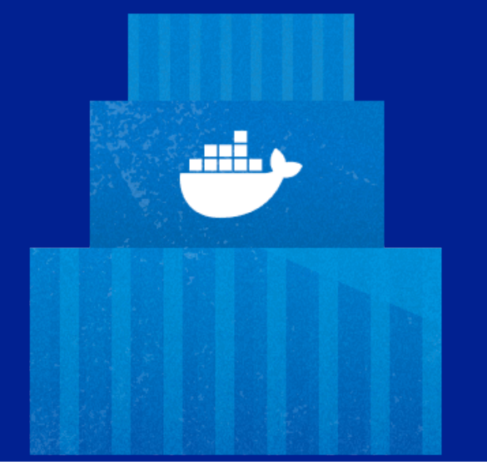
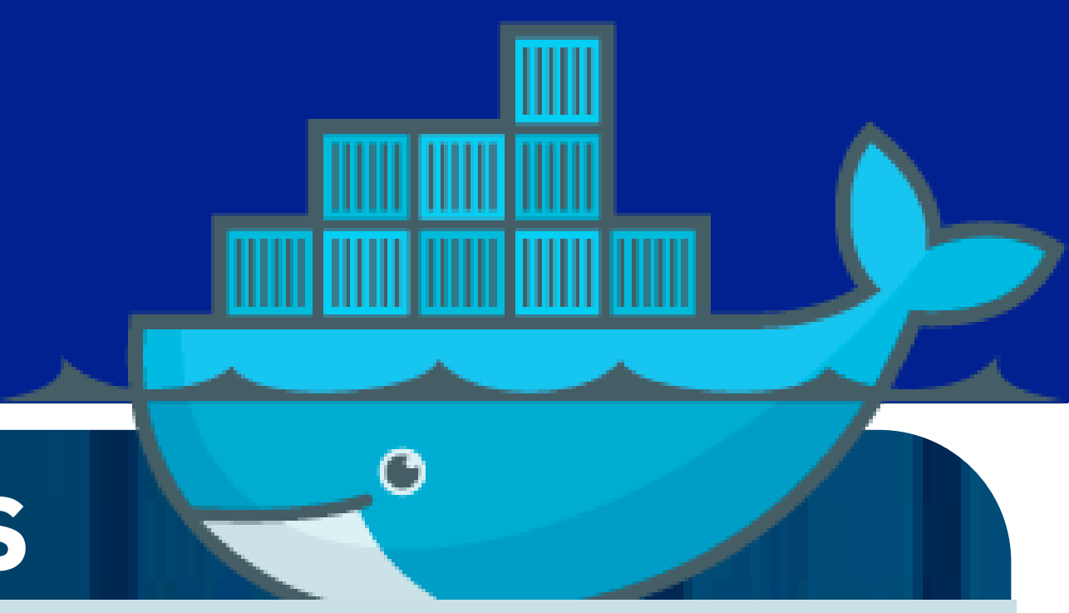
Cluster Management with Containerization on Switches

Robin Simpson, Anvitha Ramachandran, Dohyun Lee | HPC-DO
Mentors: Doug Egan, Alex Lovell-Troy, David Rich





Background

Utilizing Underutilized Resources

- Network switches are used for inter-node communication in HPC.
- Switches have underutilized resources for memory and processing
- Leverage with SONiC (Software for Open Networking in the Cloud) operating system

Integrating Containers with SONiC

- Deploying containers directly onto switches with SONiC.
- Executing various auxiliary tasks (e.g. metric logging, proxy download caching).
- Using switch capability rather than relying on running these services on a node.

Scenarios and Applications

- Streamline bootstrapping and configuration of new devices in the network with Cloud-Init
- Gather, monitor node performance metrics with Telegraf (e.g. memory utilization)
- Caching frequently accessed data closer to end users with external S3.
- Managing IPv4,6 address allocation and DNS resolution using dedicated VLAN.
- Automate client discovery on the network with Magellan and SNMP traps.

Methodology

- Scenarios validated and tested on both physical and virtual switches
- This ensures that our scenarios function as intended and uniformly.

Specifications

- Mellanox SN2100 / Arista DCS-7050QX-32
- IB MT27800/MT28908
- 9 Compute Nodes 1 Head Node Cluster Intel 6438Y+
- QEMU KVM 16 GB Memory Quad Core

Challenges

- **Network Configurations:**
  - Misconfigured virtual network brought down cluster nodes and switches.
- **Finding Ideal Software and Solutions:**
  - Proxy Caching meant to be done with Versity, had to use alternative S3FS.
- **Knowledge Stat Check:**
  - No prior knowledge on topics and scenarios. Lots of trial and error.
- **Overlap and Conflicts:**
  - All Scenarios run on the same nodes and switches. They affect each other by making some services not work on certain nodes and complicating configuration.

