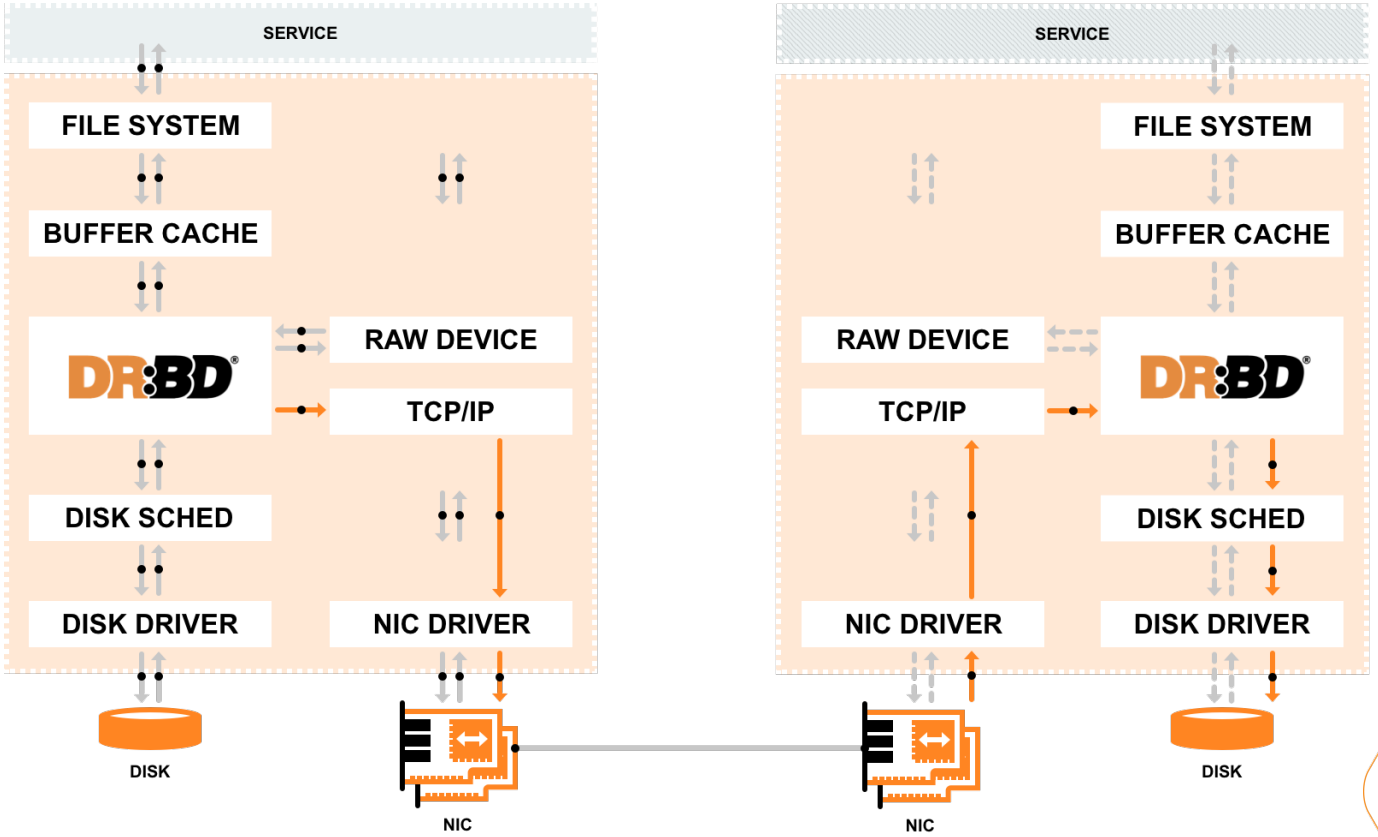
Majority in DRBD
and the LINBIT CloudStack HCI Appliance

