HOW TO BUILD AN HPC CLUSTER

What it takes to build and control an army of 200+ compute nodes

Dr. Richard Berger

HPC
High-Performance Computing

LET'S GET A HPC CLUSTER!

01

Get enough money

02

Spend the money

03

???

04

Profit

LET'S GET A HPC CLUSTER!

01

Get enough money

02

Spend the money

03

???

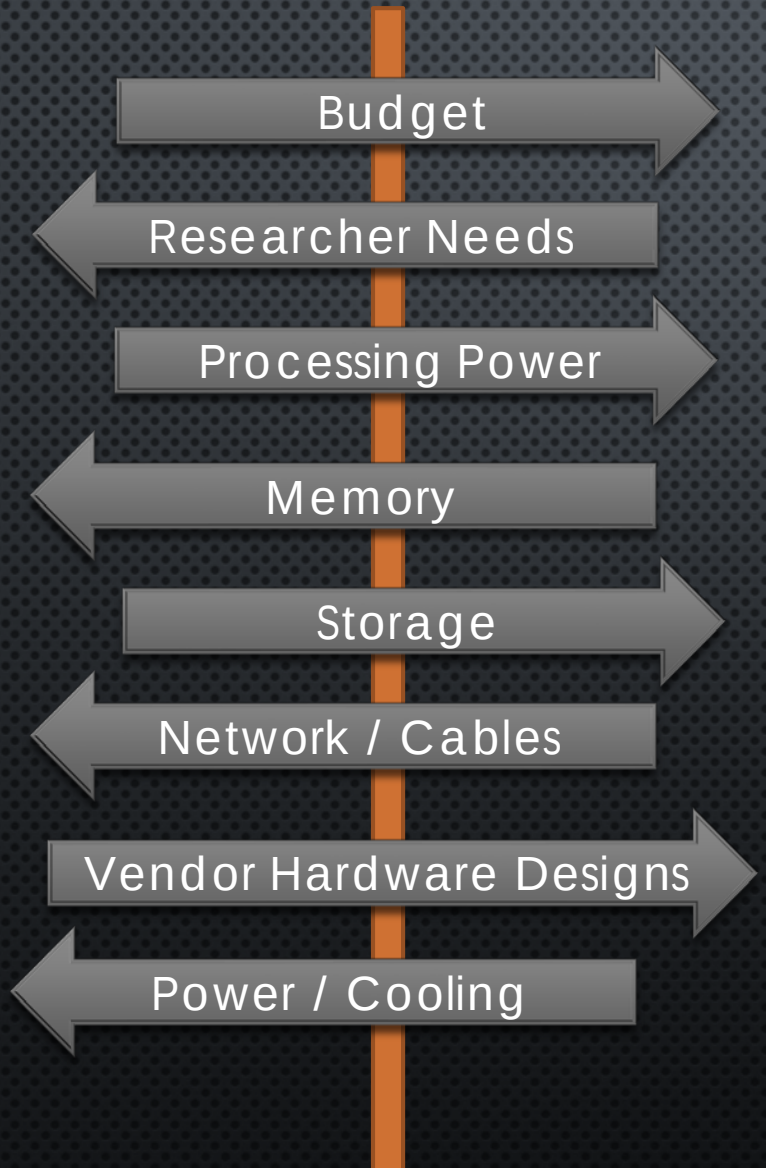
04

Profit

CLUSTER SHOPPING LIST

- Compute Servers
- Network switches
- High-Speed Interconnect
- Storage
- Management servers
- Login servers
- Cables

A MULTI-DIMENSIONAL OPTIMIZATION PROBLEM



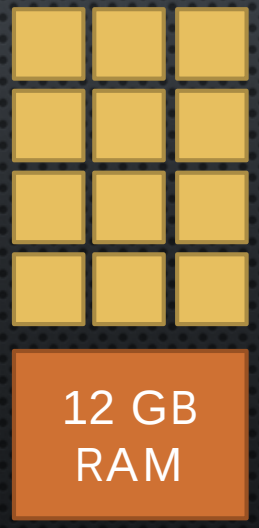
- Changing one has consequences to others
- You need to find compromises and marry different requirements
- Whatever we buy, we have to live with for up to **8 years**
- → a lot of "What if?" discussions

OWL'S NEST 2

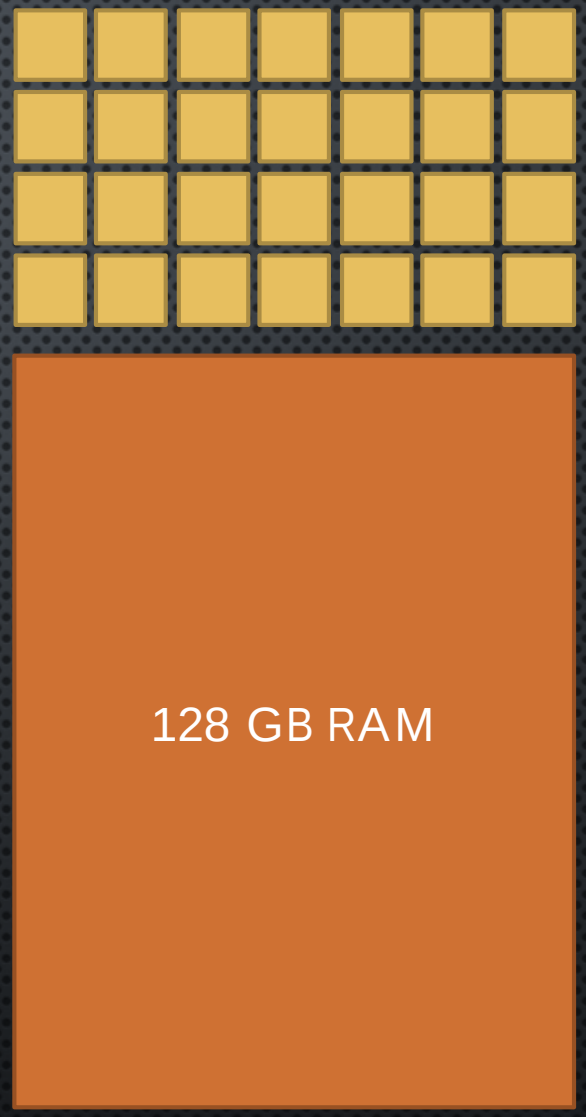
Temple University's latest HPC cluster

PROCESSING POWER & MEMORY

A typical desktop PC

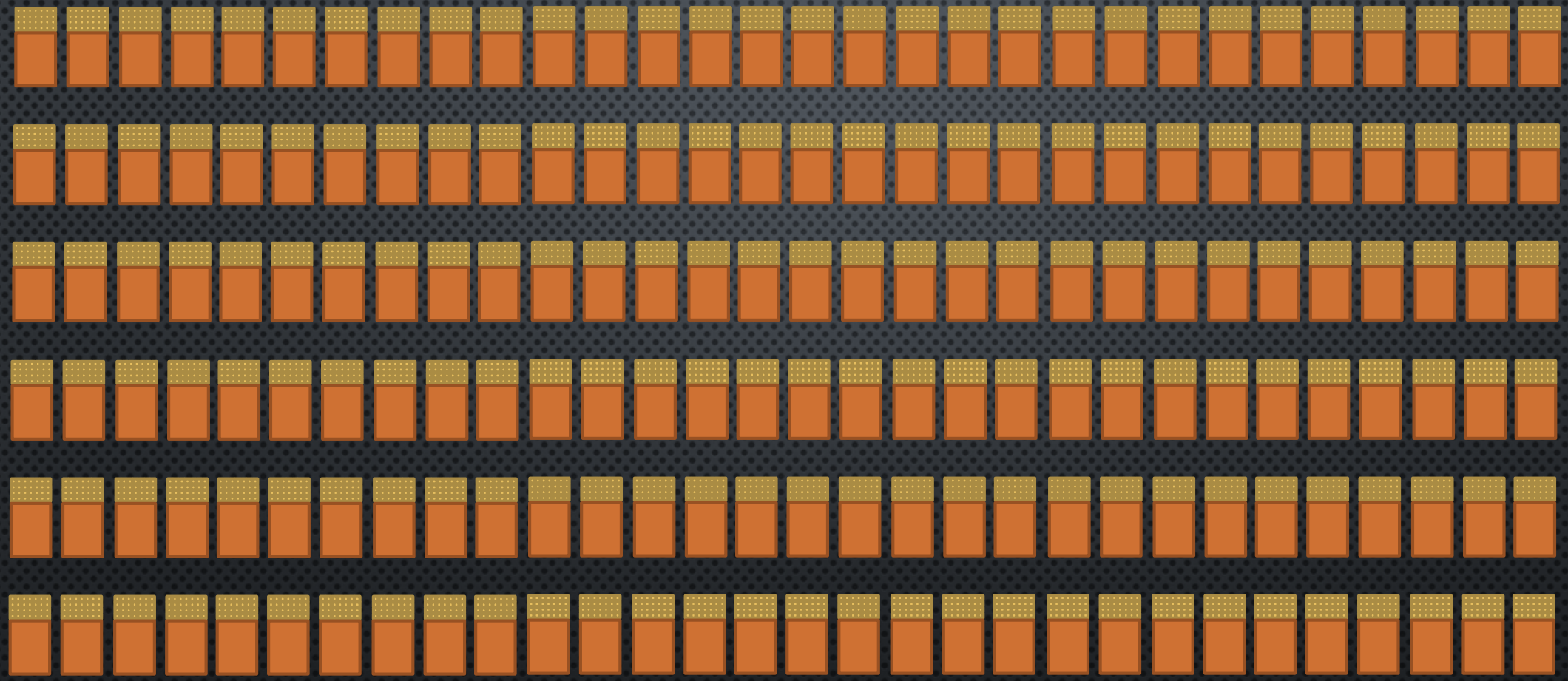


Owl's Nest 1
normal node



Owl's Nest 2
normal node

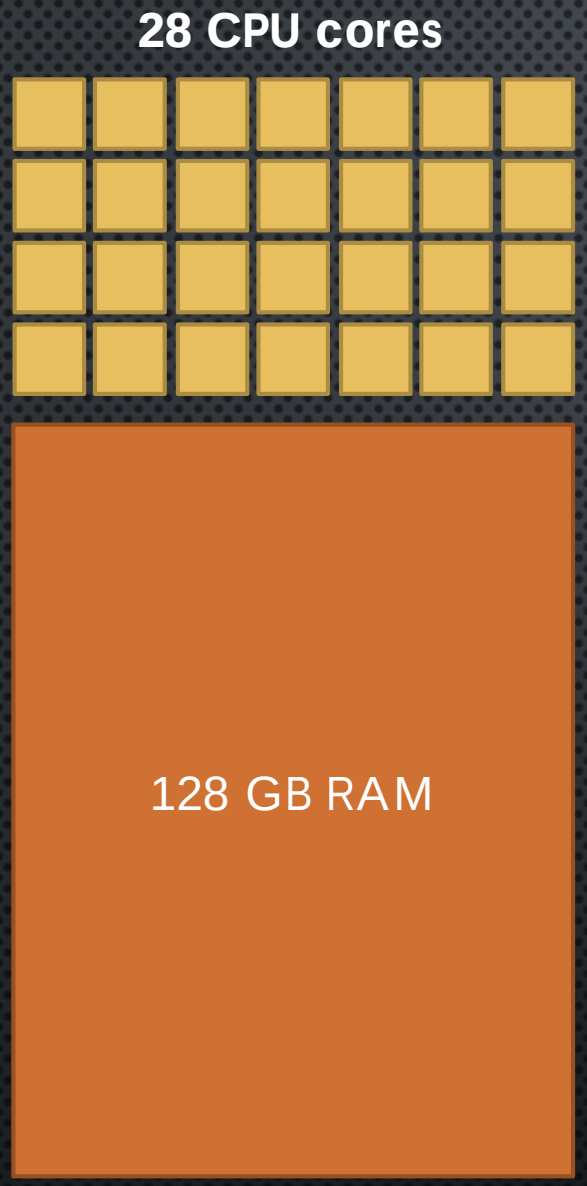
Owl's Nest 2: **180 normal** compute nodes