References

[1] D. Bautista, T. Bautista, “Supercomputer Institute Guide,” Los Alamos National Laboratory

[2] Linux Foundation, ”Software for Open Networking in the Cloud,” SONiC OS, https://sonicfoundation.dev

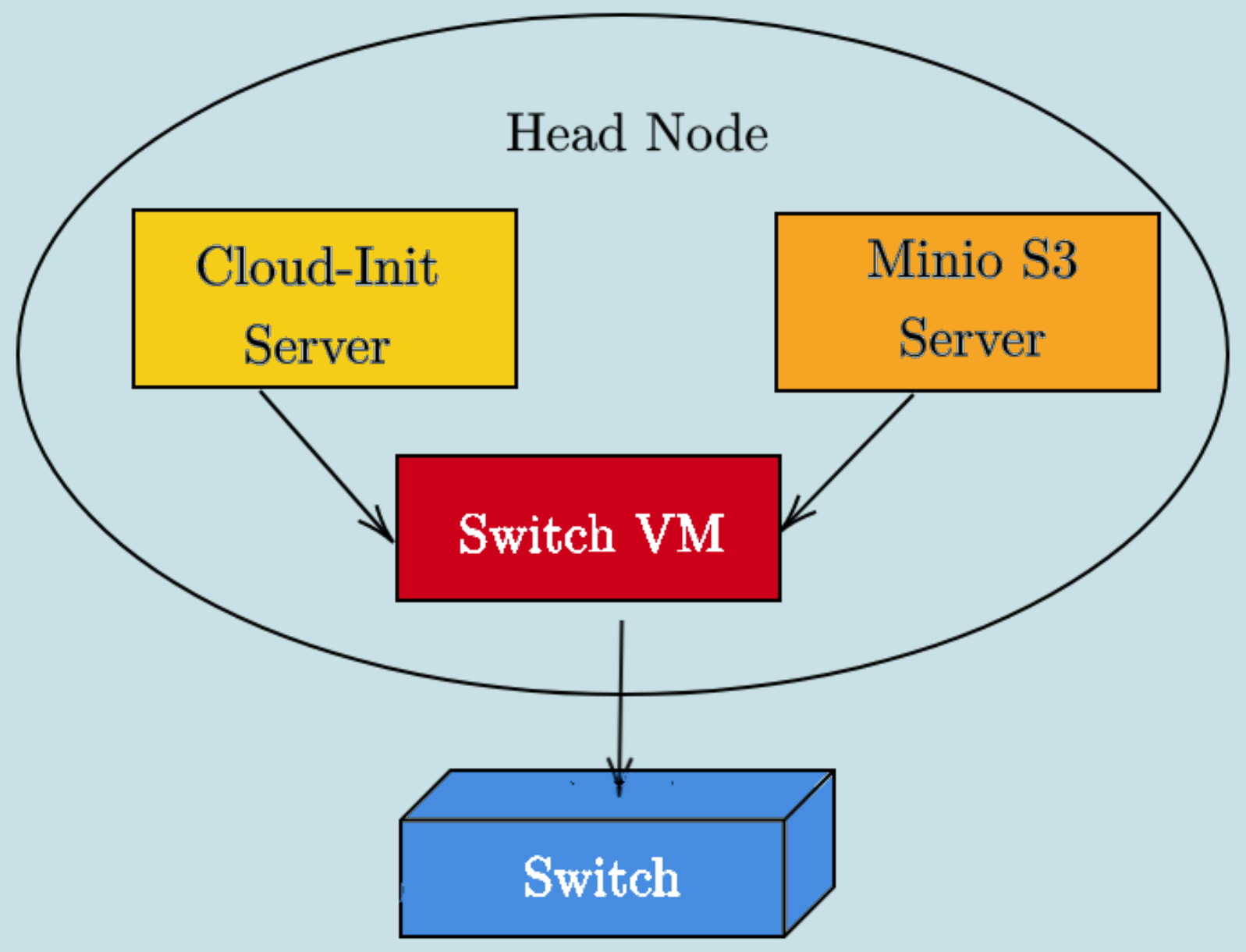
[3] “The standard for customising cloud instances,” cloud-init.io, https://cloud-init.io/

[4] “FUSE-based file system backed by Amazon S3,” S3FS, https://github.com/s3fs-fuse

[5] InfluxData, ”Telegraf Open Source Server Agent”, InfluxDB. https://www.influxdata.com/

