
Jupyter Notebooks at scale for Gateways and Workshops

Andrea Zonca SDSC

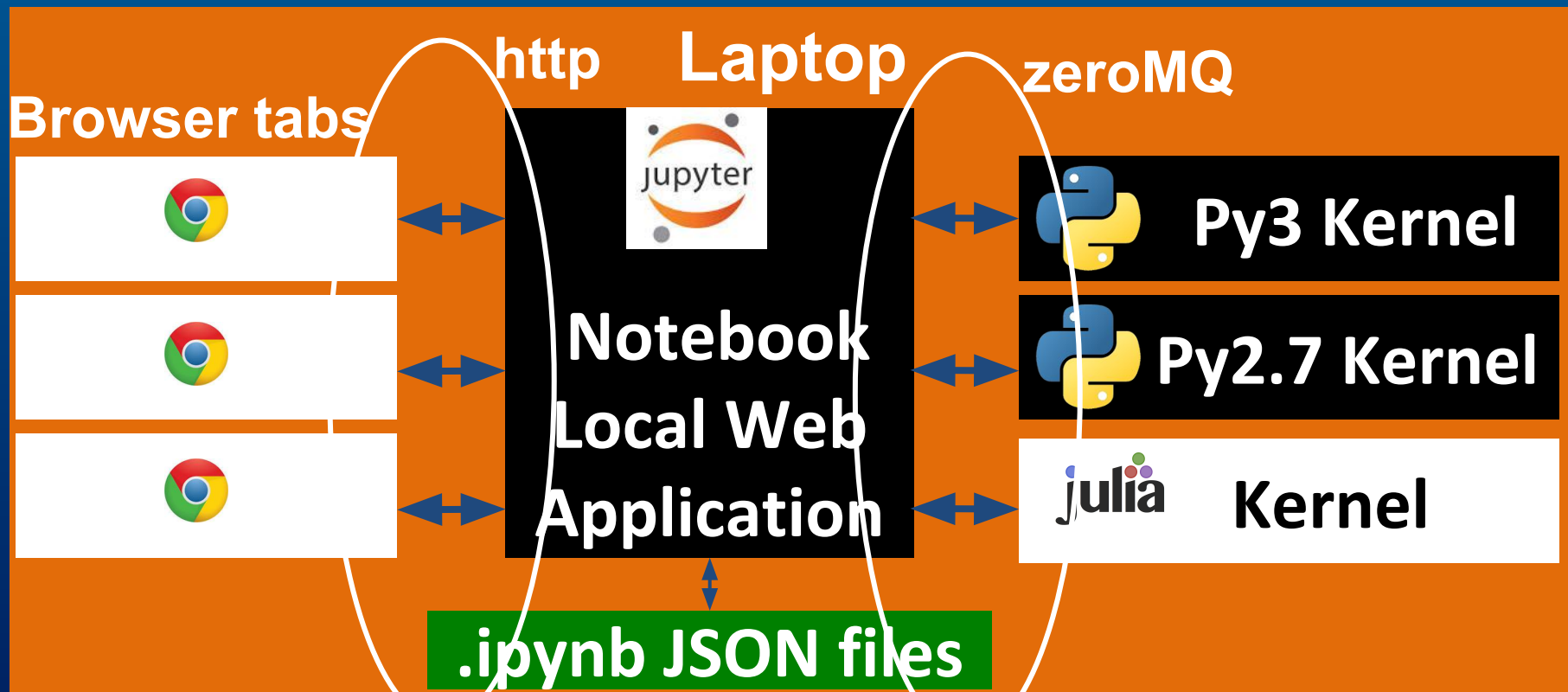
Slides and links: bit.ly/ecss_zonca

Jupyter Notebooks

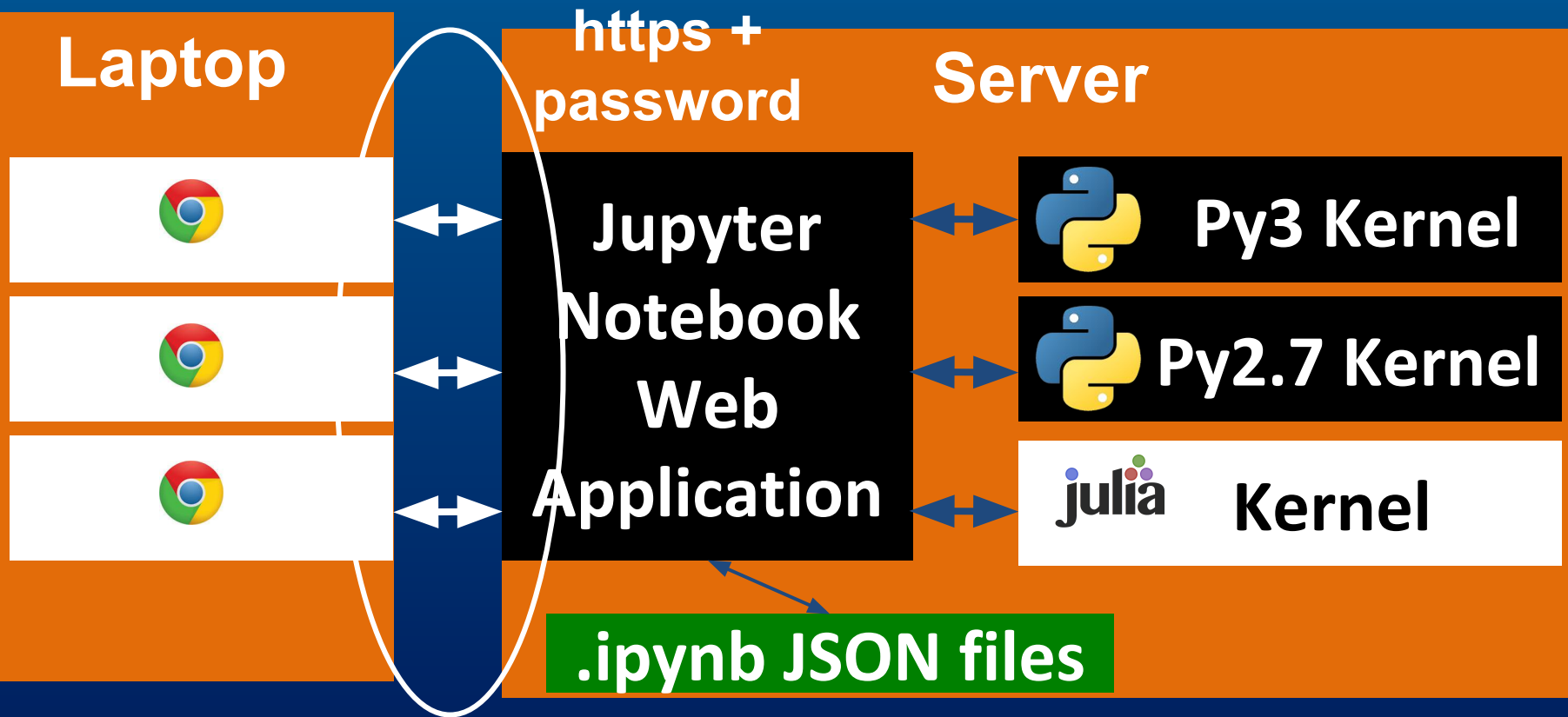
LIGO Open Science Center Gravitational Waves detection:

- [View Notebook](#)
- [Execute Notebook in the cloud](#)

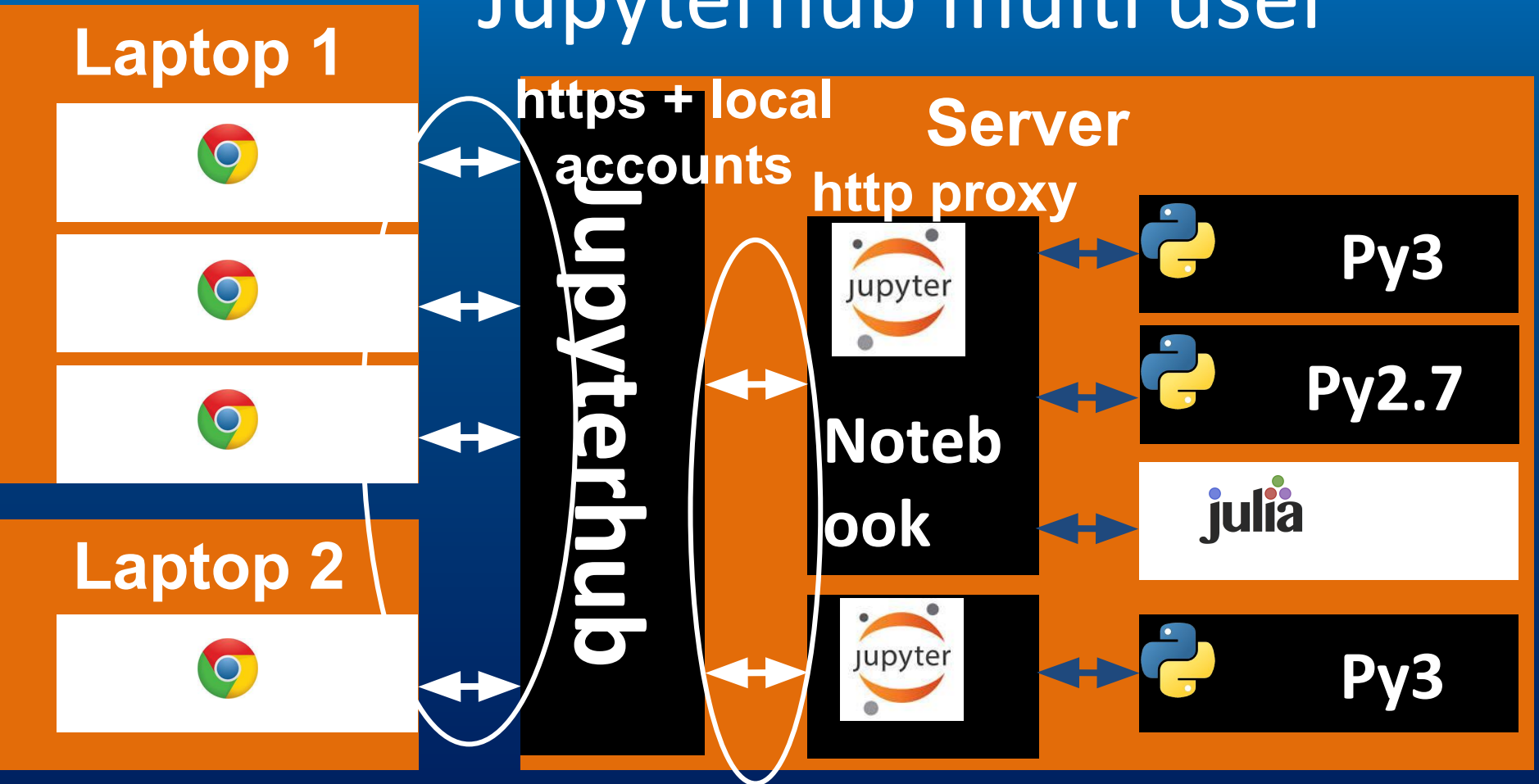
Jupyter notebook local



Jupyter notebook remote



Jupyterhub multi user



Default JupyterHub deployment

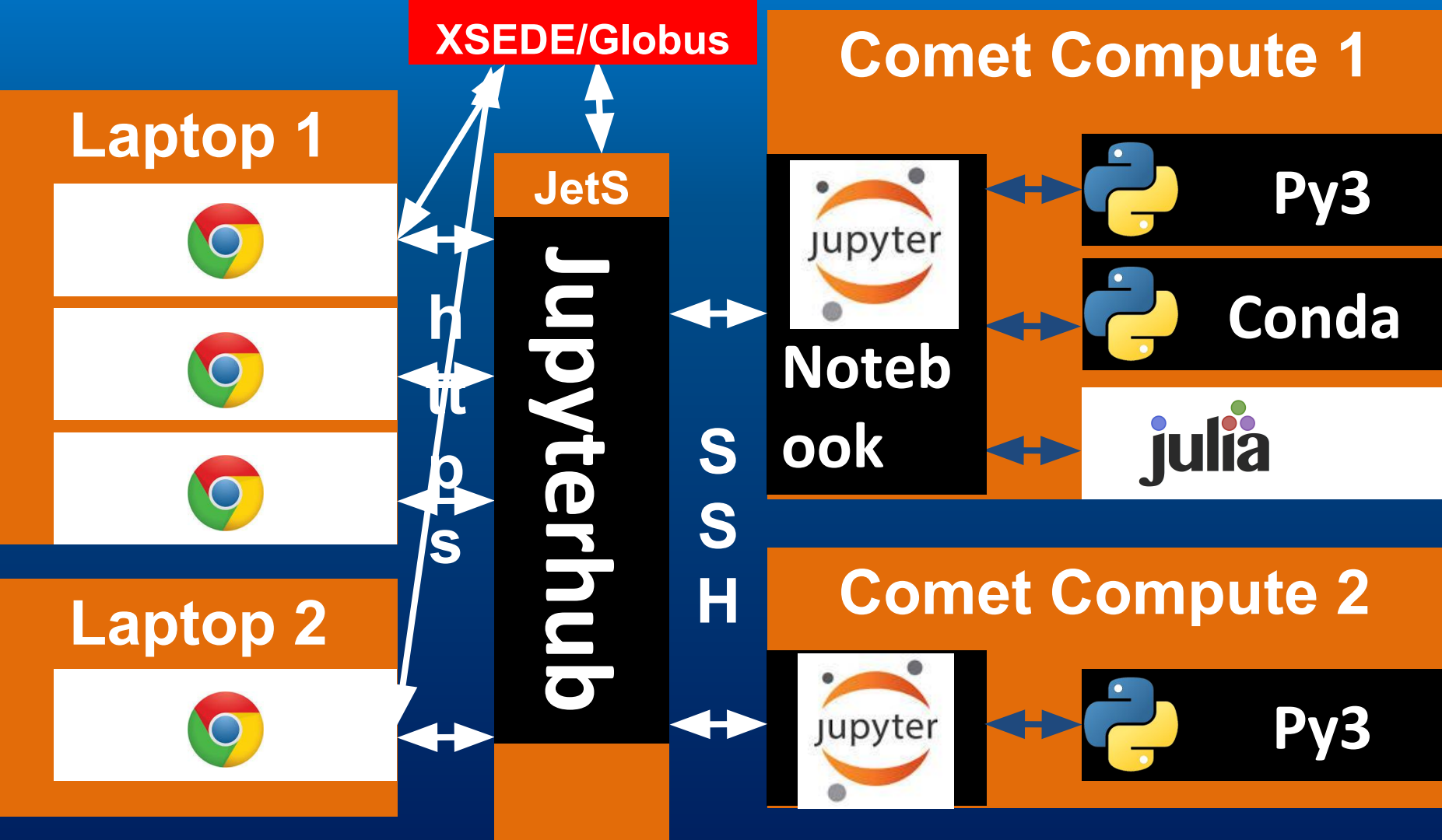
- **Automatically deploy with Ansible:**
github.com/jupyterhub/jupyterhub-deploy-teaching
- **1 server with Ubuntu**
- **NGINX proxy + HTTPS with letsencrypt**
- **Authentication with Local Unix accounts**
- **Launch Notebooks as normal processes**

Authenticators

- **XSEDE credentials**
- **Globus**
- **Github**
- **CILogon (Campus credentials)**
- **Google Accounts**
- **many more...**