5040 TOTAL CPU cores

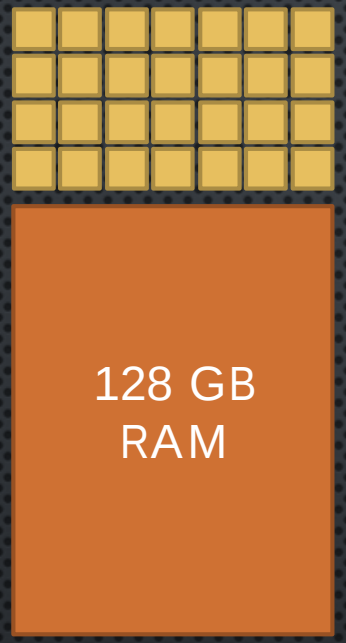
22.5 TB TOTAL RAM

Parallel Computing
Designed for Throughput



Owl's Nest 2
normal node

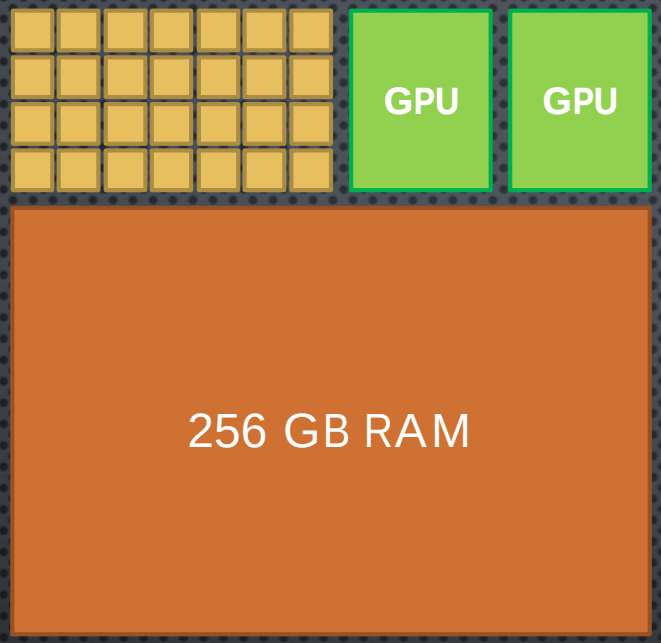
28 CPU cores



Owl's Nest 2
normal node

180x

28 CPU cores / 7168 GPU cores

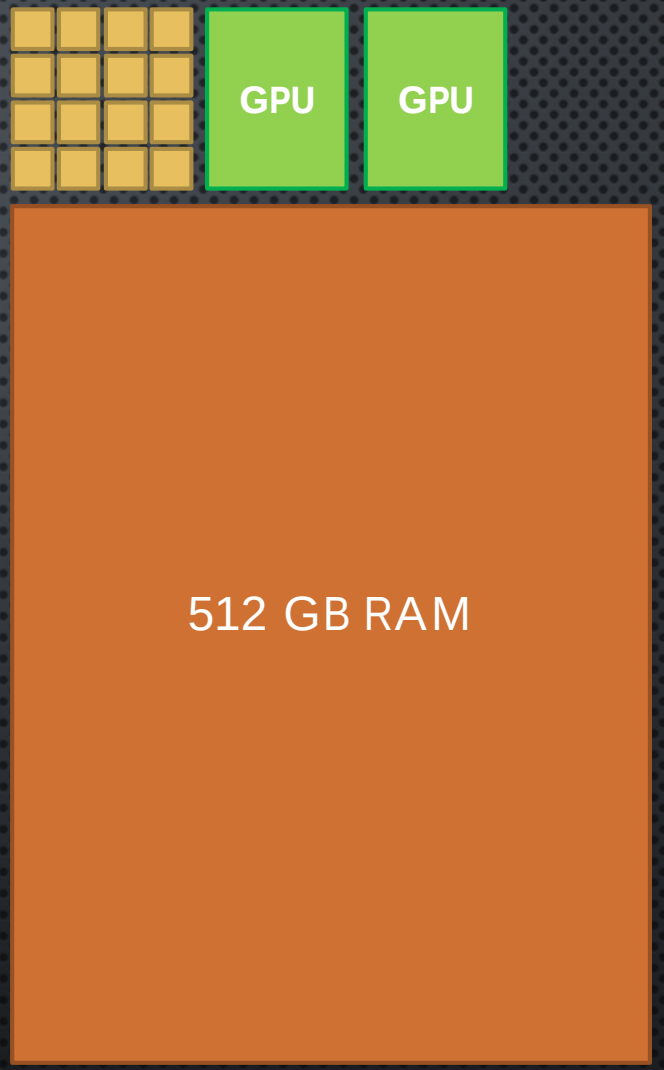


Owl's Nest 2
qcd node

12x

exclusive for
Department of Physics
(**q**uantum **c**hromo**d**ynamics)

16 CPU cores / 7168 GPU cores



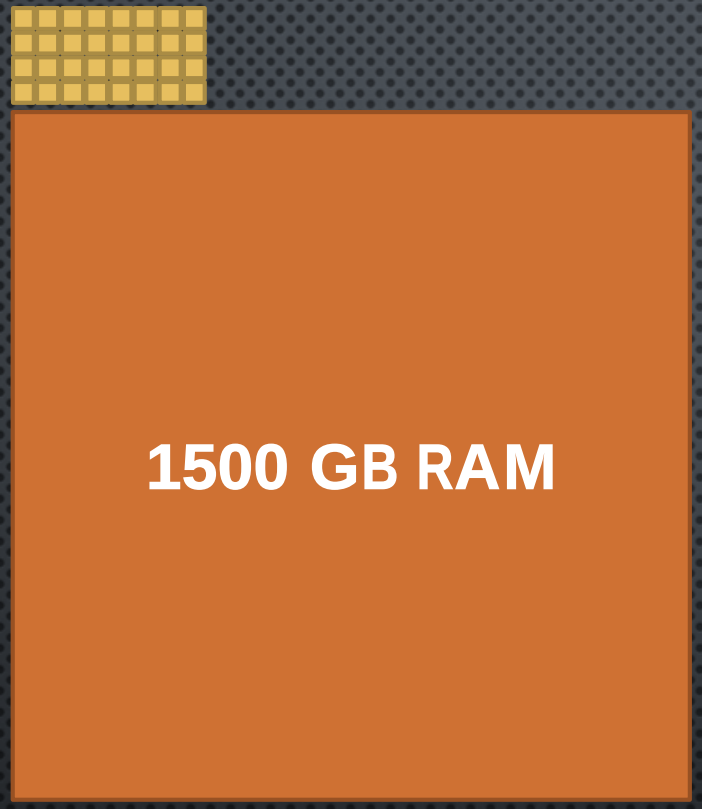
Owl's Nest 2
large node

6x

16 CPU cores /
7168 GPU cores

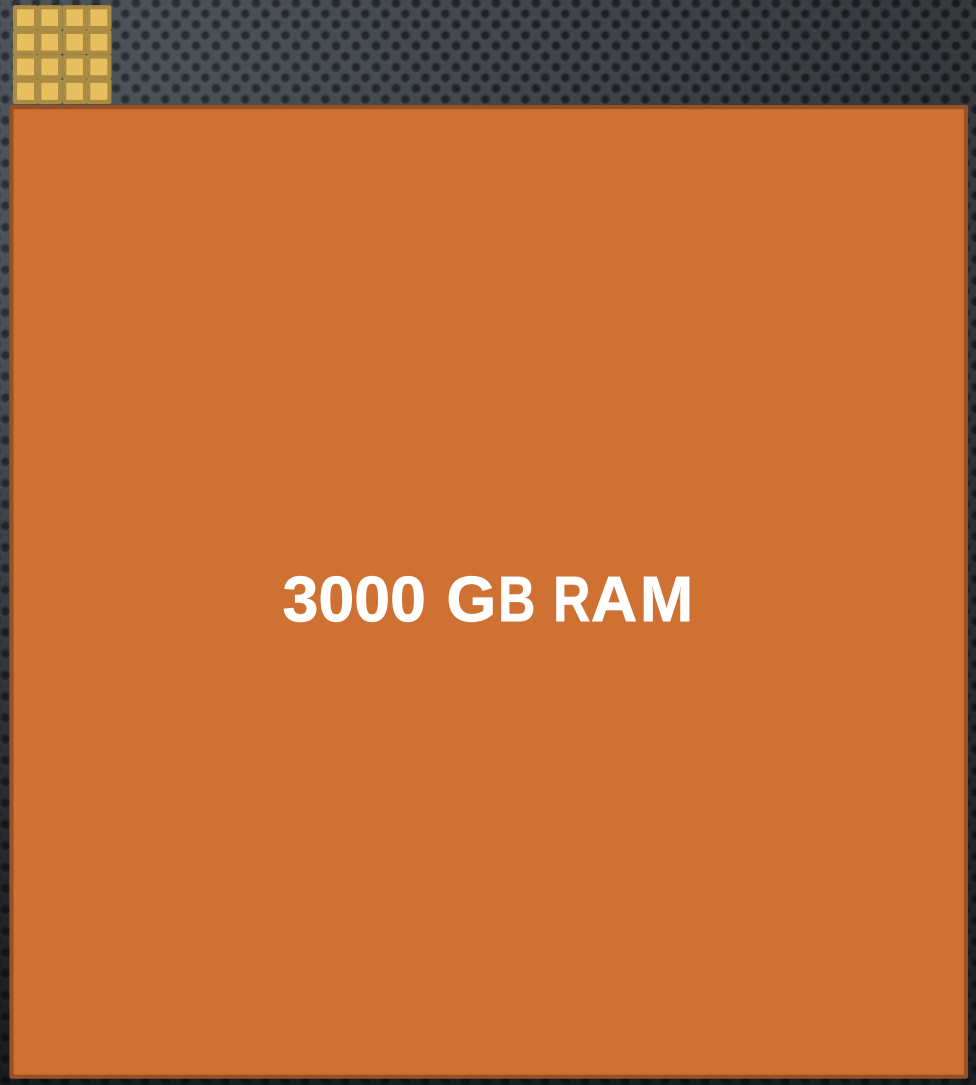
Owl's Nest 2
large node
6x

32 CPU cores



Owl's Nest 2
big node
6x

16 CPU cores



Owl's Nest 2
huge node
2x

GLOBAL DISK STORAGE

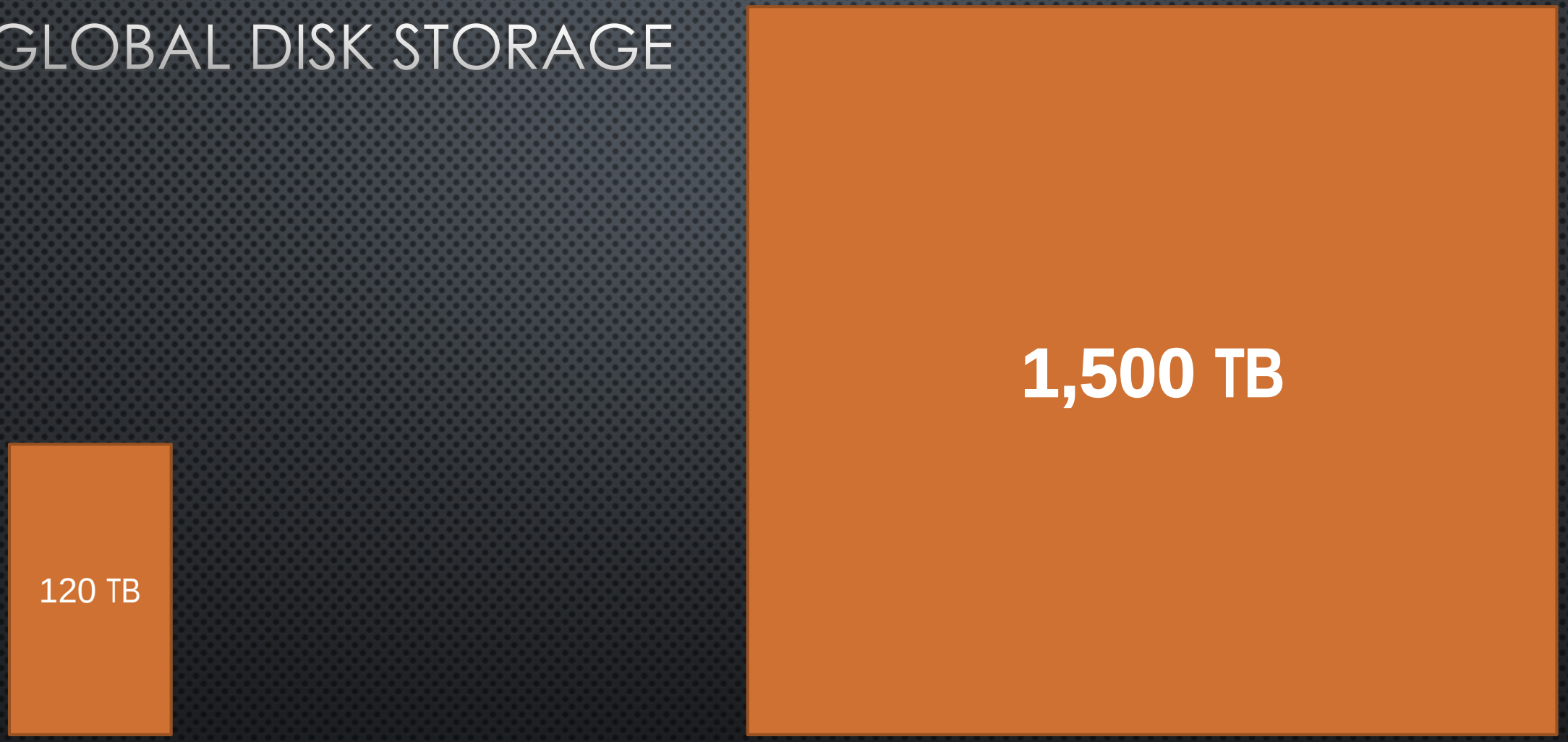
30 TB
30 TB
30 TB
30 TB

Owl's Nest 1
4x NFS Server
Bandwidth: 4x 1Gb/s

30 TB	30 TB	30 TB	30 TB	30 TB
30 TB	30 TB	30 TB	30 TB	30 TB
30 TB	30 TB	30 TB	30 TB	30 TB
30 TB	30 TB	30 TB	30 TB	30 TB
30 TB	30 TB	30 TB	30 TB	30 TB
30 TB	30 TB	30 TB	30 TB	30 TB
30 TB	30 TB	30 TB	30 TB	30 TB
30 TB	30 TB	30 TB	30 TB	30 TB
30 TB	30 TB	30 TB	30 TB	30 TB
30 TB	30 TB	30 TB	30 TB	30 TB