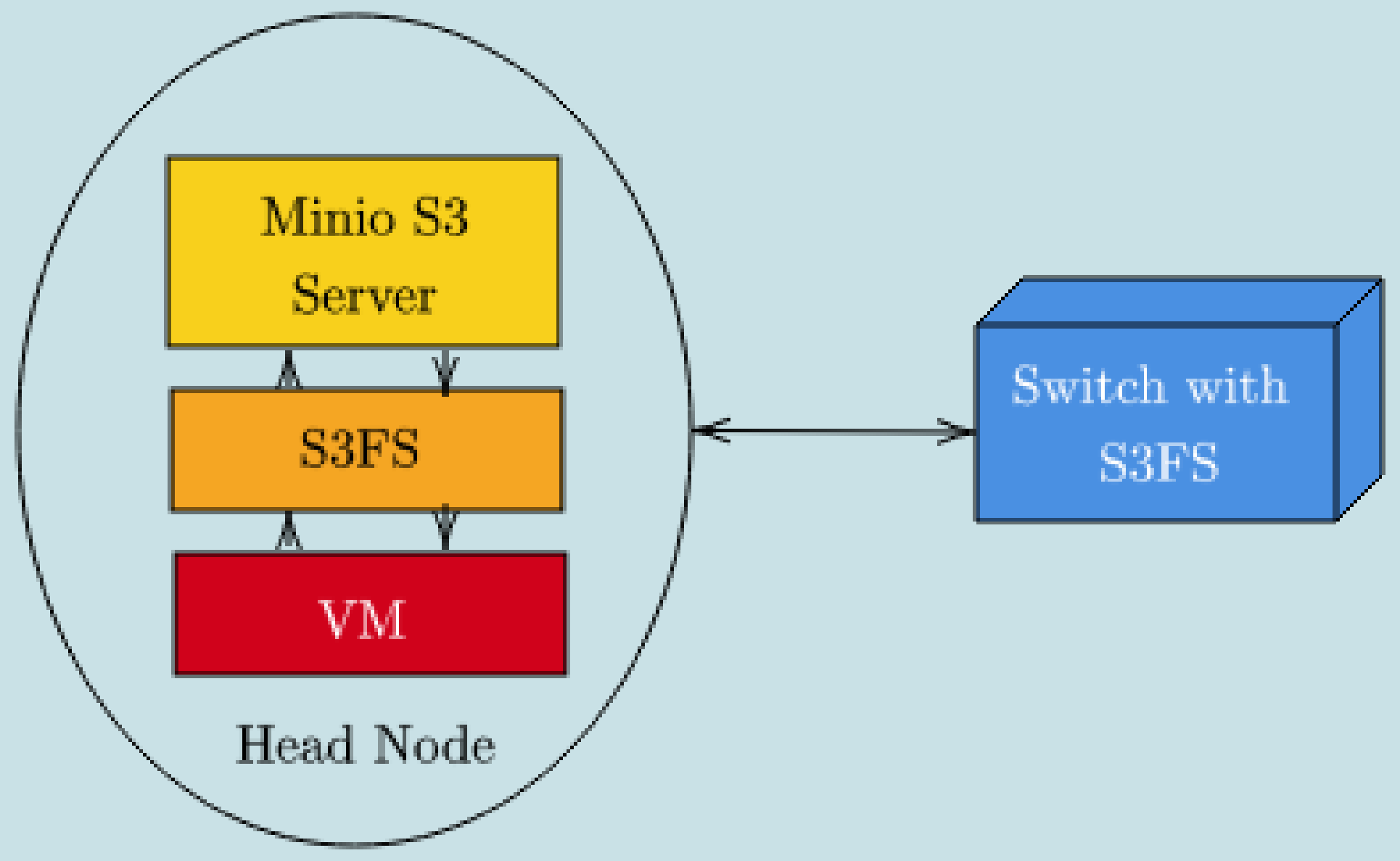
[6] Los Alamos National Laboratory, ”HPC System Management,” OpenCHAMI. https://www.ochami.org/