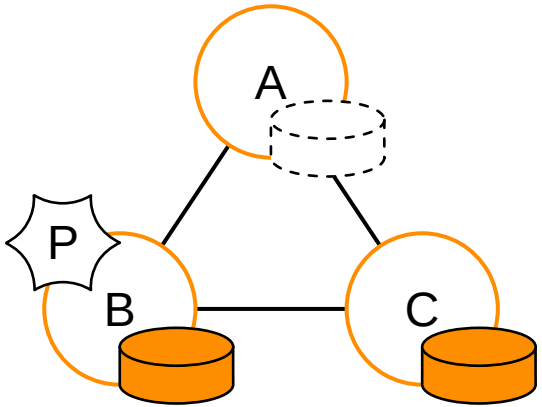
Philipp Reisner, CEO LINBIT

Protecting Data by replication



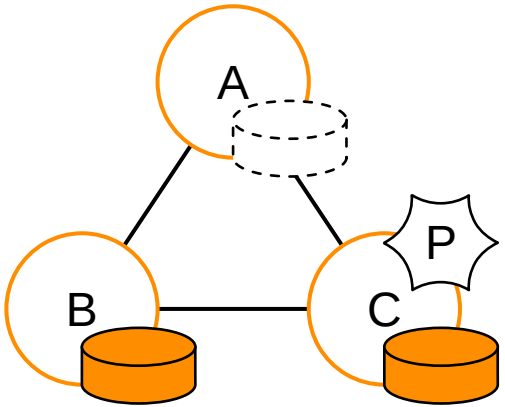
Quorum – Basics

- Three nodes minimum
- Diskless nodes act as tiebreakers
- Node B is in primary role and runs the application



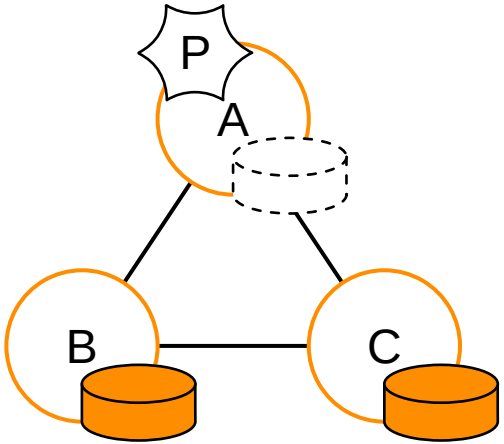
Quorum – Primary Movable

- Primary role and application are freely movable



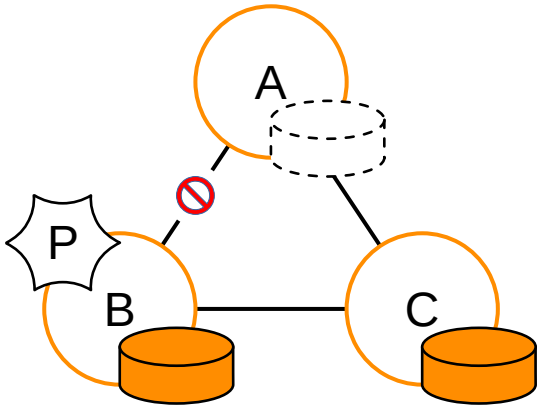
Quorum – Diskless nodes

- When it runs on a diskless node, it reads in a load balancing fashion from the nodes with backing storage
- The diskless nodes is also called the tiebreaker or the witness node