Owl's Nest 2
DDN GridScaler Parallel Storage with GPFS
Bandwidth: **4x 56 Gb/s** FDR InfiniBand

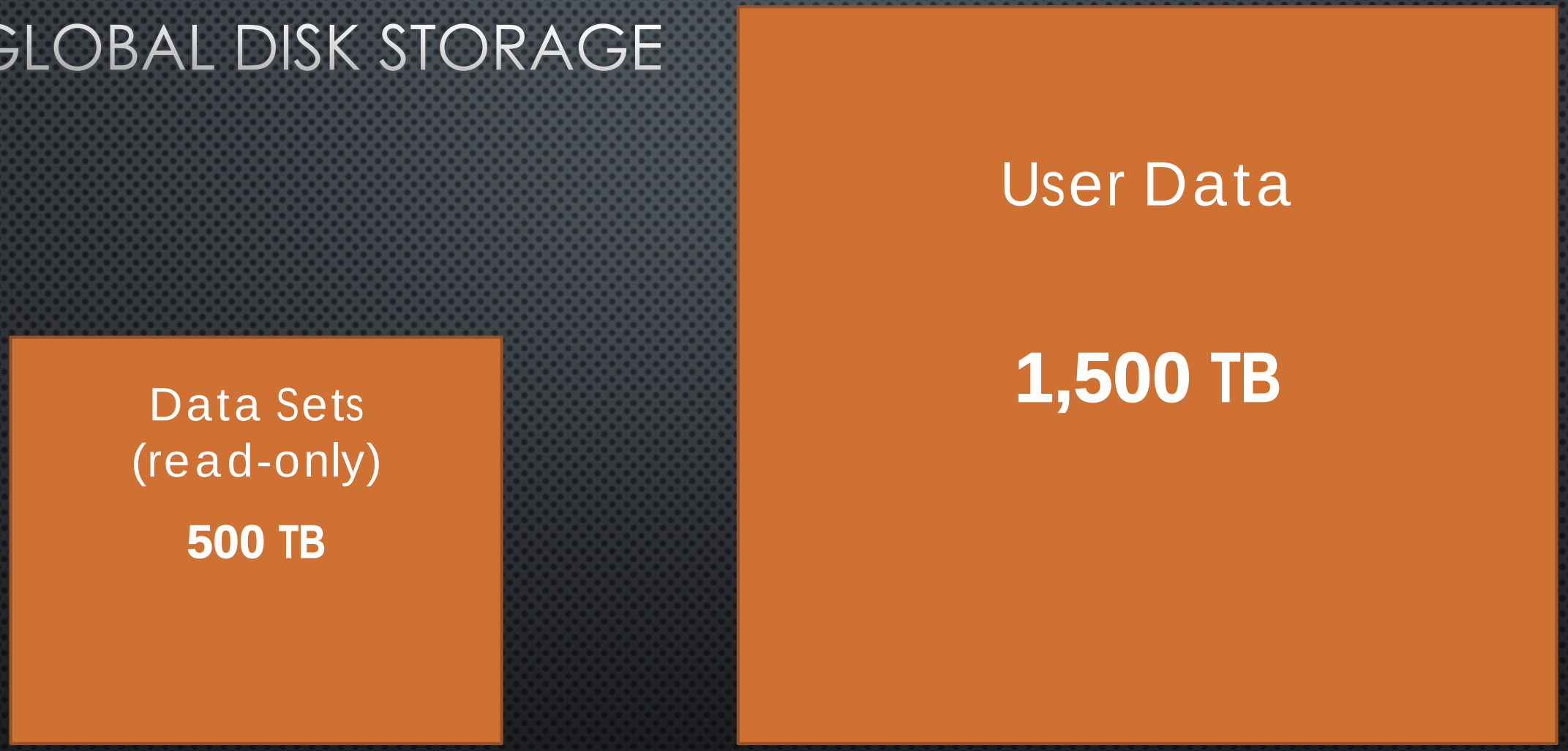
GLOBAL DISK STORAGE



Owl's Nest 1
4x NFS Server
Bandwidth: 4x 1Gb/s

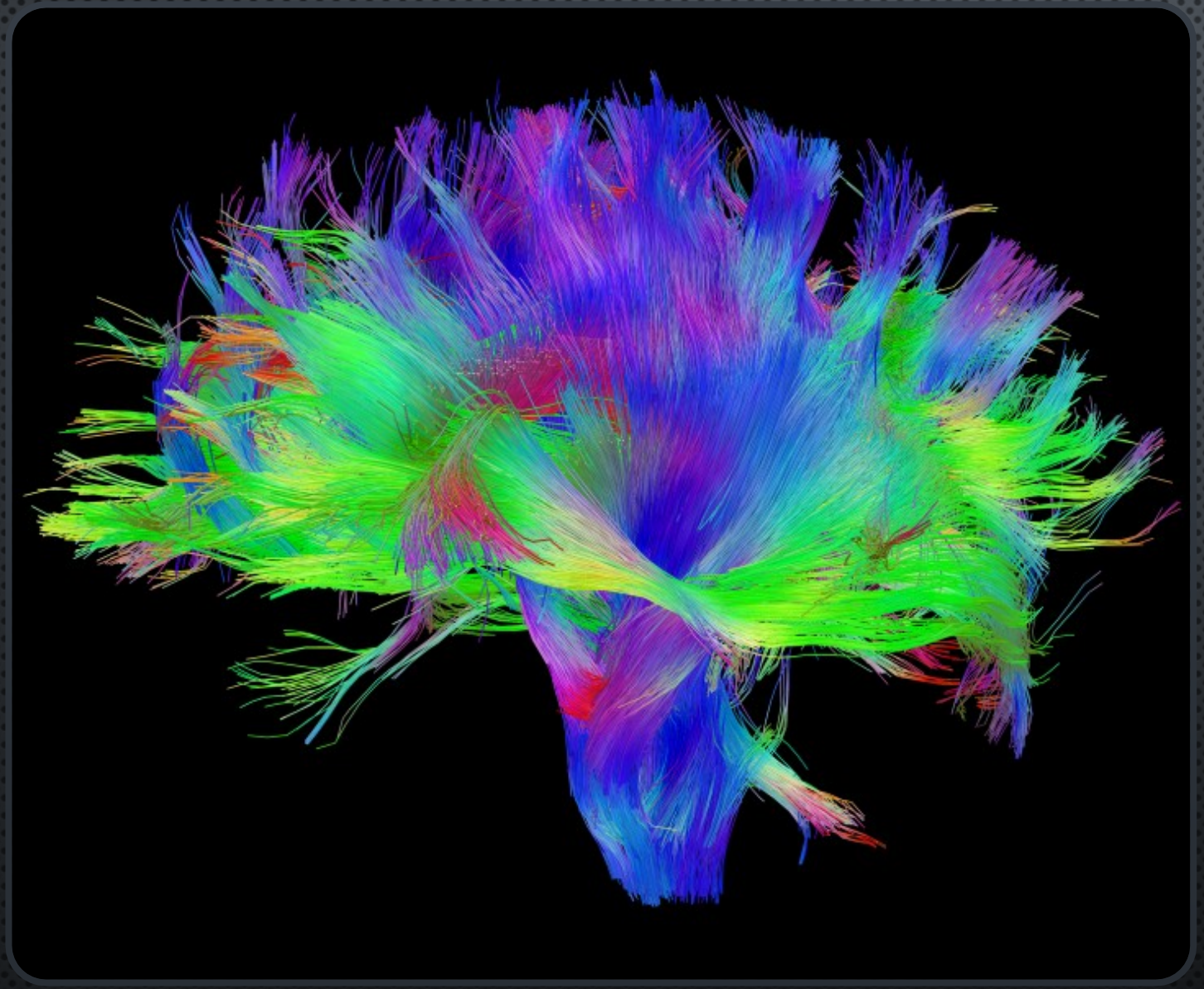
Owl's Nest 2
DDN GridScaler Parallel Storage with GPFS
Bandwidth: **4x 56 Gb/s** FDR InfiniBand

GLOBAL DISK STORAGE



Owl's Nest 2
DELL NFS server with ZFS
Bandwidth: 1x100Gb/s EDR InfiniBand

Owl's Nest 2
DDN GridScaler Parallel Storage with GPFS
Bandwidth: **4x 56 Gb/s** FDR InfiniBand



Department of Psychology

Research based on the

HUMAN CONNECTOME PROJECT

DATA SET OF **64 TB**

MRI scans of the human brain

CHALLENGES

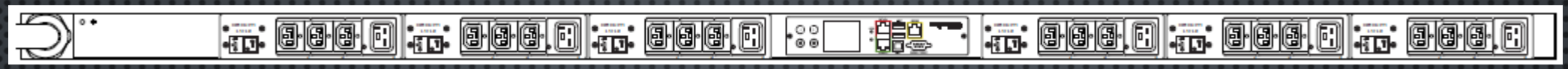
Power

Cooling

Logistics

48A/60A

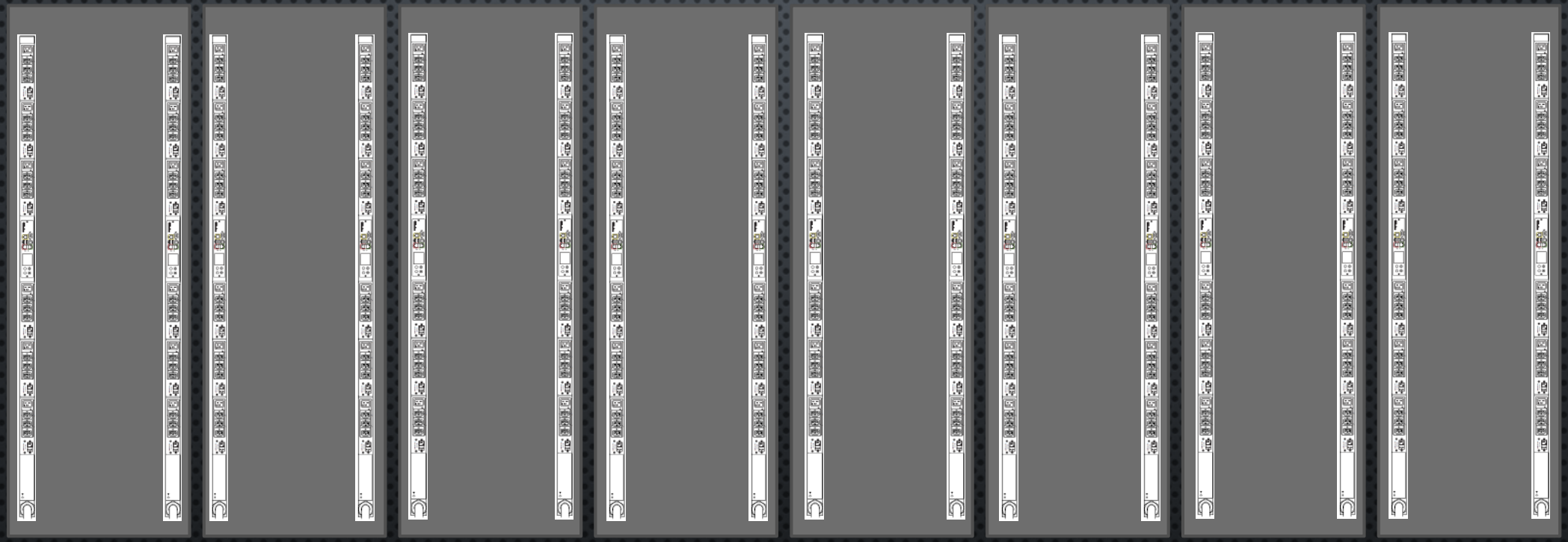
17.3kVA



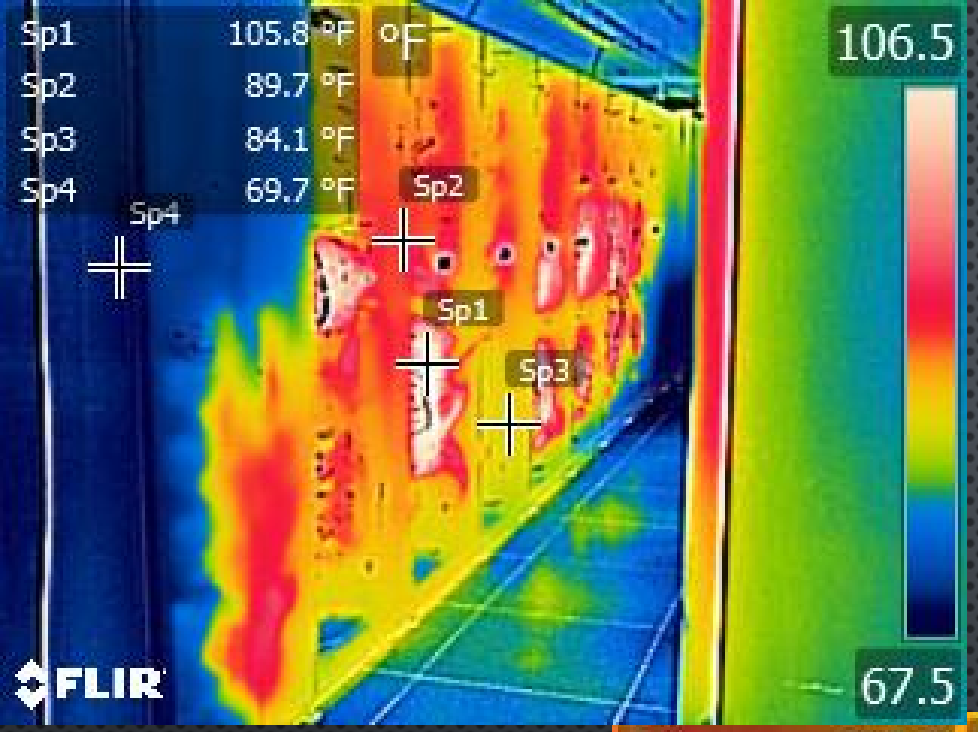
18x C14 plugs, 6x C19 plugs

POWER

Total Initial Estimate: **108KW**



Needed:	19KW	19KW	19KW	13KW	12KW	14KW	12KW	6KW
	32x C14 plugs	34x C14 plugs	32x C14 plugs	26x C14 plugs	24x C14 plugs	12x C14 plugs	15x C14 plugs 4x C19 plugs	10x C14 plugs



CHALLENGES

Power

Cooling

Logistics


Logistics: How to turn this...