Cloud-Init Services



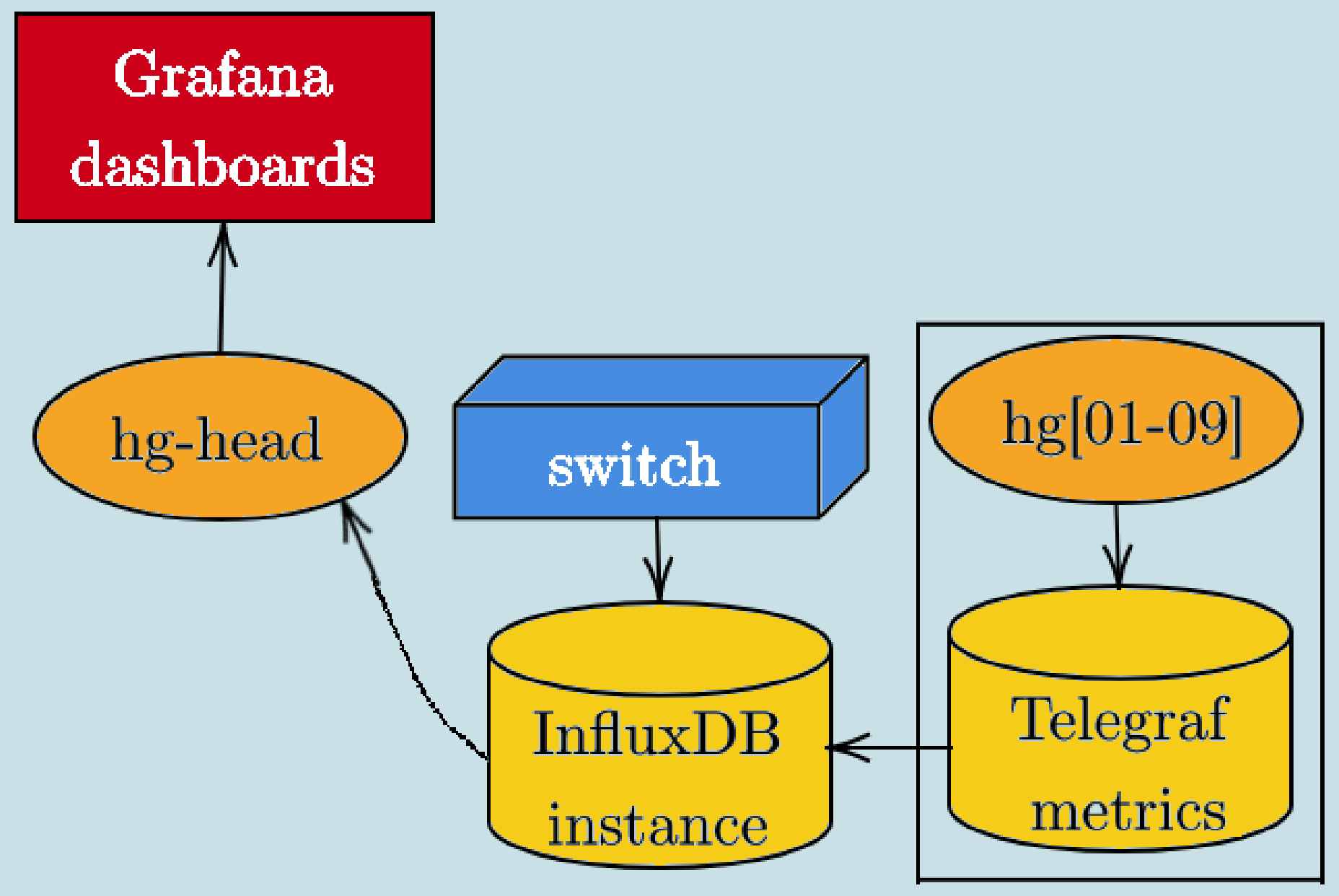
- Run cloud-init services on a SONiC switch.
- Use MinIO S3 as the external storage for config payload.
- MinIO and Cloud-Init are Docker containers on the head node linked by virtual nets.