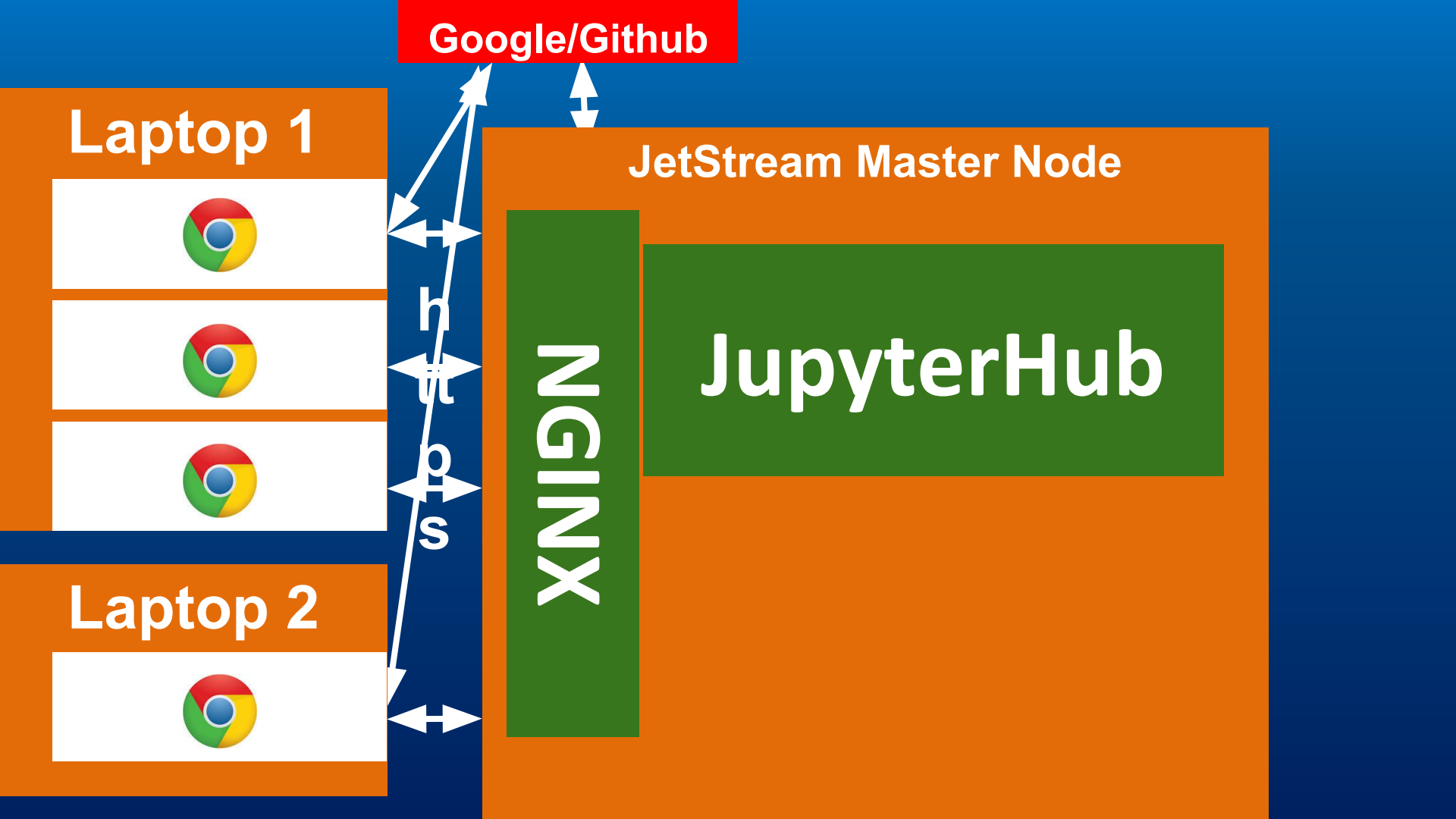
Spawners

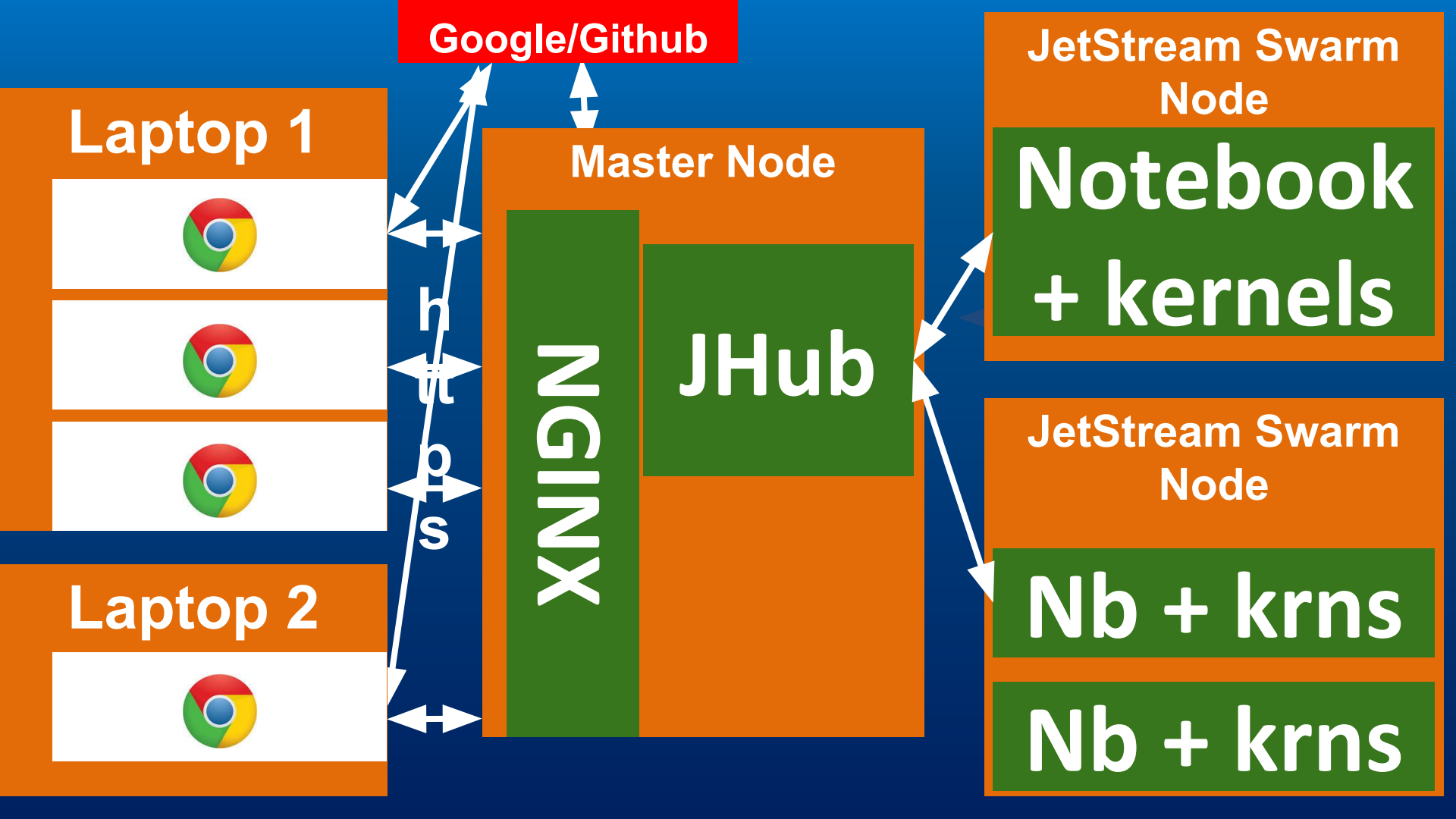
- **batchspawner: launch via SLURM/PBS/SGE...**
- **SwarmSpawner: launch in Docker containers via Swarm**
- **KubeSpawner: launch on Kubernetes pods**