...unpack it all...

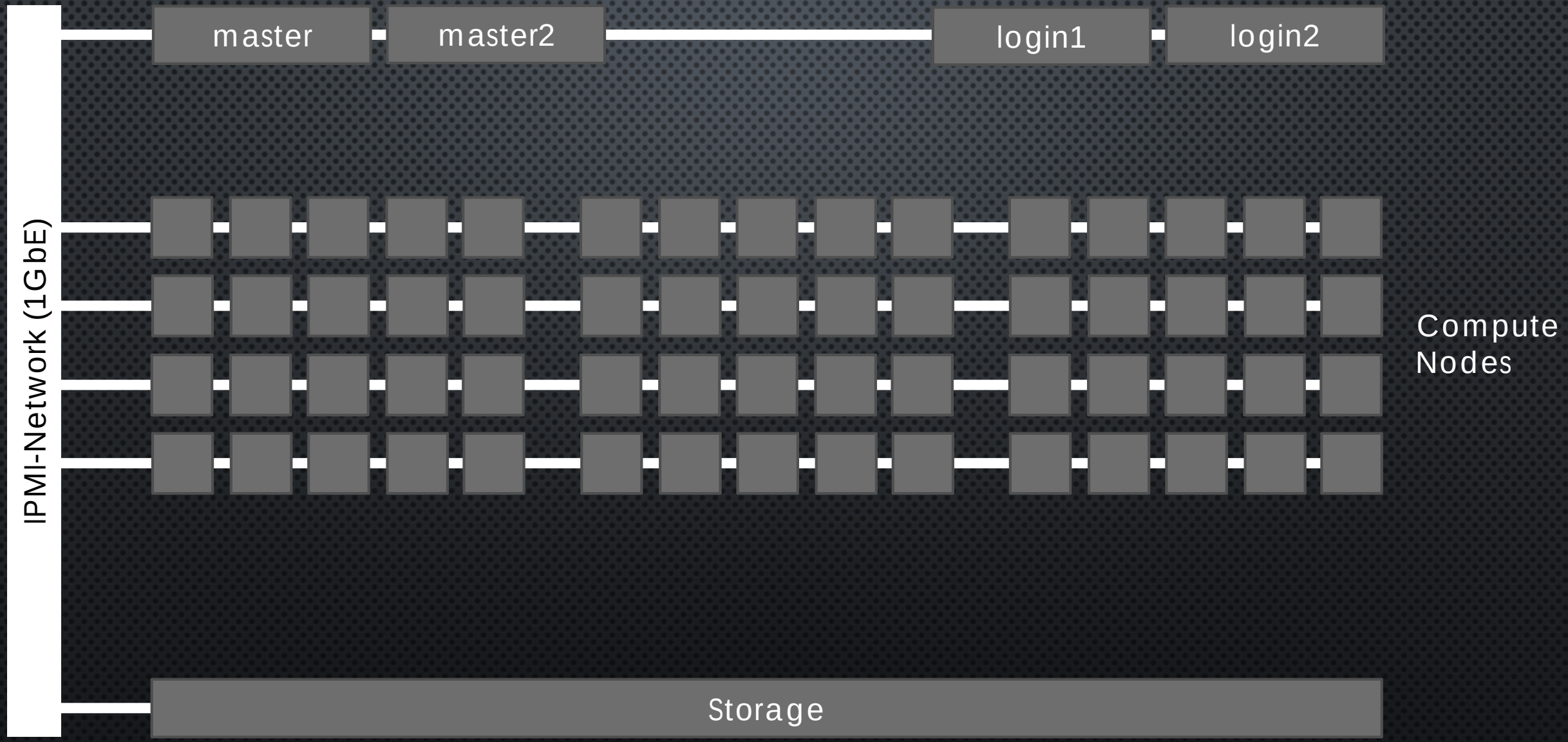
...move around multiple metric tons of equipment...


...and turn it all into a fully functioning cluster

Building the new Owl's Nest

HPC
High-Performance Computing

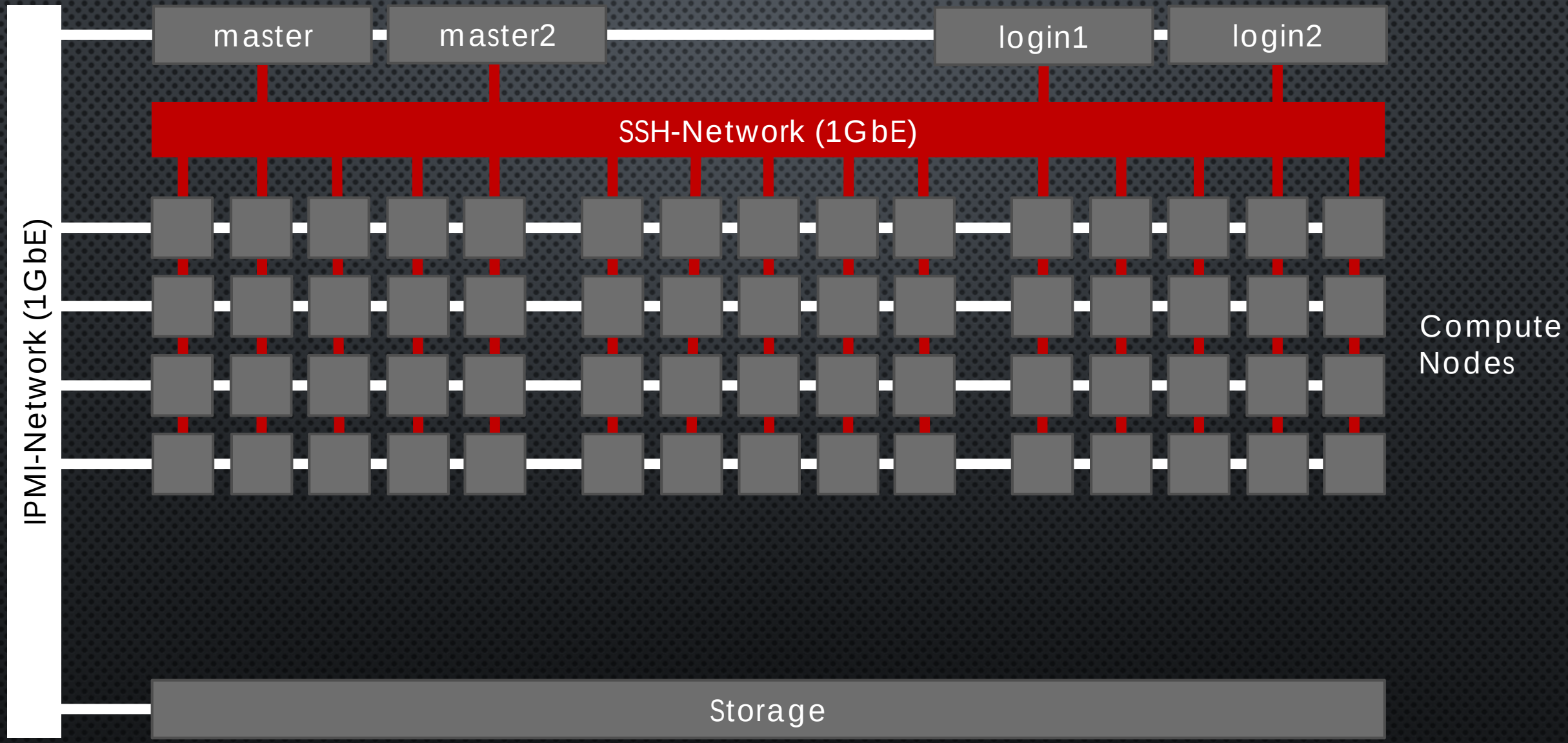
NETWORK OVERVIEW



IPMI-NETWORK

- An isolated 1GbE ethernet network
- Only master node can access it
- All servers have a **B**oard **M**anagement **C**ontroller (BMC) connected to this network
- BMC is always accessible, **even if server is off**
- You can control a server via its BMC
  - Power on/off/cycle
  - Read sensor data
  - Update BIOS/Firmware
  - Change BIOS settings

NETWORK OVERVIEW

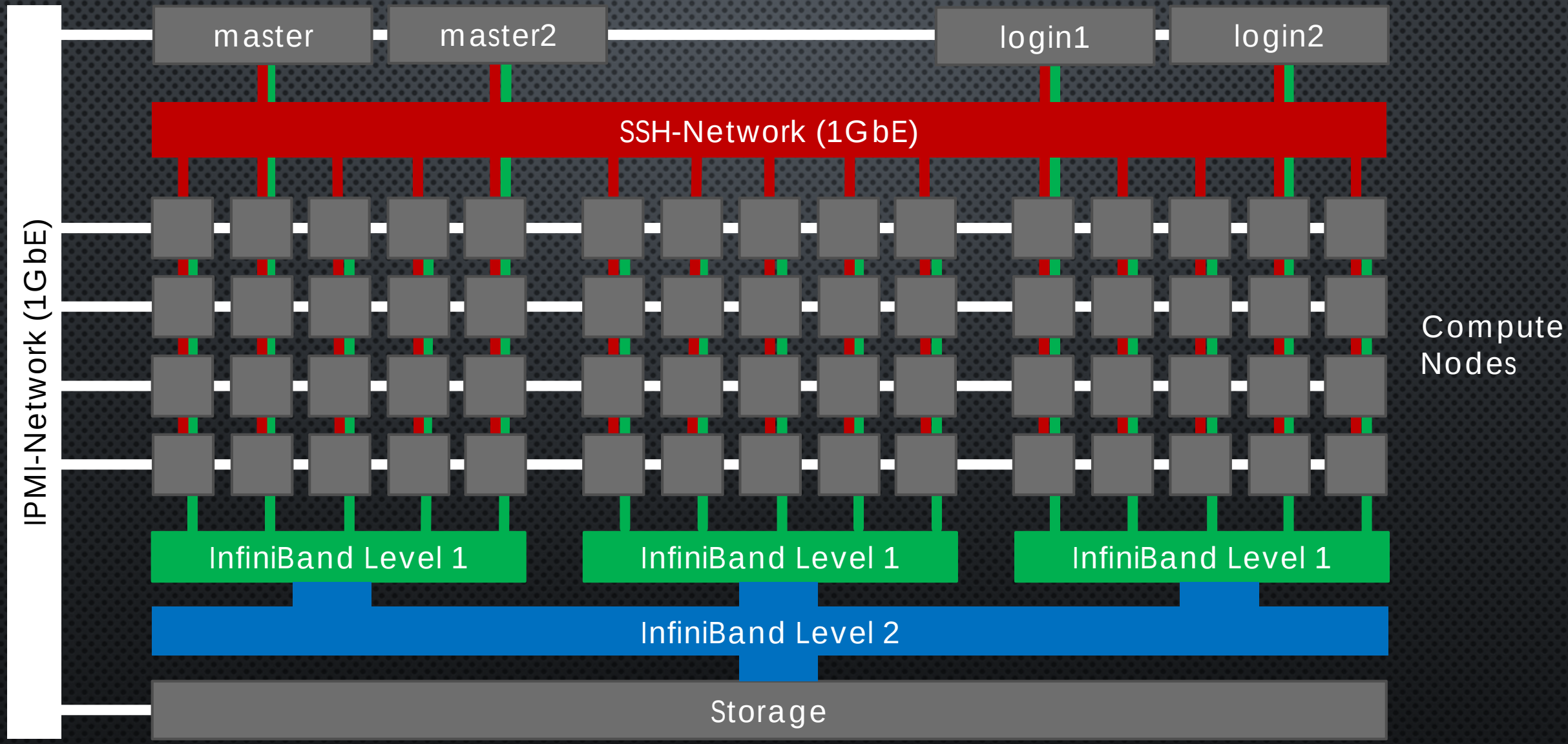


SSH-NETWORK

- A cluster internal 1GbE ethernet network
- Used by Batch-System to control nodes
- Users can log into nodes that run their jobs