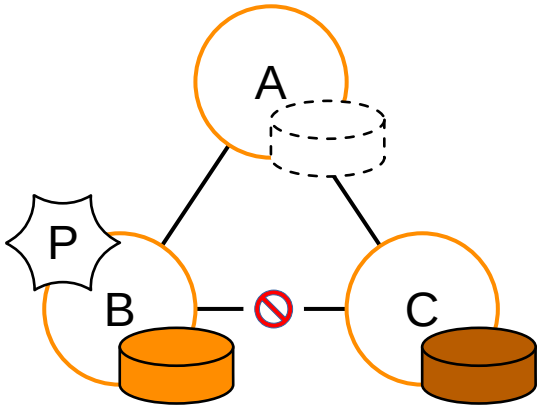
Network partitioning

- Network partitioning between A and B has no impact



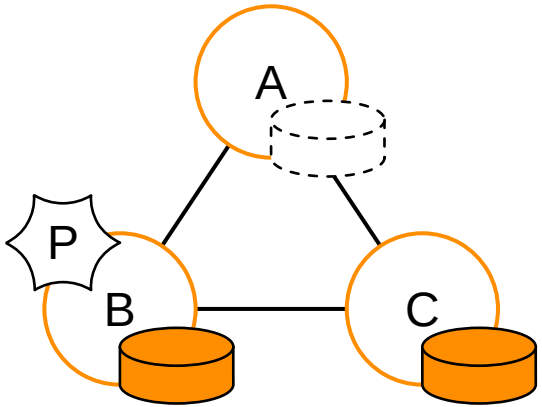
Network partitioning

- Network partitioning between A and C leads to C's replica becoming Outdated



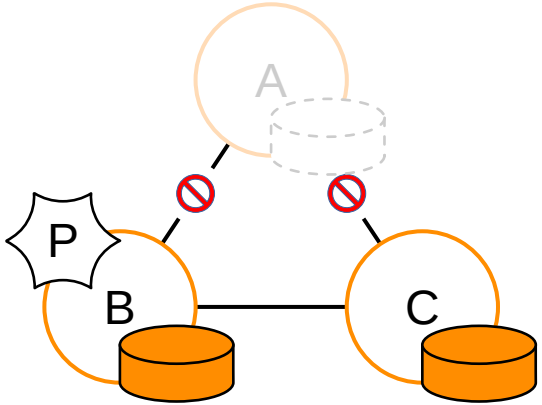
Network partitioning

- DRBD resyncs C upon reconnect
- Only blocks touched in the meantime on the primary



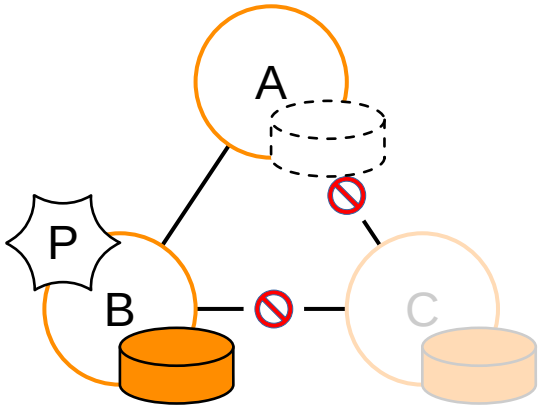
Network partitioning

- Isolate A
- A loses quorum – can no longer promote – it also has no access to data
- B & C keep quorum



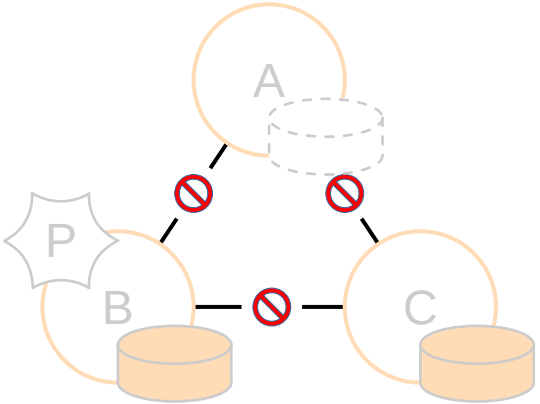
Network partitioning

- Isolate C
- C loses quorum – can no longer promote
- B & A keep quorum



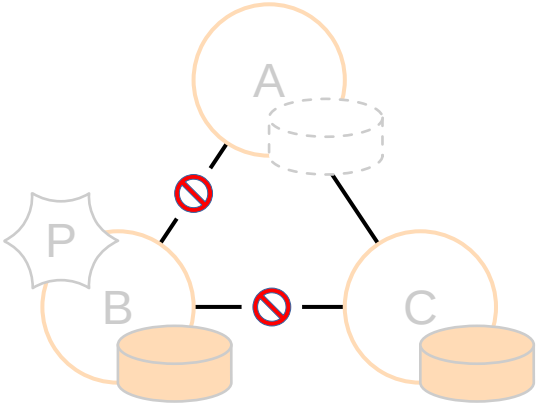
Diskless switching partition

- A switches to C, step 1
- All nodes without quorum
- Primary B, freezes I/O or completes I/O requests with errors, depends on config



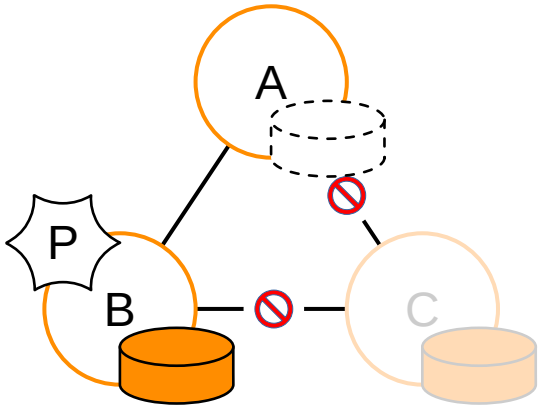
Diskless switching partition

- A switches to C, step 2
- C can not regain quorum by connecting to the tiebreaker node(s)