Proxy Caching



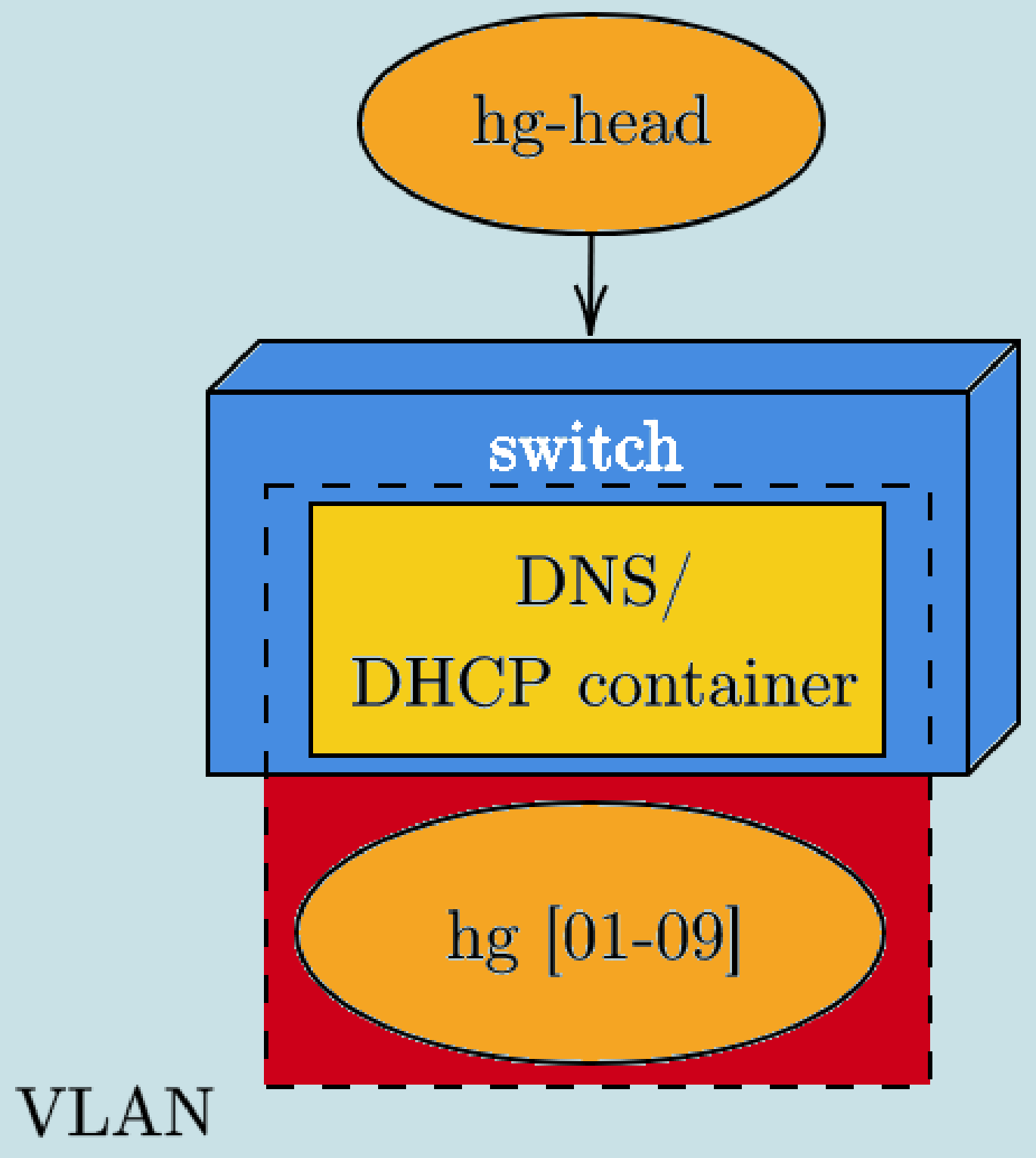
- Configure s3fs with necessary S3 bucket credentials.
- Mount the S3 to a local directory on the switch.
- Enable the switch to cache files locally.

Telegraf Collection



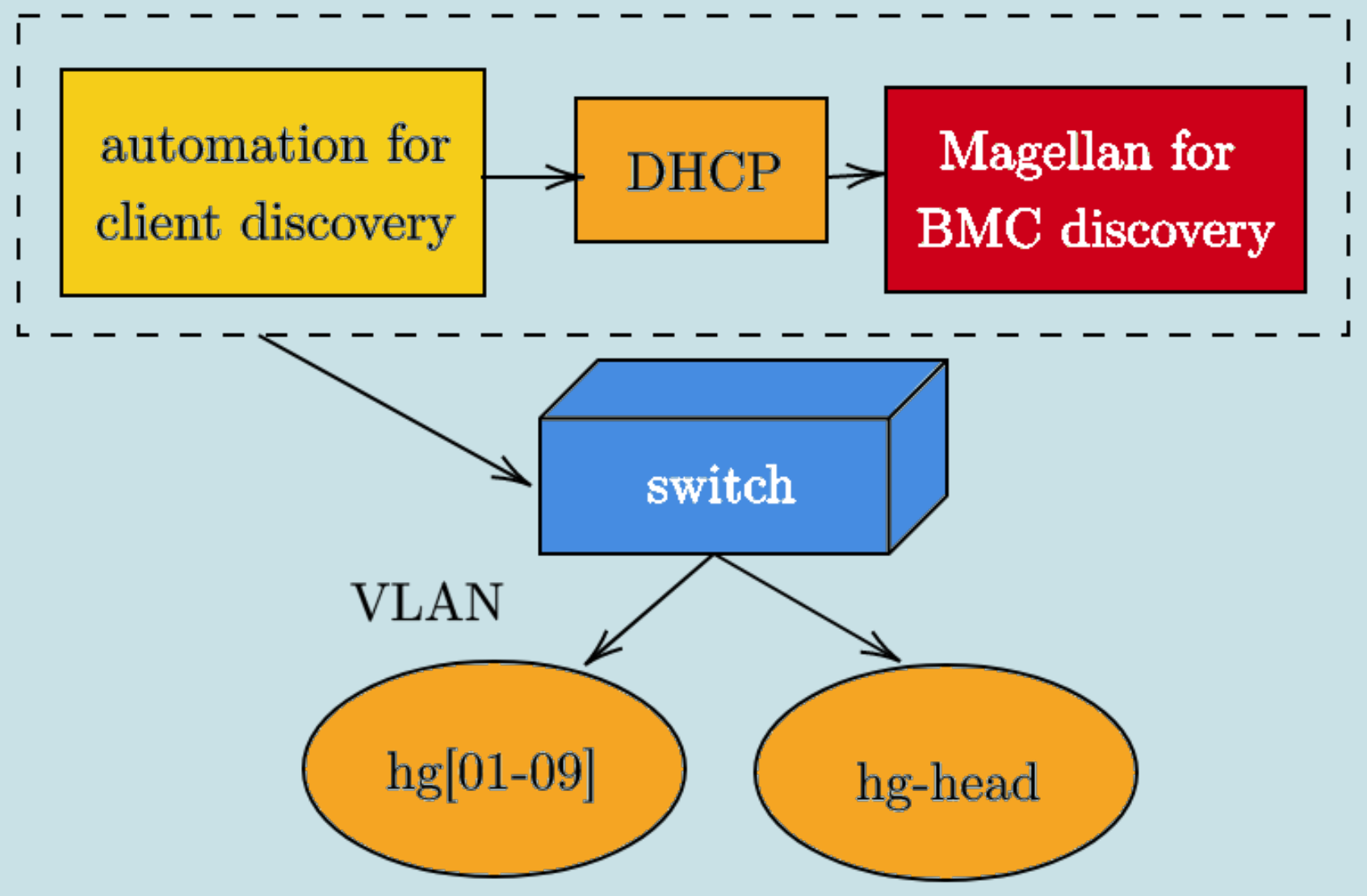
- Collect node metrics (e.g., CPU usage, disk I/O).
- Use containers to aggregate and display metrics.
- Create visualizations of nodes' health.