- Used for provisioning / Network install of nodes

NETWORK OVERVIEW



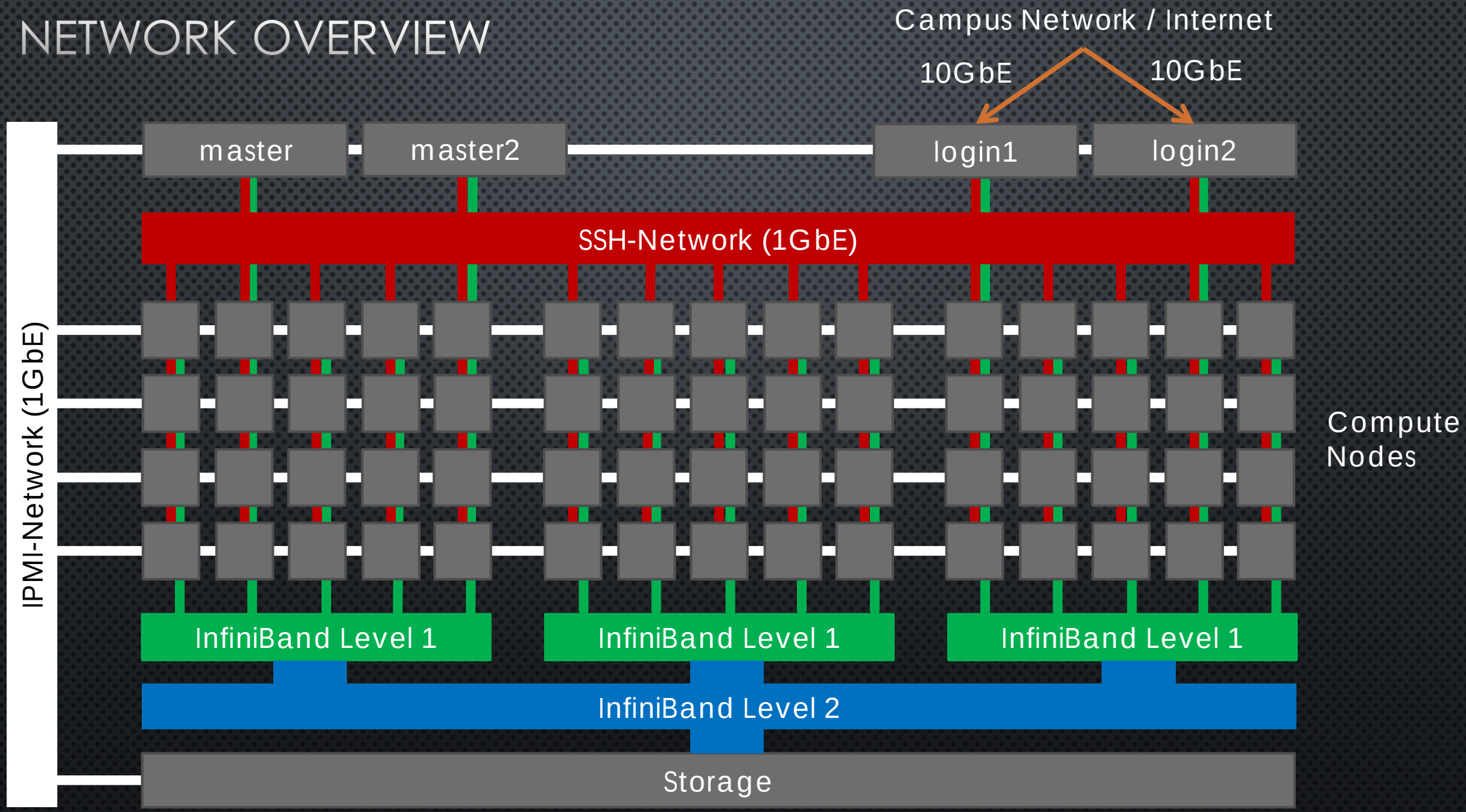
INFINIBAND FABRIC

- Low latency (0.7 microseconds), High Bandwidth (100 Gb/s)
- For parallel communication during computation
- Also used for Storage traffic

- 2 levels
  - 3x Core switches
  - 9x LEAF switches

- Nodes are connected to one leaf switch via copper cables
- All leaf switches are connected to all core switches via multiple fiber cables

NETWORK OVERVIEW

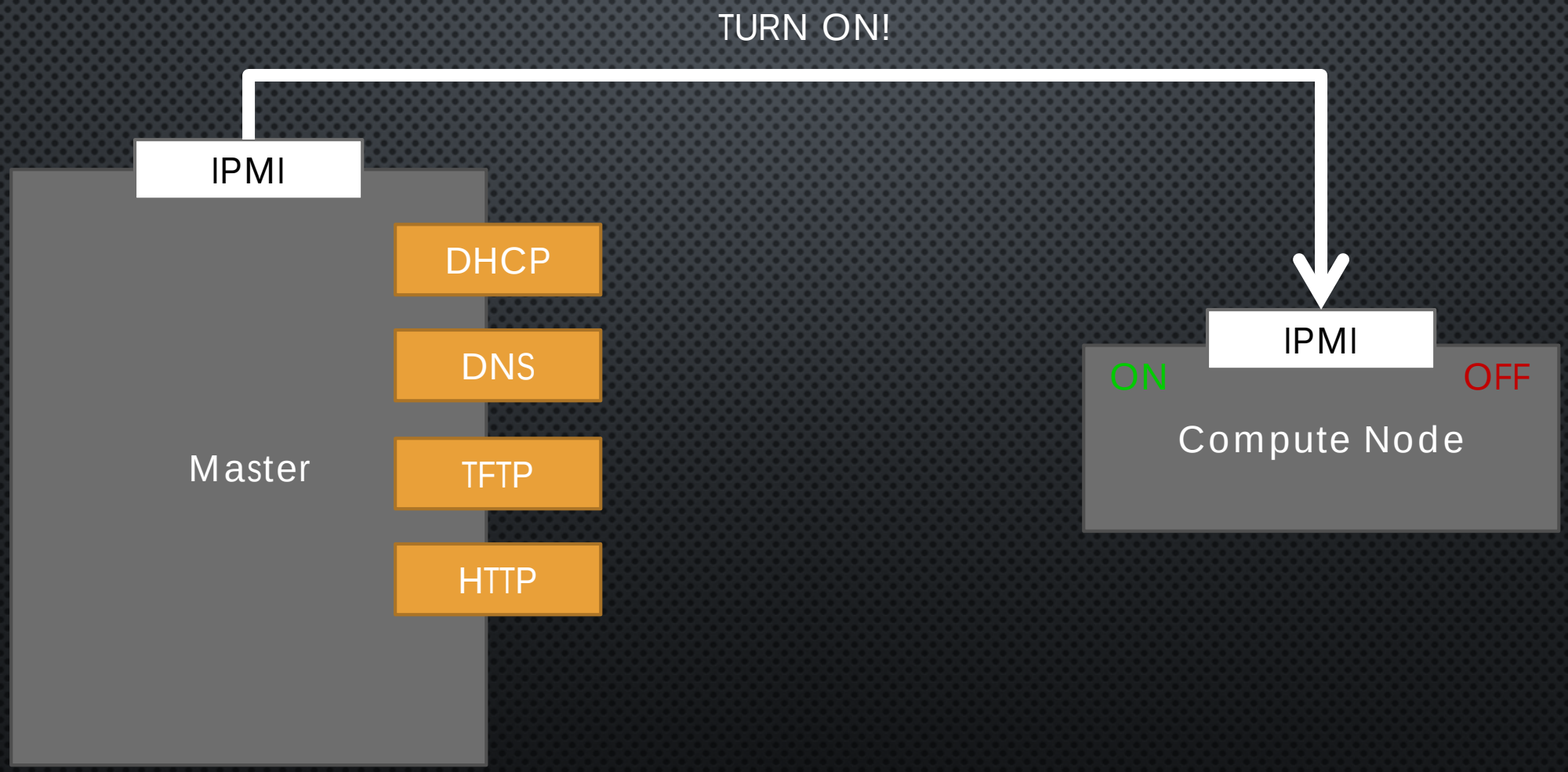


HARDWARE IS ONLY HALF OF THE STORY

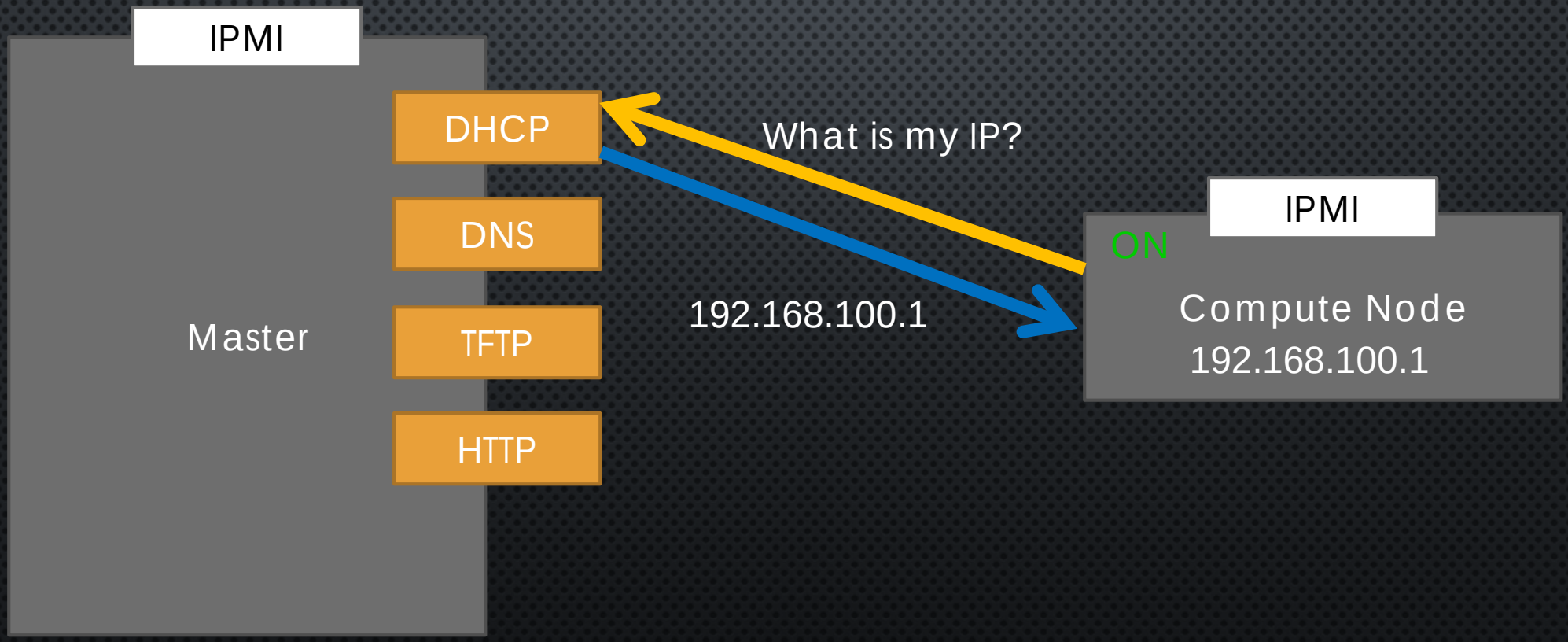
Provisioning - Install 200+ Servers

- Automated Installation
  - Kickstart Scripts for automating server installation
  - Used to set up filesystems and partitions on server
  - Install base operating system (e.g. CentOS)
  - Configure Network interfaces
  - Base packages (RPMs)
- Cobbler (http://cobbler.github.io/)
  - A tool to manage a group of machines and provision them using kickstart
  - Stores and configures daemons for DHCP, TFTP (for network boot), DNS and HTTPD
  - After exploring many alternatives, for us this has the best balance low level access and convenience