Spawn Notebooks on Supercomputers

- **Authentication with XSEDE or Globus**
- **batchspawner**
 - SSH with community account **or** GSISSSH with XSEDE token
 - Submit job to SLURM
 - Wait for job to start
 - Connect back to JupyterHub with SSH tunnel
- **[Link to tutorial](#)**





Google/Github

Laptop 1



Laptop 2



Master Node

NGINX

JHub

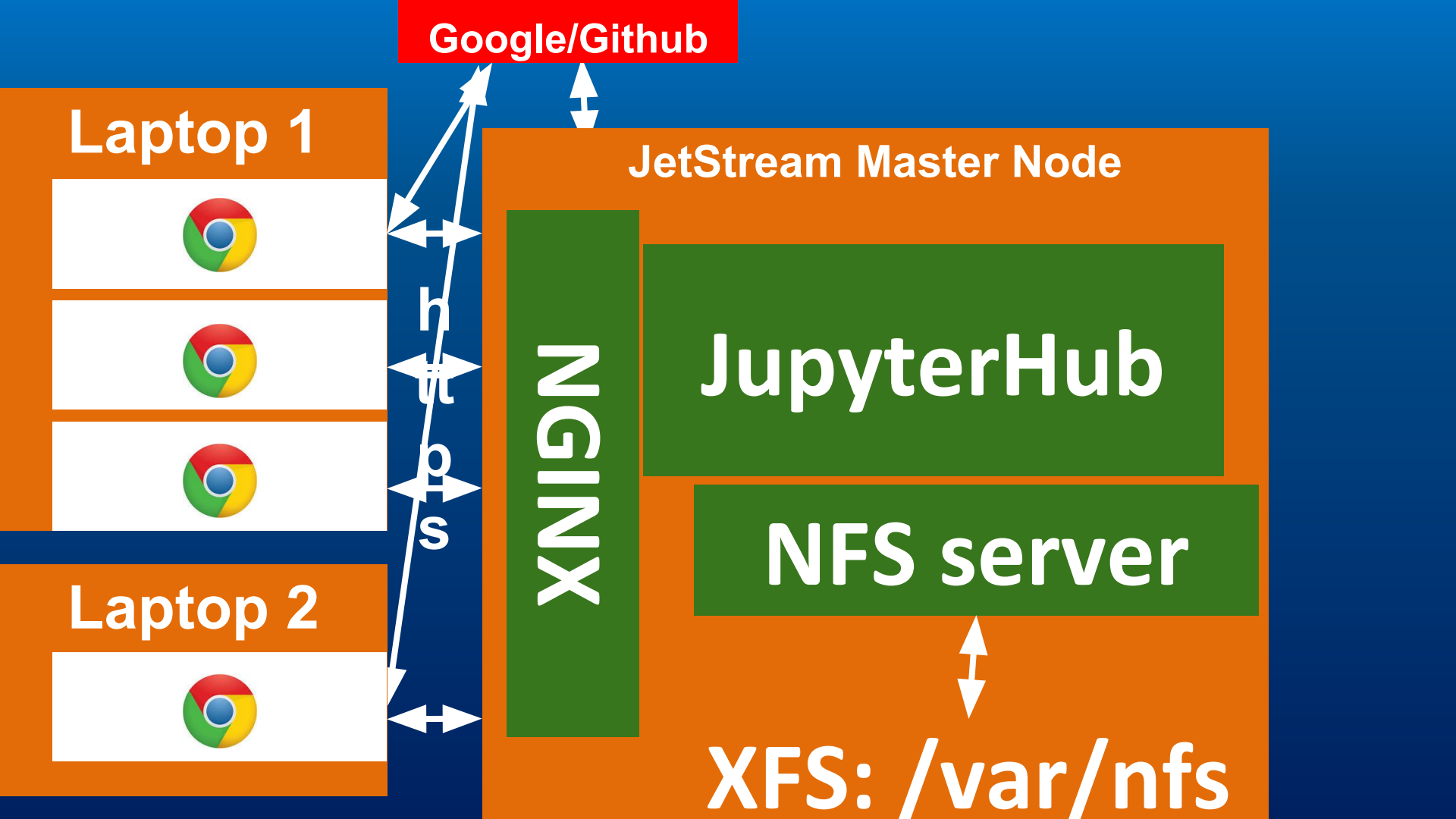
JetStream Swarm Node

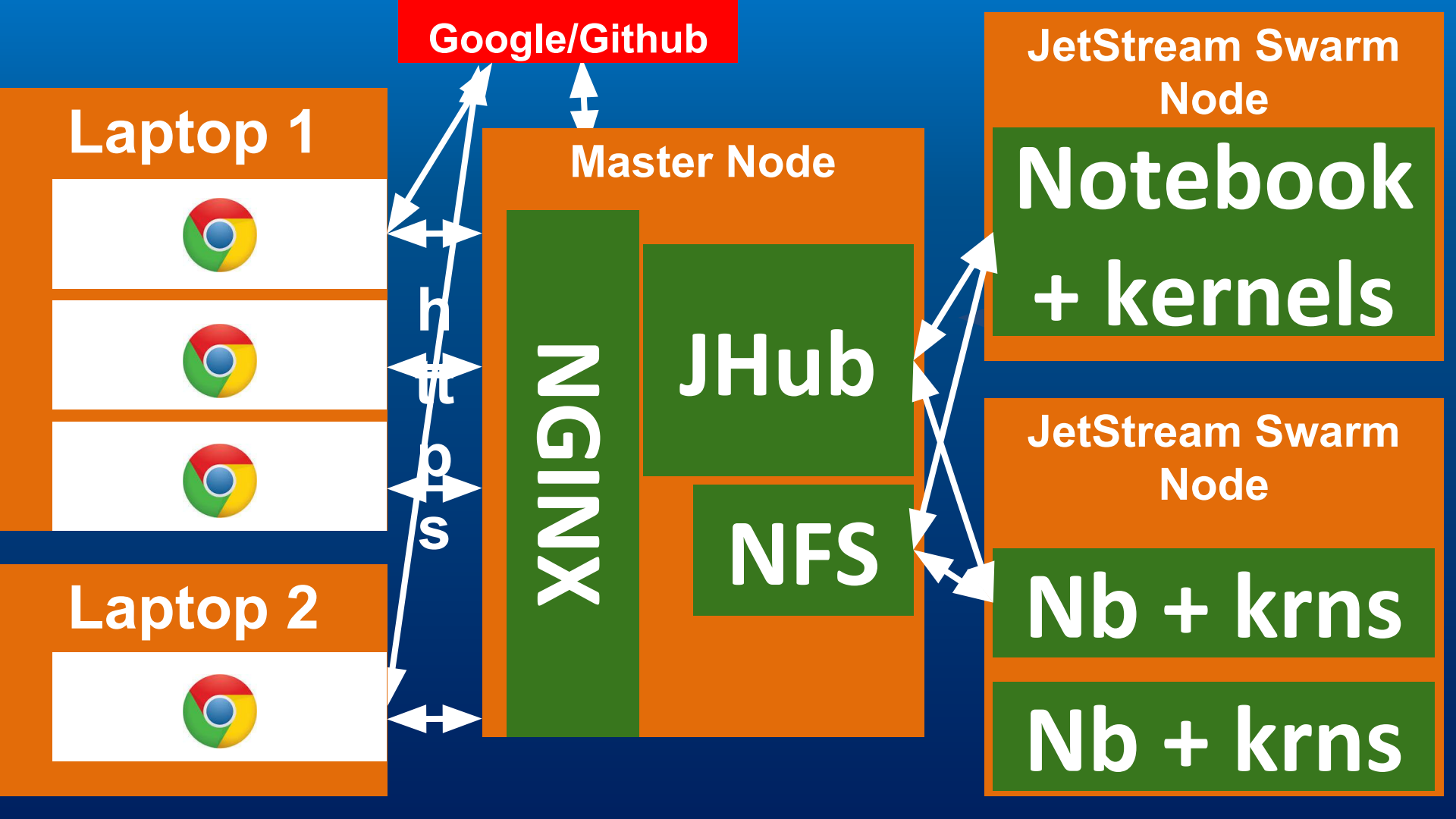
Notebook + kernels

JetStream Swarm Node

