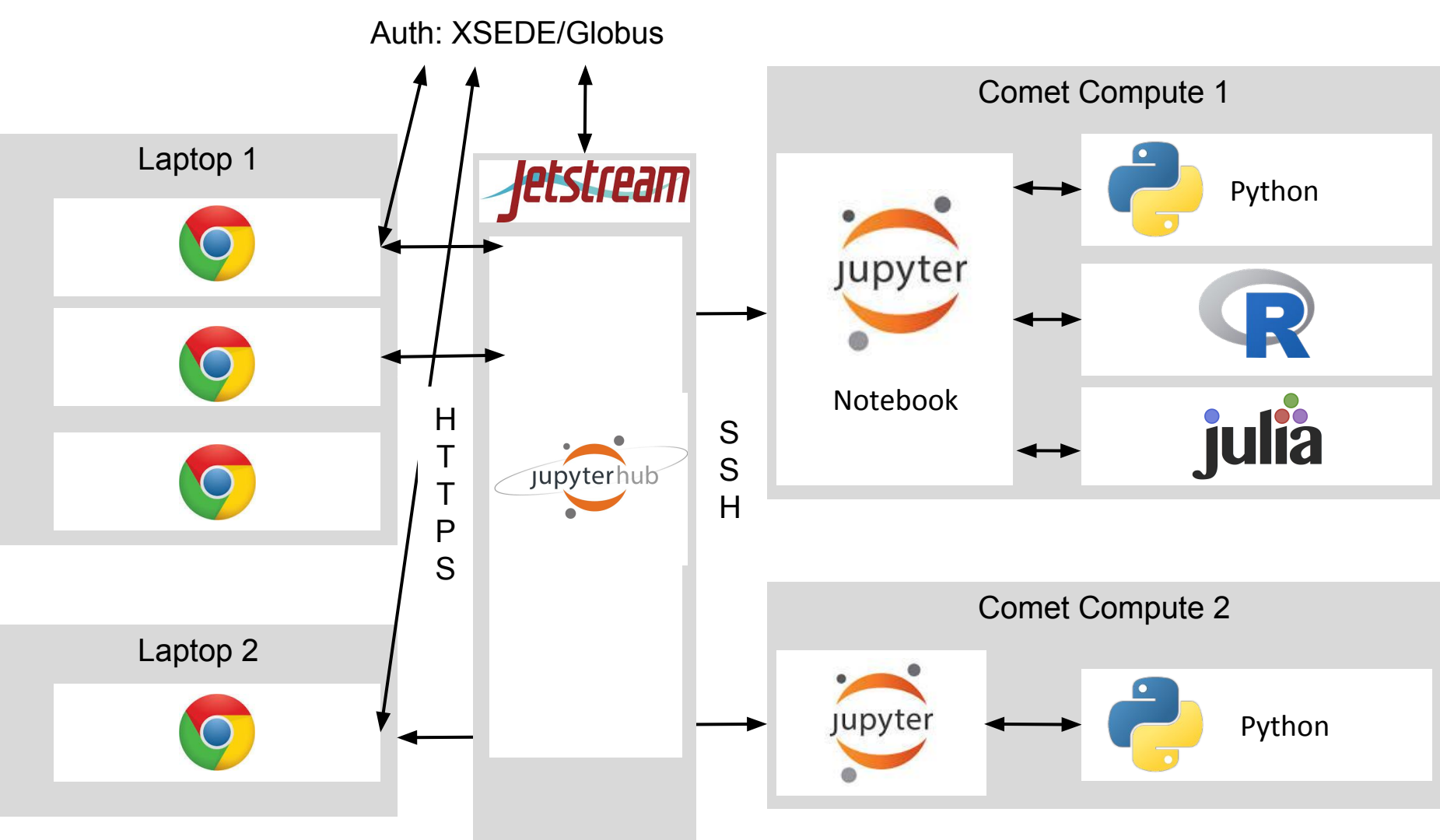
How?