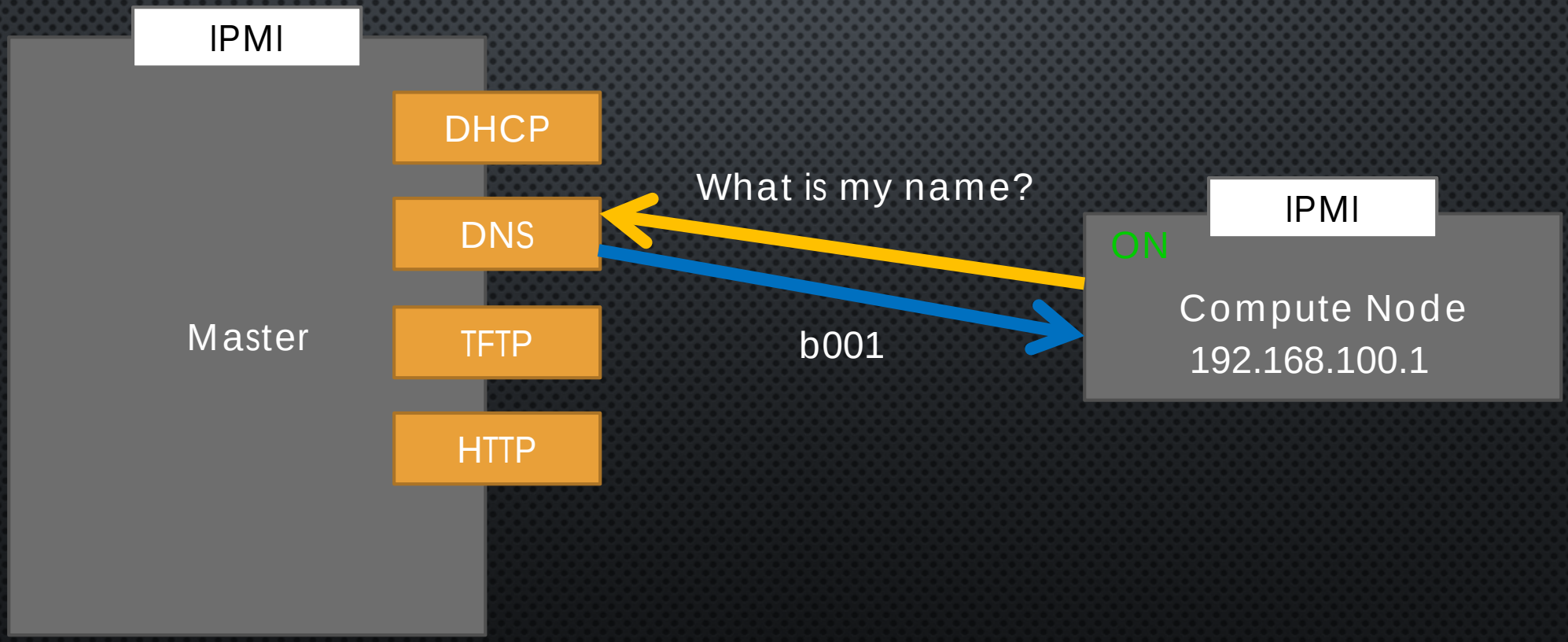
PROVISIONING



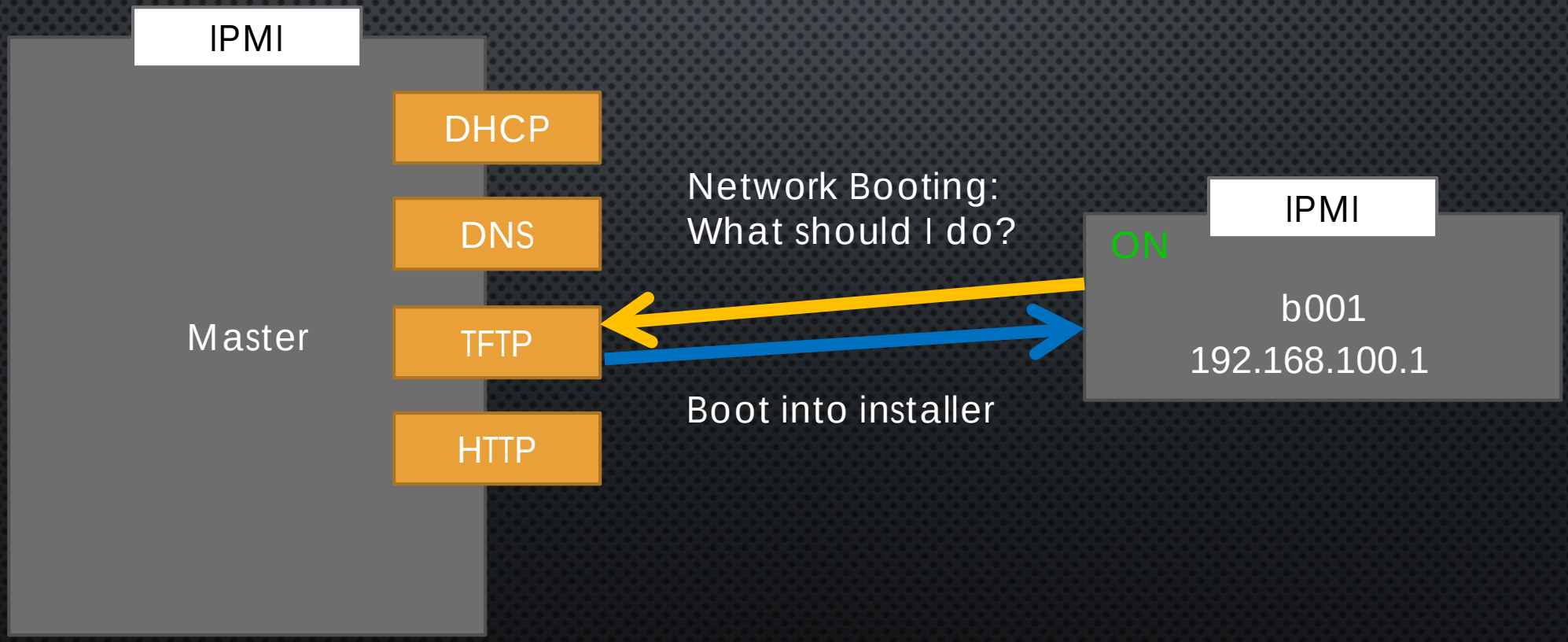
PROVISIONING



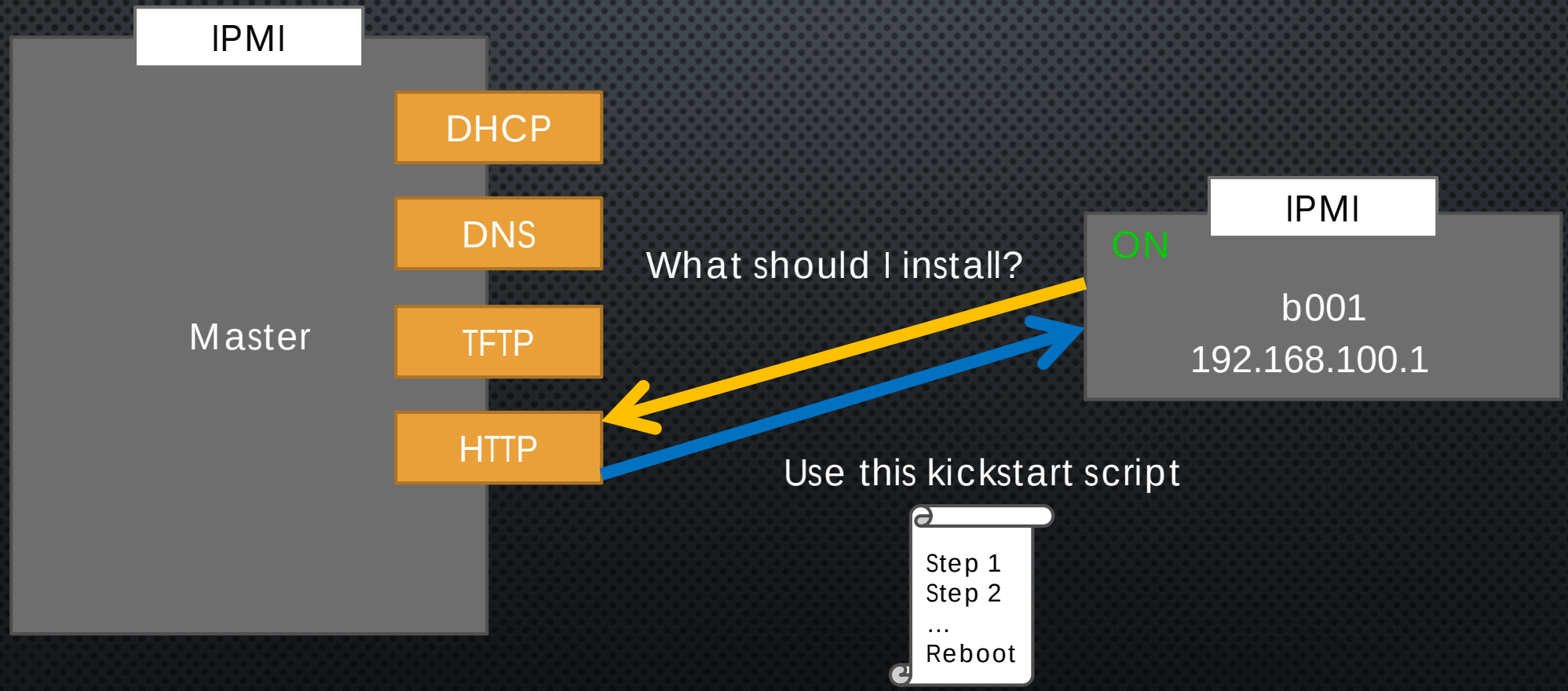
PROVISIONING



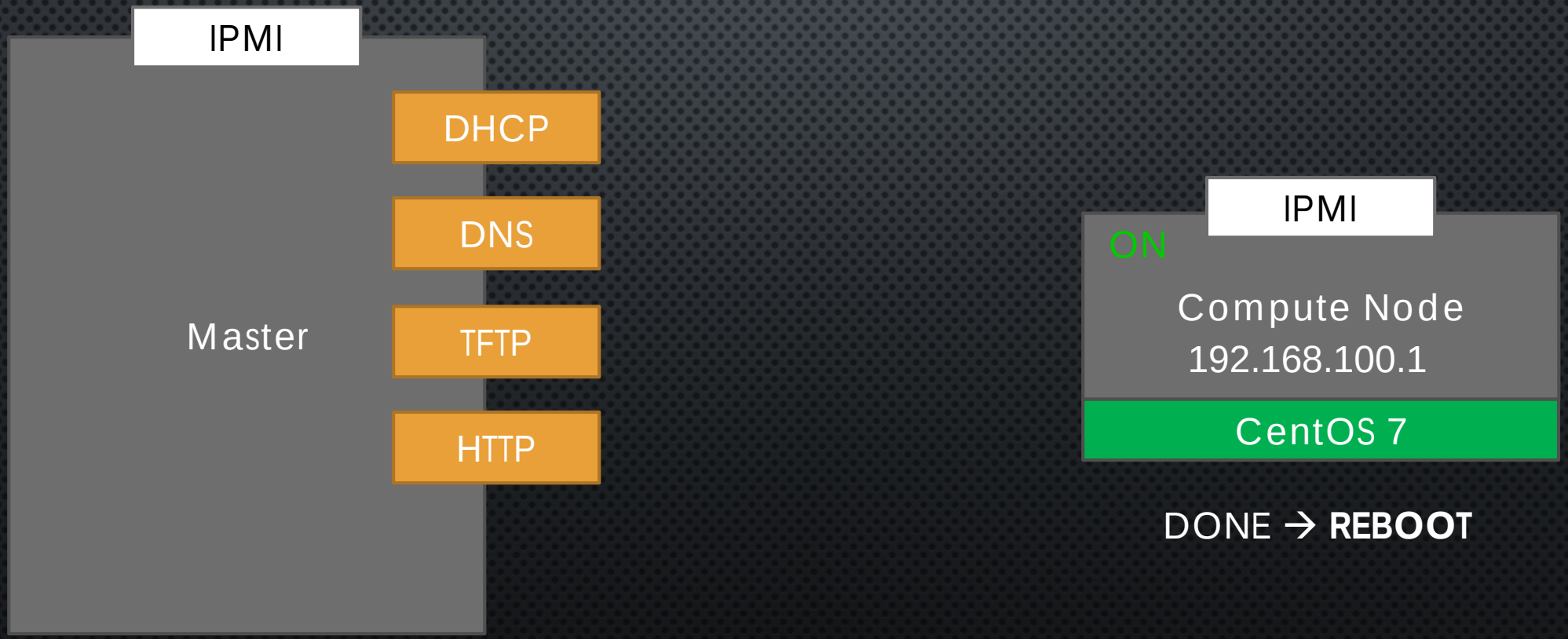
PROVISIONING



PROVISIONING

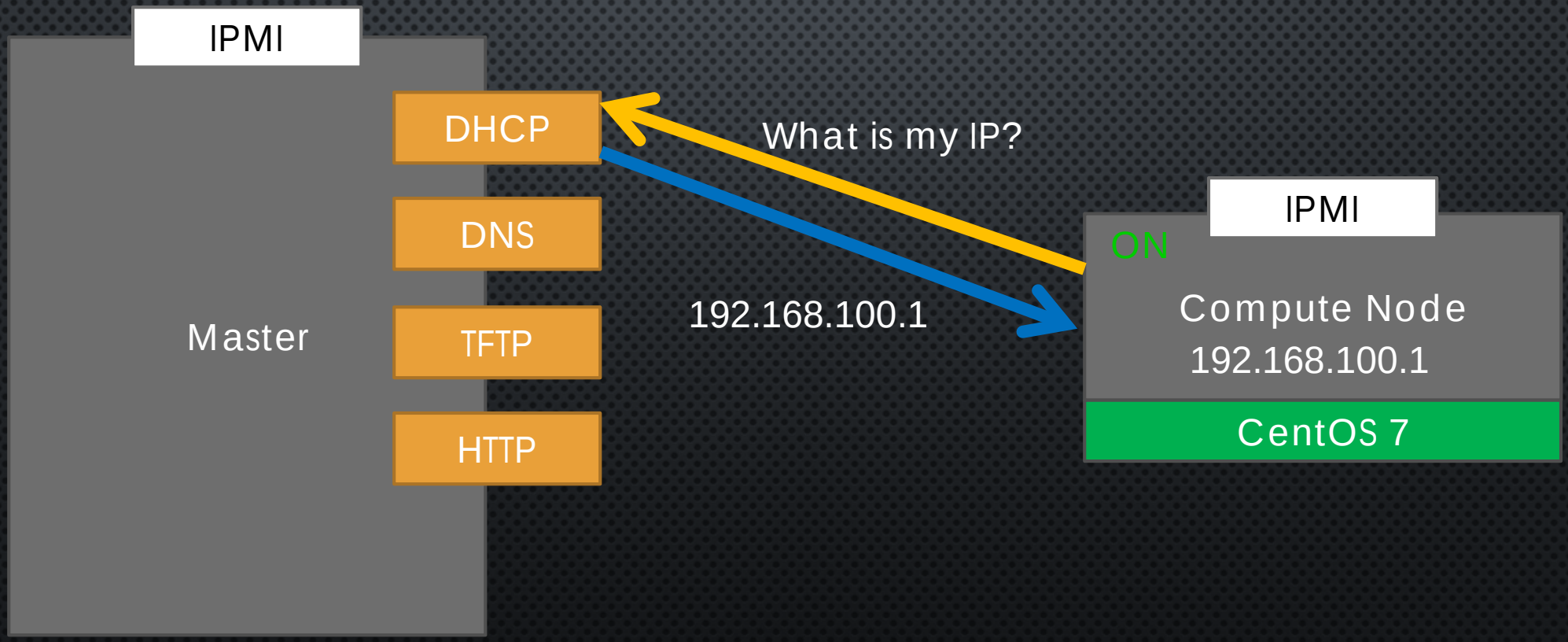


PROVISIONING

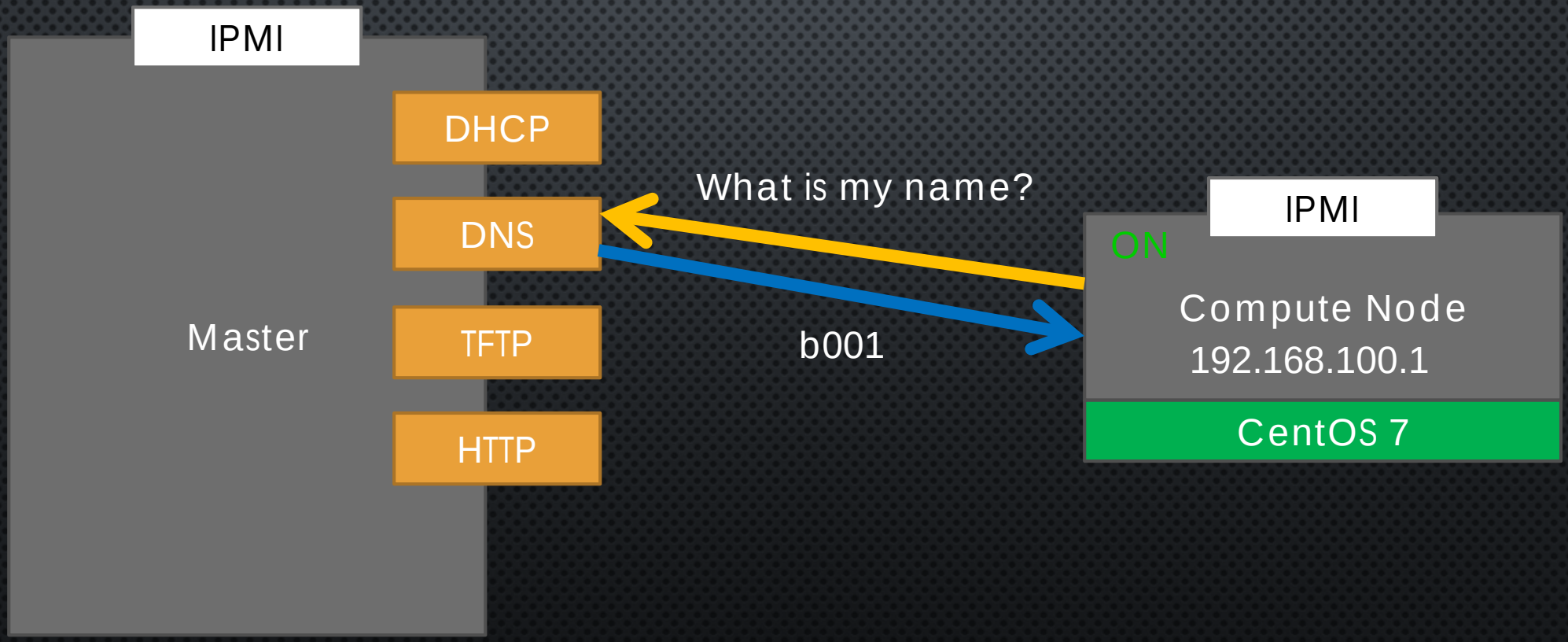


DONE → **REBOOT**

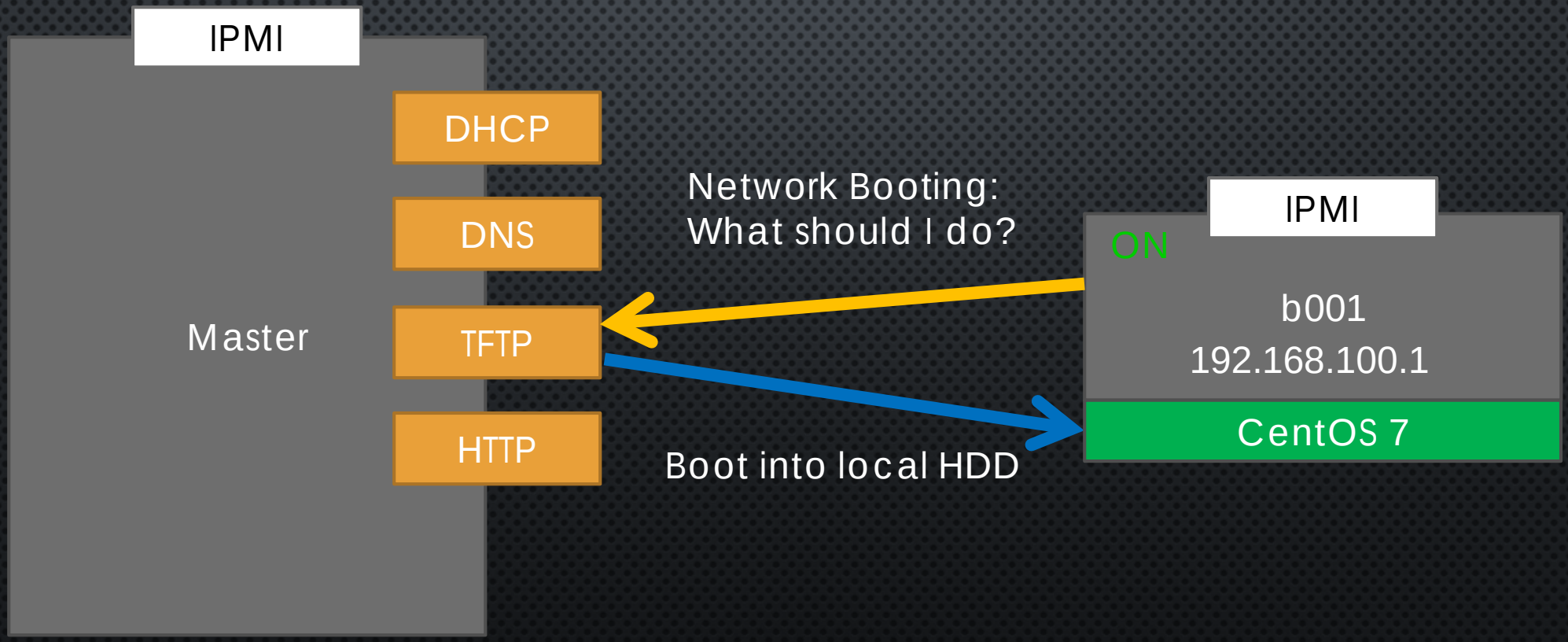
PROVISIONING



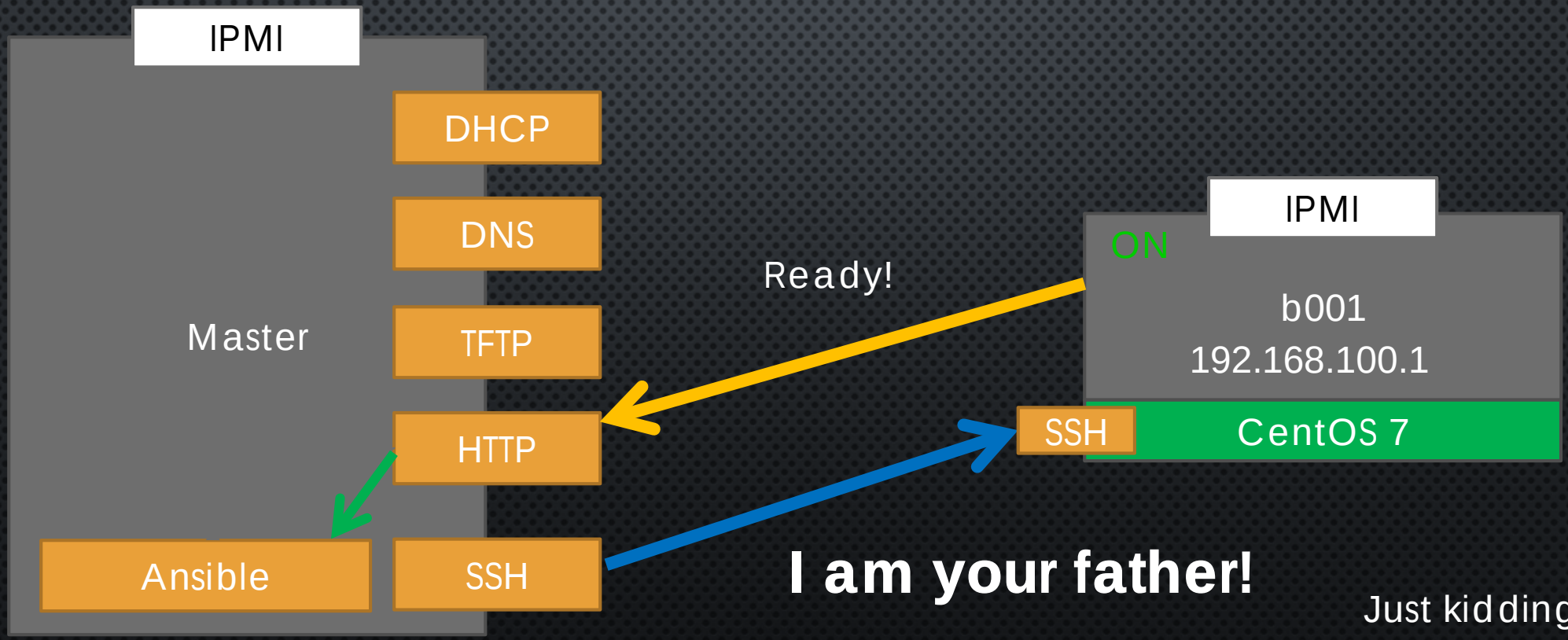
PROVISIONING