Nb + krns

Nb + krns





Google/Github

Laptop 1



Laptop 2



Master Node

NGINX

JHub

NFS

JetStream Swarm Node

Notebook + kernels

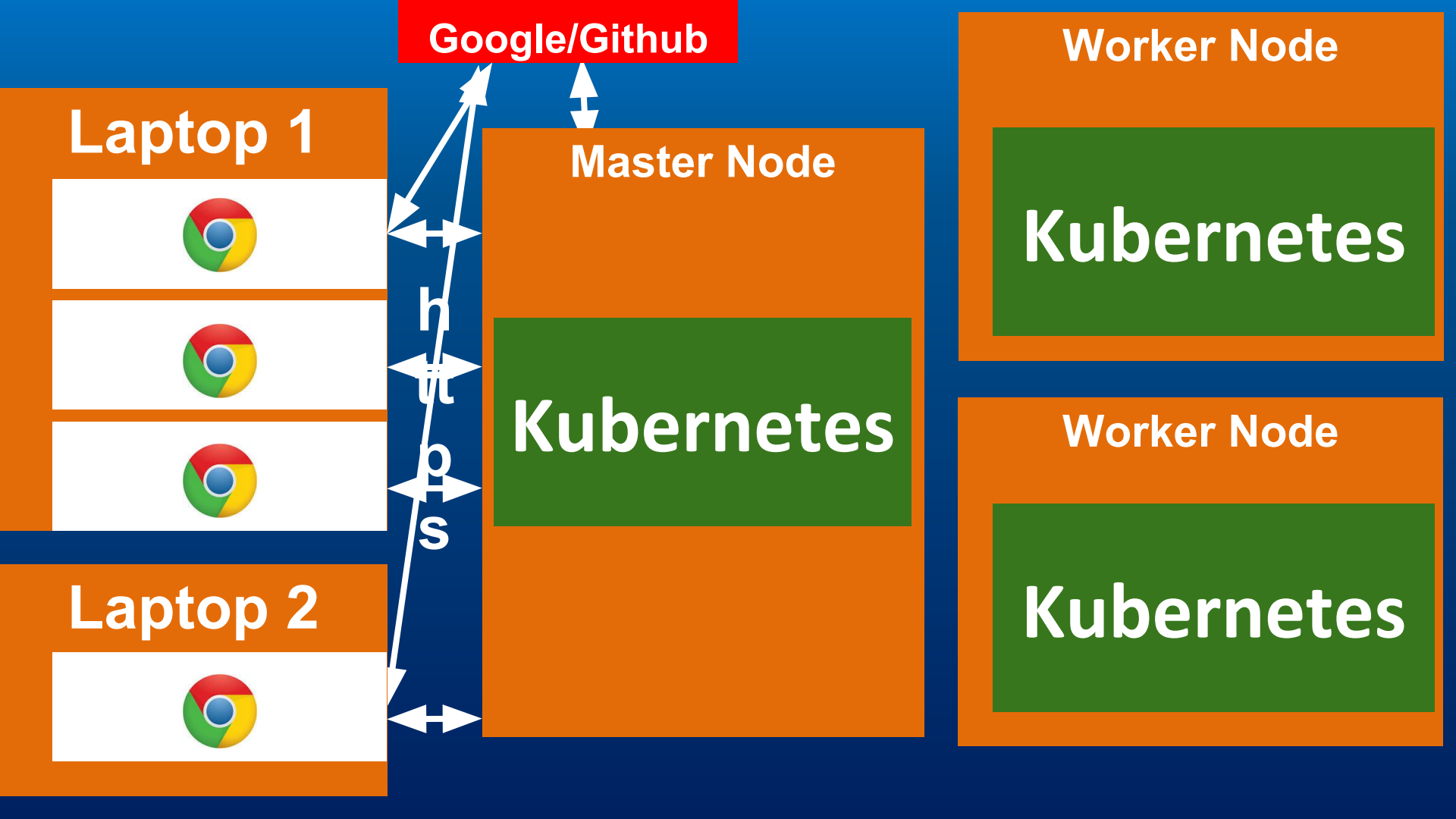
JetStream Swarm Node

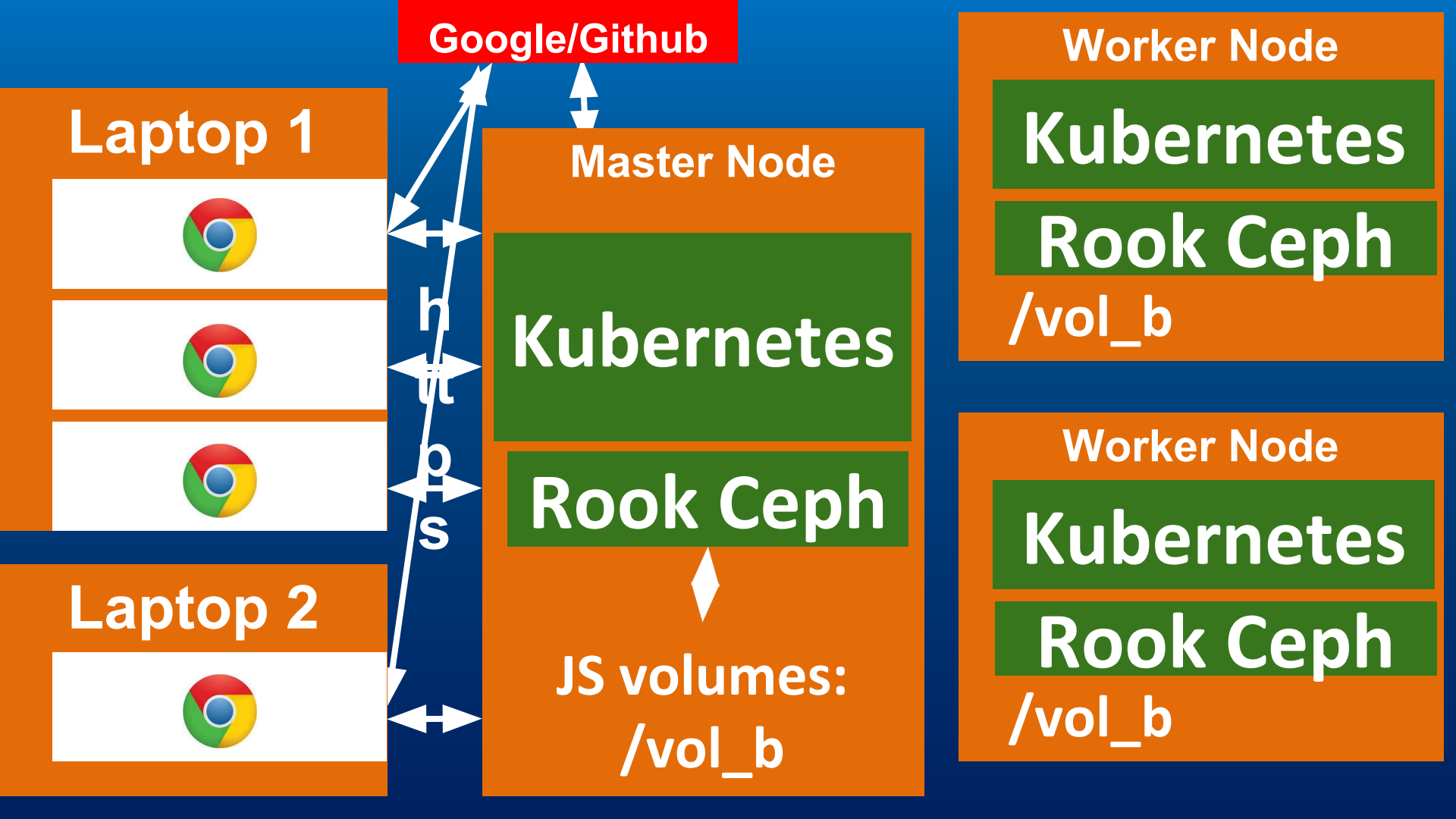
Nb + krns

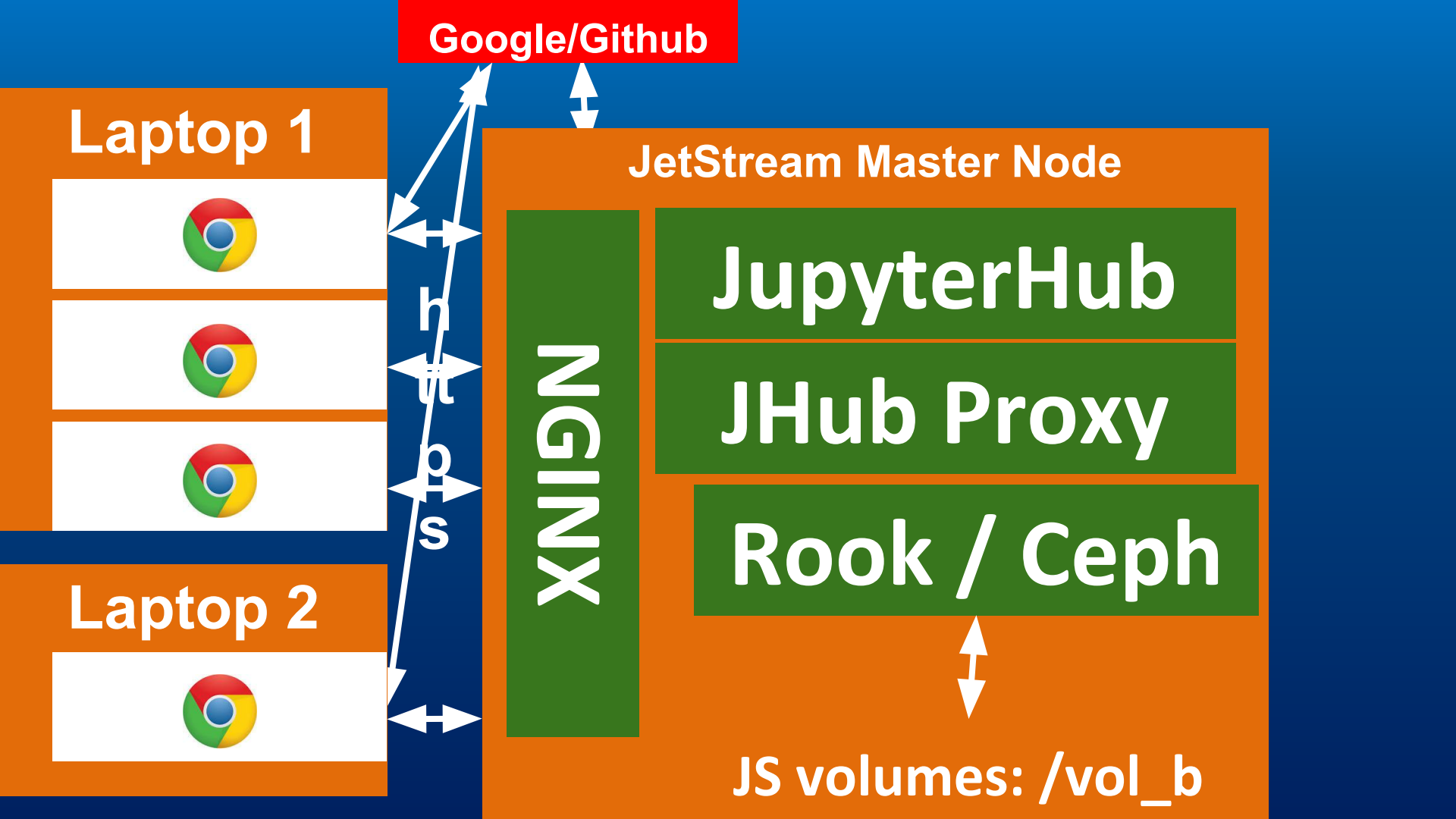
Nb + krns

Cloud deployment JetStream/Swarm

- **Master node, Docker containers:**
 - NGINX with letsencrypt
 - Jupyterhub
 - NFS server - persistent homes, user quota
- **SwarmSpawner**
 - Managed by Docker Swarm over many servers
- **Scales to tens of servers, <100 users**
- **[Link to tutorial](#)**







Google/Github

Laptop 1



requests

JetStream Master Node

NGINX

JupyterHub

JHub Proxy

Rook / Ceph

Laptop 2



JS volumes: /vol_b

JetStream Master Node

NGINX

JupyterHub

JHub Proxy

Worker Node

Notebook

Worker Node

Notebook

Notebook

JetStream Master Node

NGINX

JupyterHub

JHub Proxy

Rook / Ceph



JS volumes: /vol_b

Worker Node

Notebook

Rook Ceph

/vol_b

Worker Node

Notebook

Rook Ceph

/vol_b

JetStream with Kubernetes

- **Kubernetes services (zero-to-jupyterhub):**
 - NGINX with letsencrypt HTTPS
 - Jupyterhub
 - Jupyterhub Proxy
 - Automatically restarted if die
- **KubeSpawner**
 - Launches user Notebooks containers
 - Persistent data with Ceph distributed File System
- **Scales to hundreds of servers, ~1000 users**
- **[Link to tutorial](#)**

Thanks

- **Contact me for anything related to Jupyter on XSEDE**
- **zonca@sdsc.edu**