IPV6 Provider



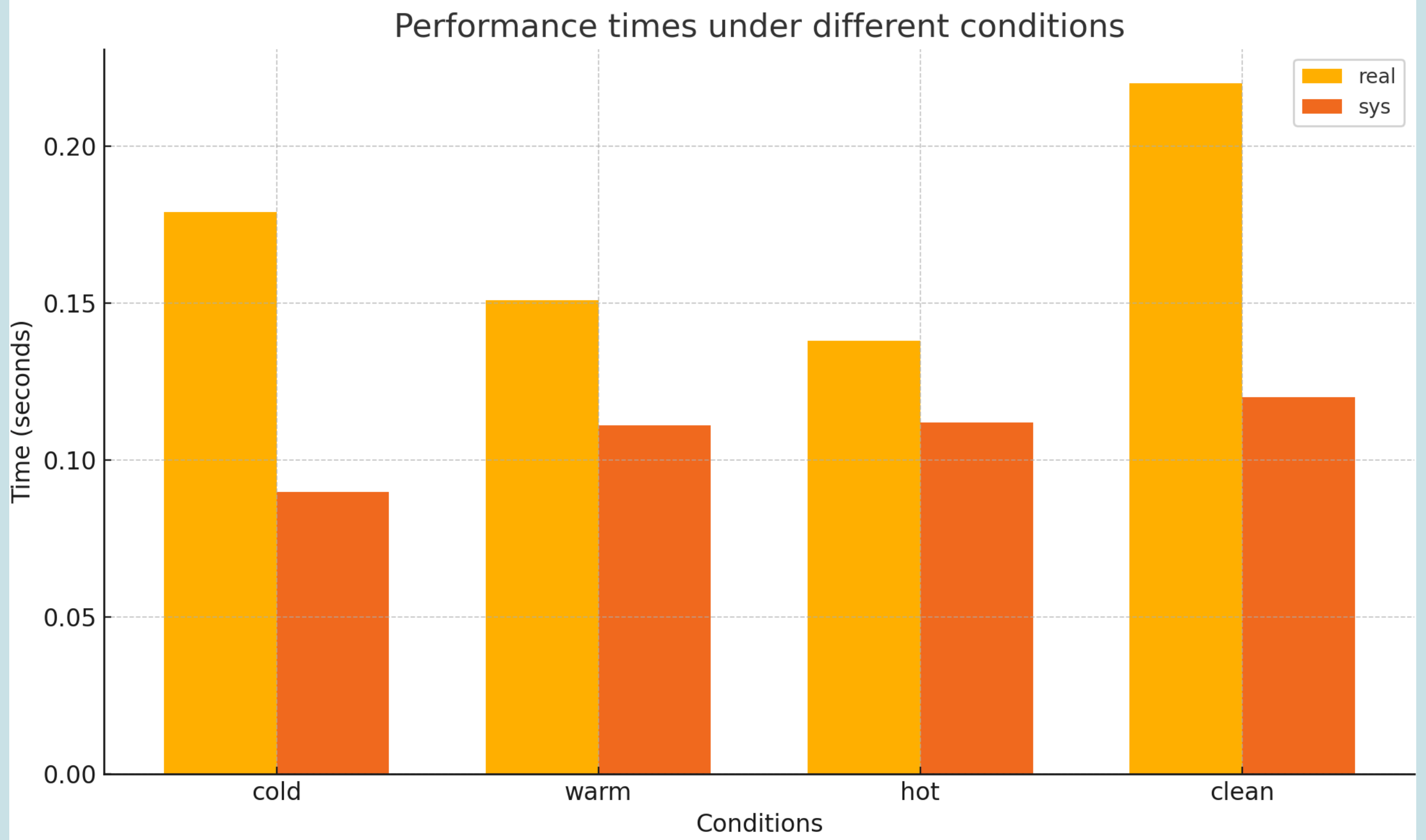
- Enable automatic dynamic IP assignment to nodes.
- Utilize IPv6 to expand the number of nodes.
- Automate configuration without a switch restart.