Paper provides 3 strategies:

- **Launch Notebooks on Comet computing nodes**
- **Launch Notebooks on multiple Jetstream instances:**
 - with Docker Swarm
 - with Kubernetes

JupyterHub on Comet

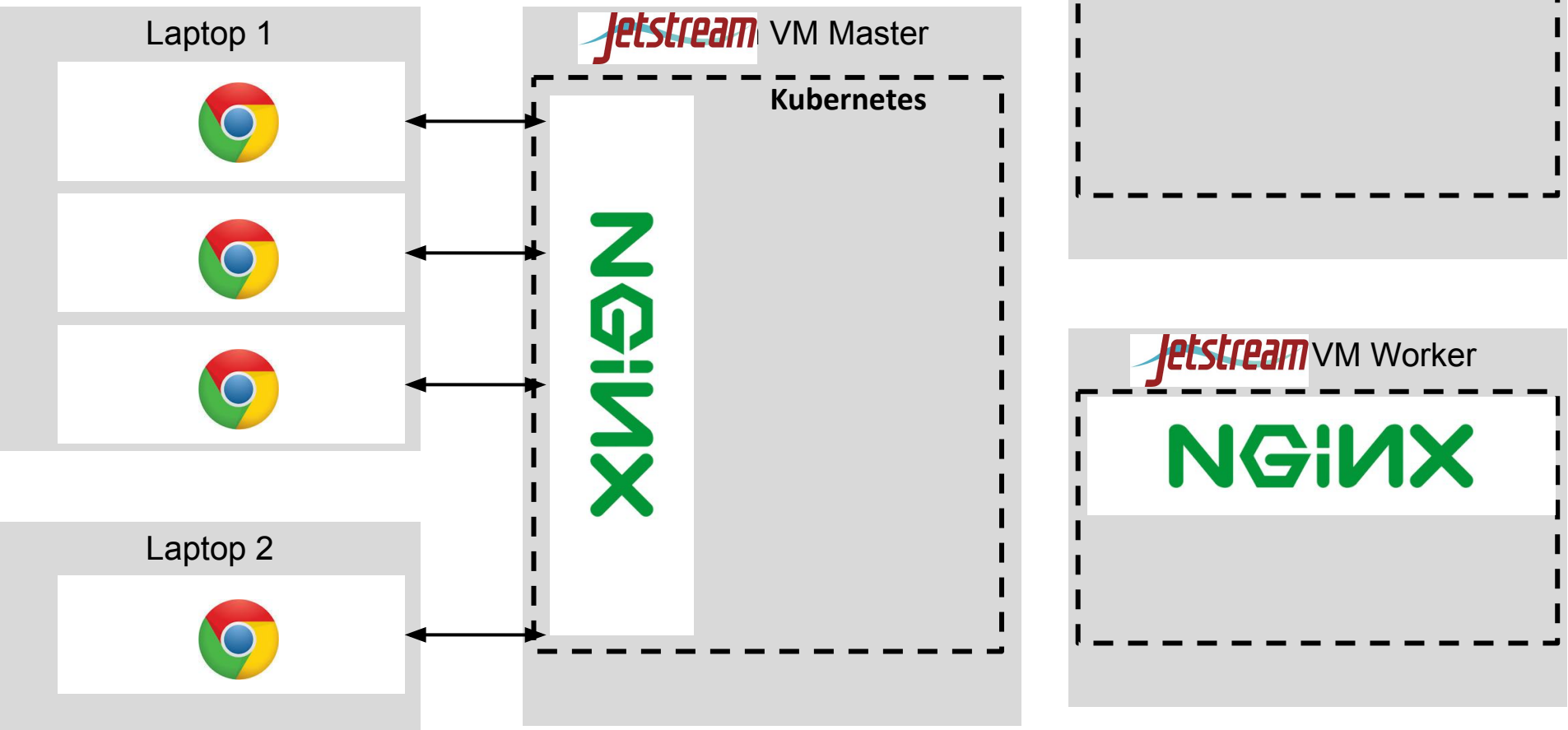
- **JupyterHub on a single node on Jetstream**
- **Workflow:**
 - User logs in with XSEDE credentials
 - Jupyterhub submits SLURM job
 - Jupyter Notebook running on Comet computing node proxied to the user



Kubernetes on Jetstream

- **Think of Kubernetes like SLURM**

- Installed on tens of Jetstream Virtual Machines
- Connect to the master node
- Launch "services" instead of jobs, services are processes running inside Docker containers
- Kubernetes launches containers somewhere in the cluster
- Kubernetes monitors, balances, restarts, kills services
- i.e. Launch a web server service of 2 containers



Jupyterhub/Kubernetes on Jetstream

- **Jupyterhub runs as a Kubernetes service**
- **Users authenticate with Github/Google/XSEDE**
- **Users Jupyter Notebooks run as other containers**
- **Persistent home folders with Rook distributed filesystem**