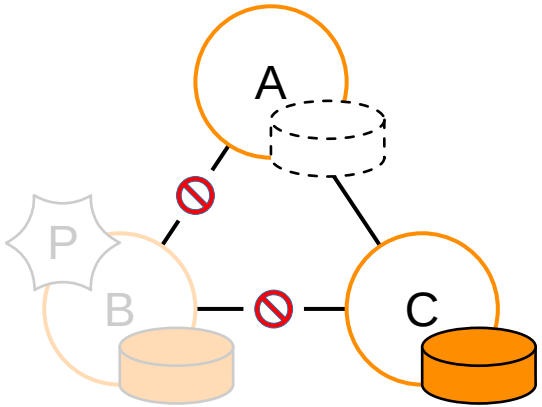
Diskless not switching partition

- A restores quorum in the partition A & B
- I/O on B resumes
- Application on B thaws
- The partition preserves quorum over a reboot of B



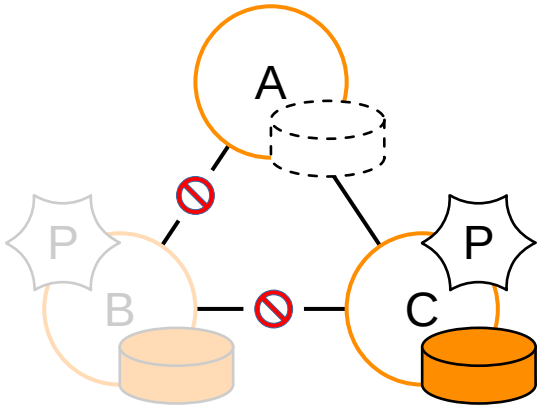
Network partitioning

- B gets isolated
- A & C are quorate
- B freezes I/O and the application
- A & C advertise a positive promotion_score



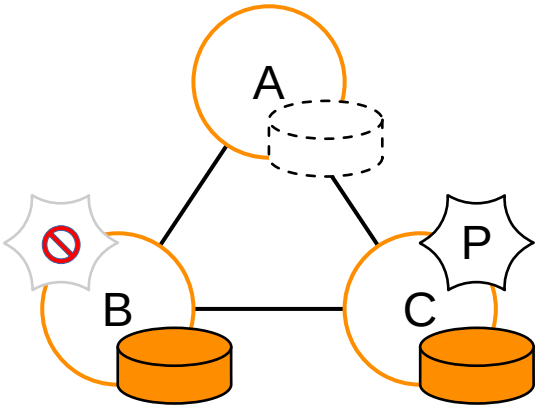
Network partitioning

- A & C are quorate
- A clustermanager promotes C to primary and starts the application
- B does not know, and stays frozen



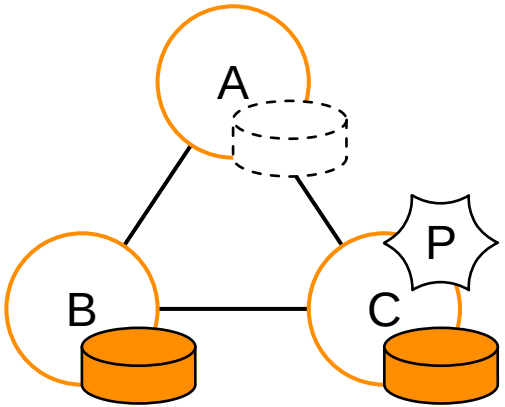
Network partitioning

- B can recover by reconnecting with the other partition
- Or by `drbdadm –force secondary`
- The application gets I/O errors and needs to close the DRBD device (terminate)



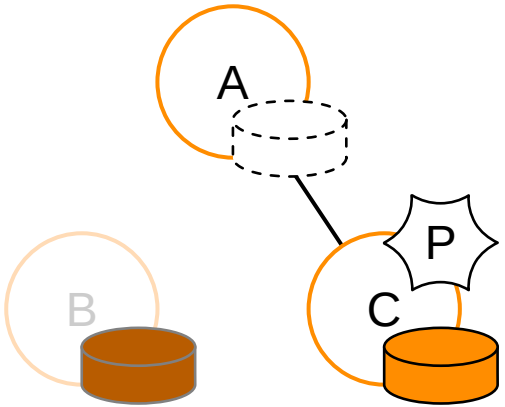
Network partitioning

- After the application terminated DRBD clears the force-io-errors flag
- Resync from C to B
- Recovery completed