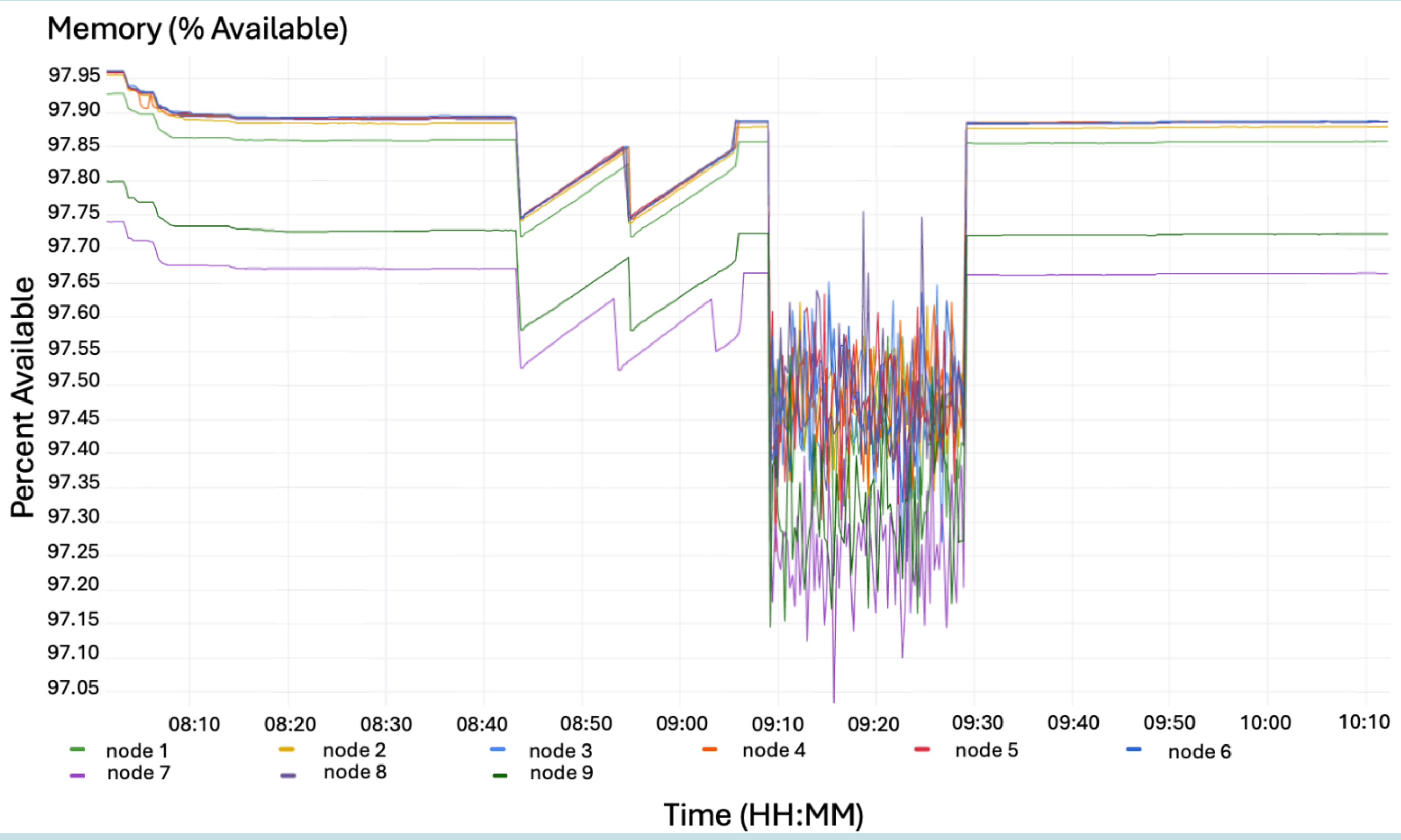
Magellan Discovery

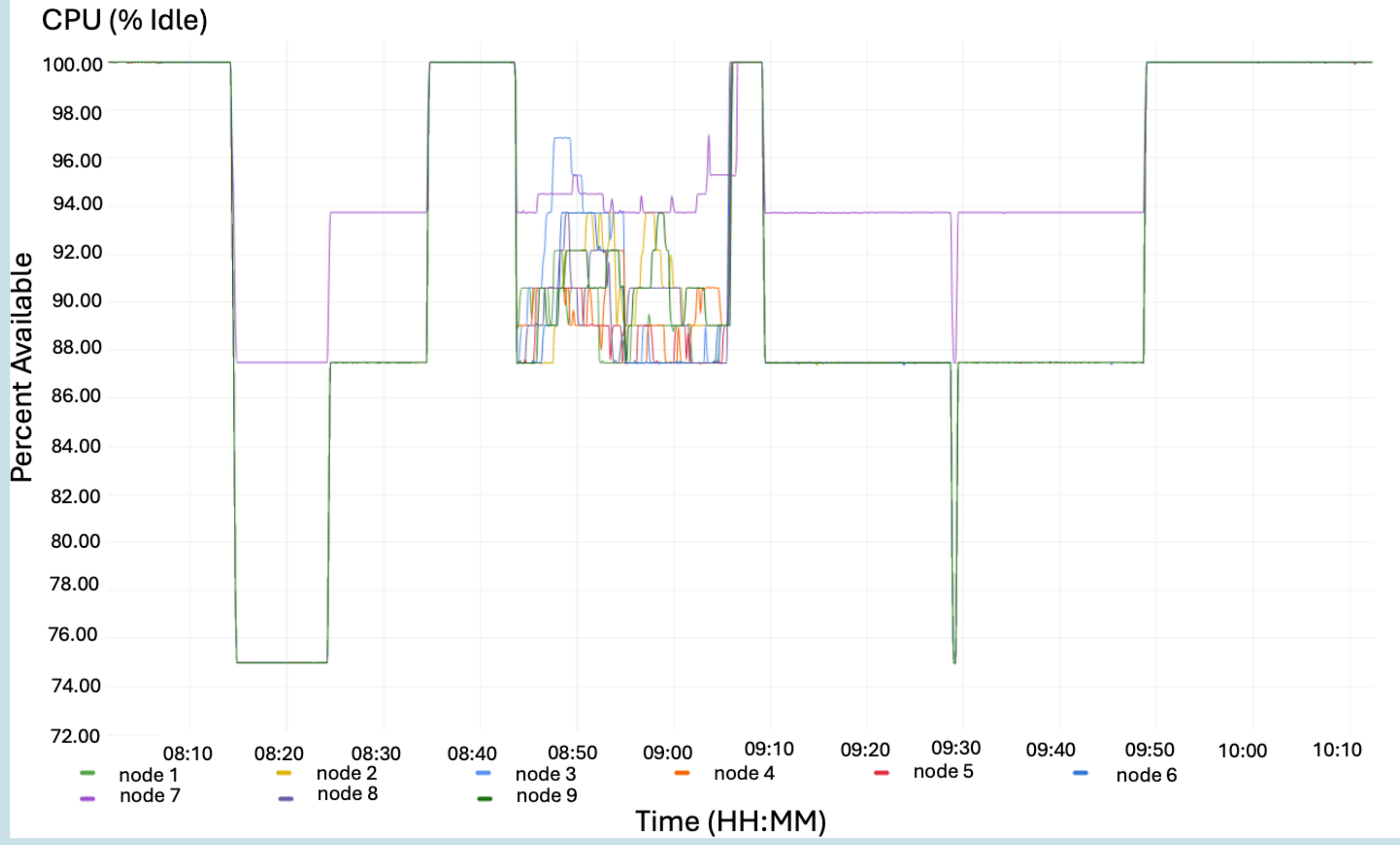


- Create and deploy a container or automated system to manage new client devices on a VLAN.
- Assign IP addresses and initiate the "discovery" process for new clients using Magellan.
- Use Magellan to map the network topology, identifying and monitoring new devices.
- Streamline integration of new clients, ensuring efficient IP allocation and network management.

Results







Future Work

Explore Additional Scenarios and Use Cases:
Security Services, Load Balancing, Network Management Tools

Improve Efficiency:
Explore more ‘intensive’ processes to test the limits of Network Switches

Try on Newer Hardware:
Mellanox and Arista Switches used are 10+ years old; compare performance

Los Alamos NATIONAL LABORATORY NNSA National Nuclear Security Administration

Managed by Triad National Security, LLC, for the U.S. Department of Energy's NNSA.

LA-UR-24-28076

ILLINOIS USC CAL POLY