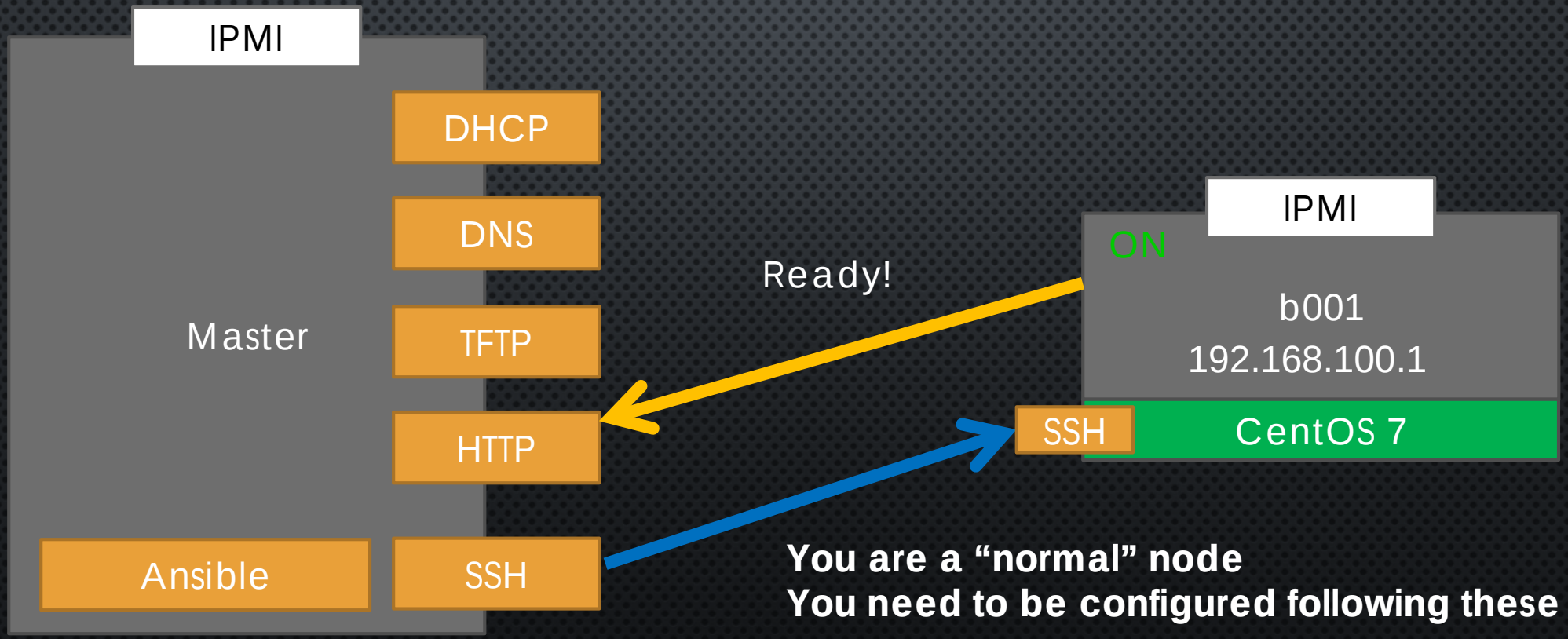
PROVISIONING



PROVISIONING



PROVISIONING



What happens during provisioning

- Each server boots into installer via network boot
- Based on selected profile, server downloads kickstart script
- HDD is wiped and new filesystems are created
- After completion of the installation → **reboot**
- Local boot into installed OS
- Server reports "I'm ready"
- Second stage of provisioning via configuration management (ansible)

ANSIBLE

- https://www.ansible.com/
- Tool for configuration management
- Language to write configuration changes as code
  - Install packages
  - Changes files
  - Control services
- Useful to enforce a consistent configuration among multiple servers
- You can define roles anD apply them to multiple servers using "playbooks"
- Alternative: Puppet

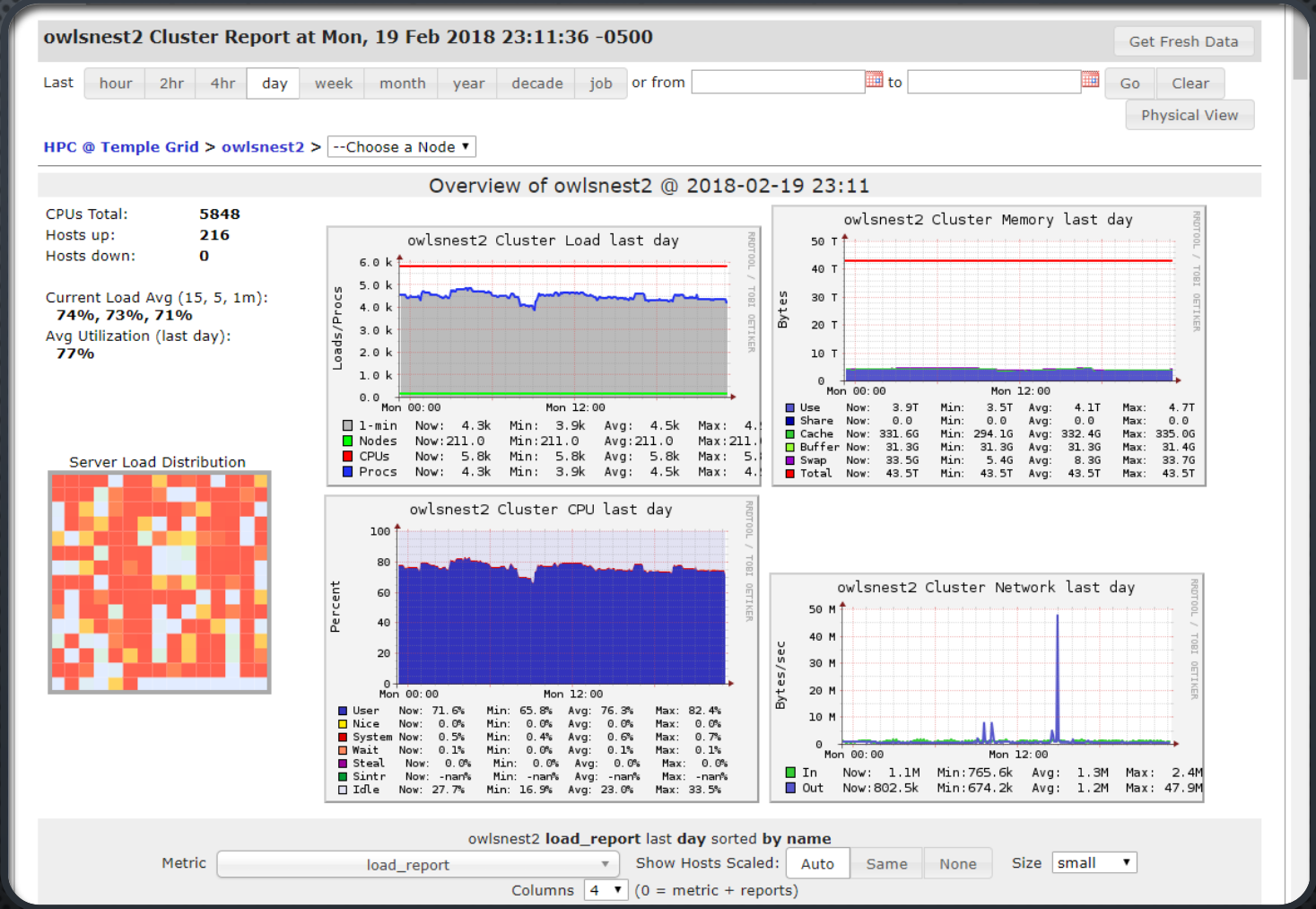
```
---
- name: Install ntpd daemon
  yum:
     name: ntp
     state: latest

- name: Disable default ntp
  lineinfile:
    dest: /etc/ntp.conf
    regexp: '^(server [0-3]\.centos\.pool\.ntp\.org iburst)'
    backrefs: yes
    line: '#\1'
    state: present

- name: Set up default default ntp server
  lineinfile:
    dest: /etc/ntp.conf
    regexp: '^server'
    insertafter: '^#server '
    line: 'server ntp.hpc prefer'

- name: Restart ntpd daemon
  service:
    name: ntpd
    enabled: yes
    state: restarted
```

When things go wrong

- Cluster nodes automatically detect if hardware has problems
- Health check scripts are run before and after a job
- → node is marked offline if there is a problem
- Admins can take action later, **but rest of the cluster remains operational**
- Thanks to automation, reimaging (reinstalling) a node is done within 10 minutes

MONITORING

- COLLECT INFORMATION ABOUT YOUR INFRASTRUCTURE!
- BENEFITS:
  - DETECT ANOMALIES
  - MAY HELP YOU DURING PROBLEM INVESTIGATIONS
- WE USE GANGLIA MONITORING SYSTEM
  - HTTP://GANGLIA.SOURCEFORGE.NET/
  - EVERY NODE HAS A GMOND DAEMON
    - CPU
    - MEMORY USAGE
    - GPU USAGE
    - NETWORK TRAFFIC
  - DATA IS COLLECTED ON MASTER
  - VISUALIZED ON EXTERNAL WEBSITE



WHY USE A BATCH SYSTEM?

→ **A BATCH SYSTEM ALLOWS YOU TO IMPLEMENT A USAGE POLICY**

→ MAXIMIZE UTILIZATION OF RESOURCES

→ GIVE USERS A **FAIR** AMOUNT OF COMPUTING RESOURCES

→ ENABLES USER MONITORING AND USAGE REPORTING

Batch system

We use: **PBS/Torque + Maui scheduler** (Alternative: SLURM)

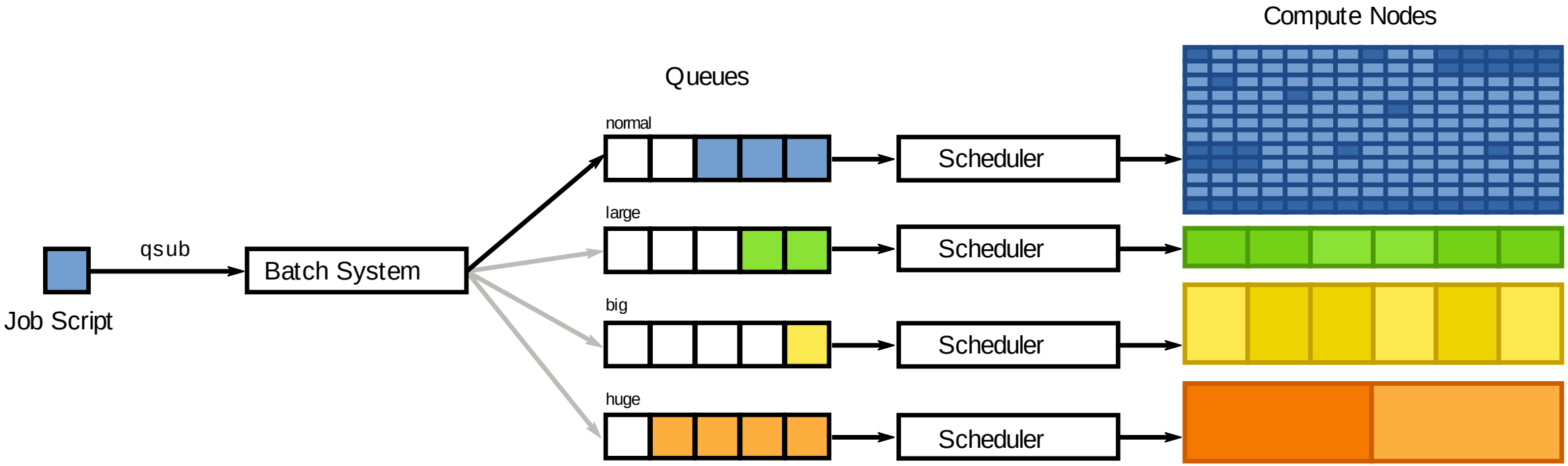
Nodes are configured in queues

- jobs can run on a list of compute nodes
- How many jobs may a user run at once in this queue?
- How many nodes can be allocated in one job?
- How many processors/how much ram are allowed per job?
- How long can jobs run?
- Etc.

We use queues to separate different hardware

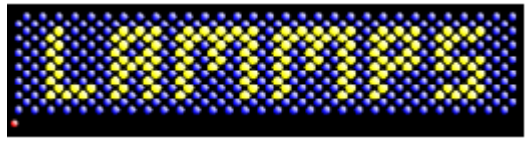
- Normal (28 core, 128GB RAM)
- Large (16 cores, 512GB RAM, 2x NVIDIA P100 GPU)
- Big (32 cores, 1500GB RAM)
- Huge (16 cores, 3000GB RAM)

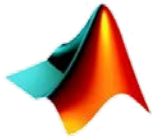
BATCH SYSTEM



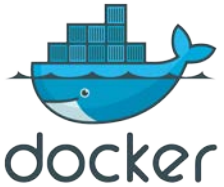
USER SOFTWARE

- Software is installed globally on shared storage and provided via modules
  - We use LMOD, which is compatible to original modules but has better dependency management
- We are exploring new capabilities to enable workflows (e.g., containers for machine learning and bioinformatics)
- We constantly work with users to improve cluster software over time

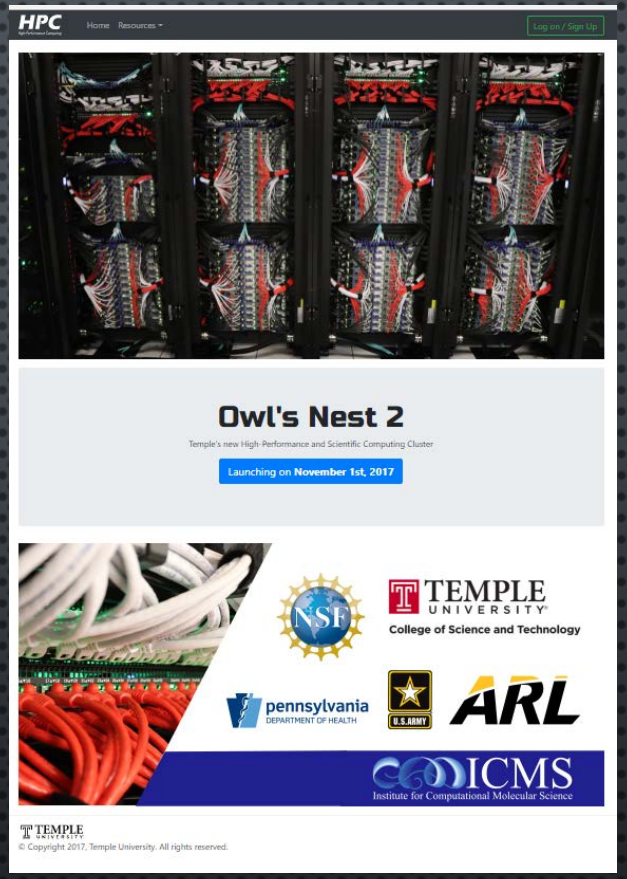



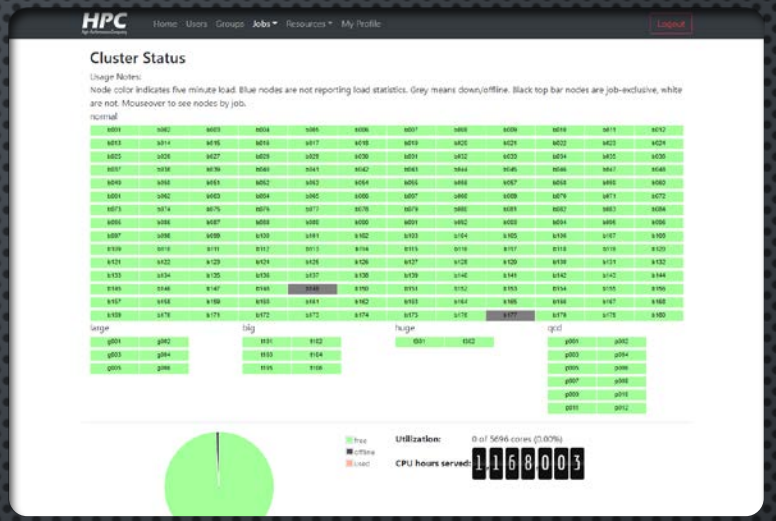




WEBSITE

- GATEWAY TO THE CLUSTER
- SIGN UP FOR CLUSTER ACCOUNT
- DOCUMENTATION OF THE CLUSTER
- MONITORING OF CLUSTER STATE AND USAGE
- USERS CAN KEEP TRACK OF THEIR JOBS
- DATA COLLECTION FOR REPORTING





Q & A