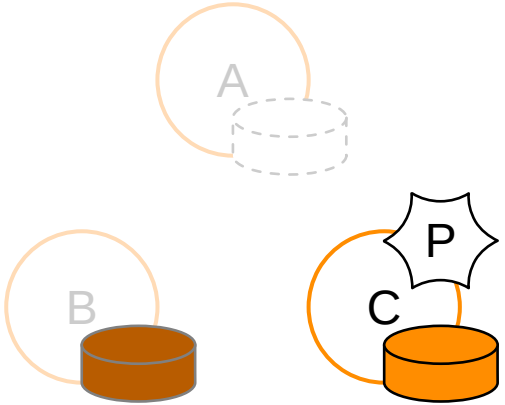
Gracefull disconnect

- B gracefully disconnects
- B outdates its disk when it leaves a partition with a primary node
- A graceful shutdown behaves in the same way



Gracefull disconnect

- C knows that B is Outdated and A is diskless
- It keeps quorum although it is just one out of three



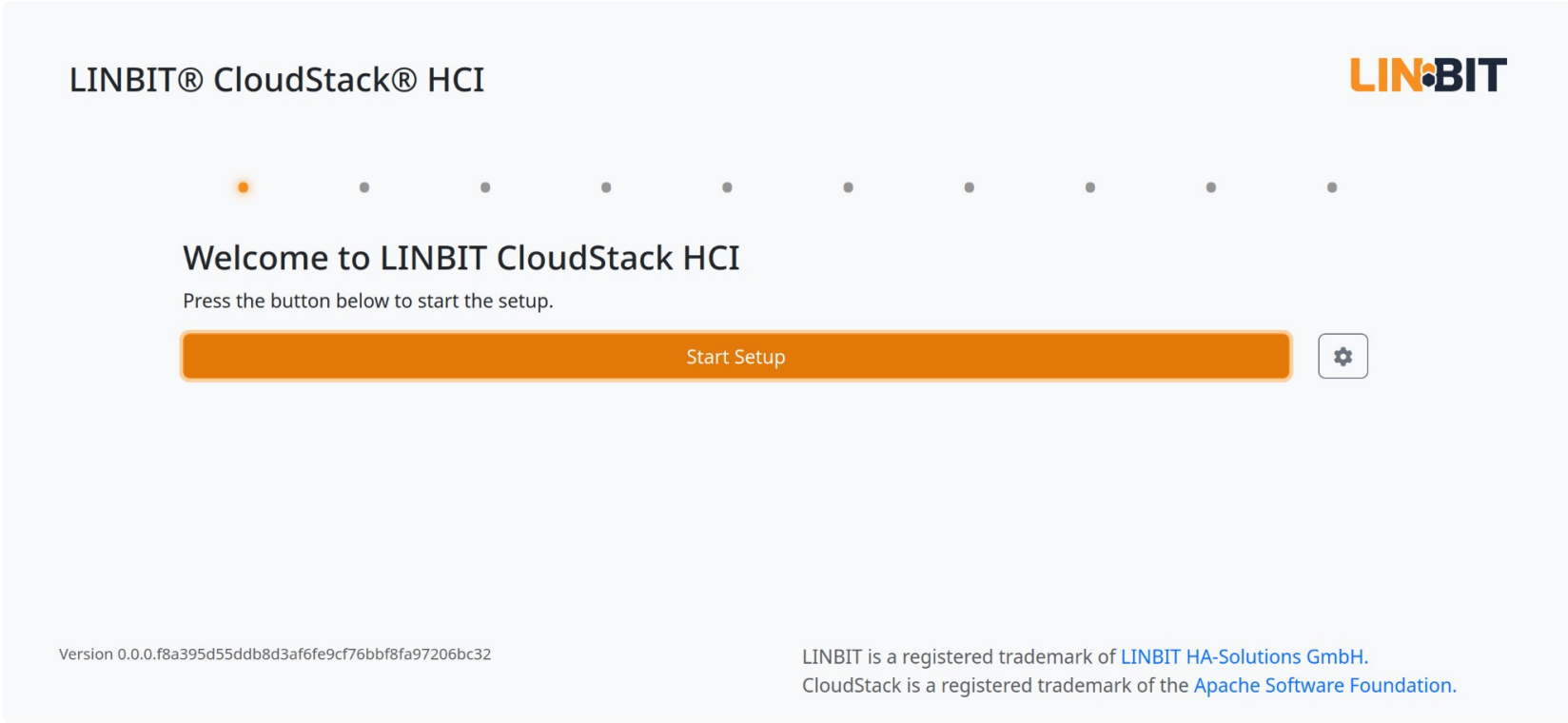
LINBIT CloudStack HCI Appliance

LINBIT CloudStack HCI Appliance

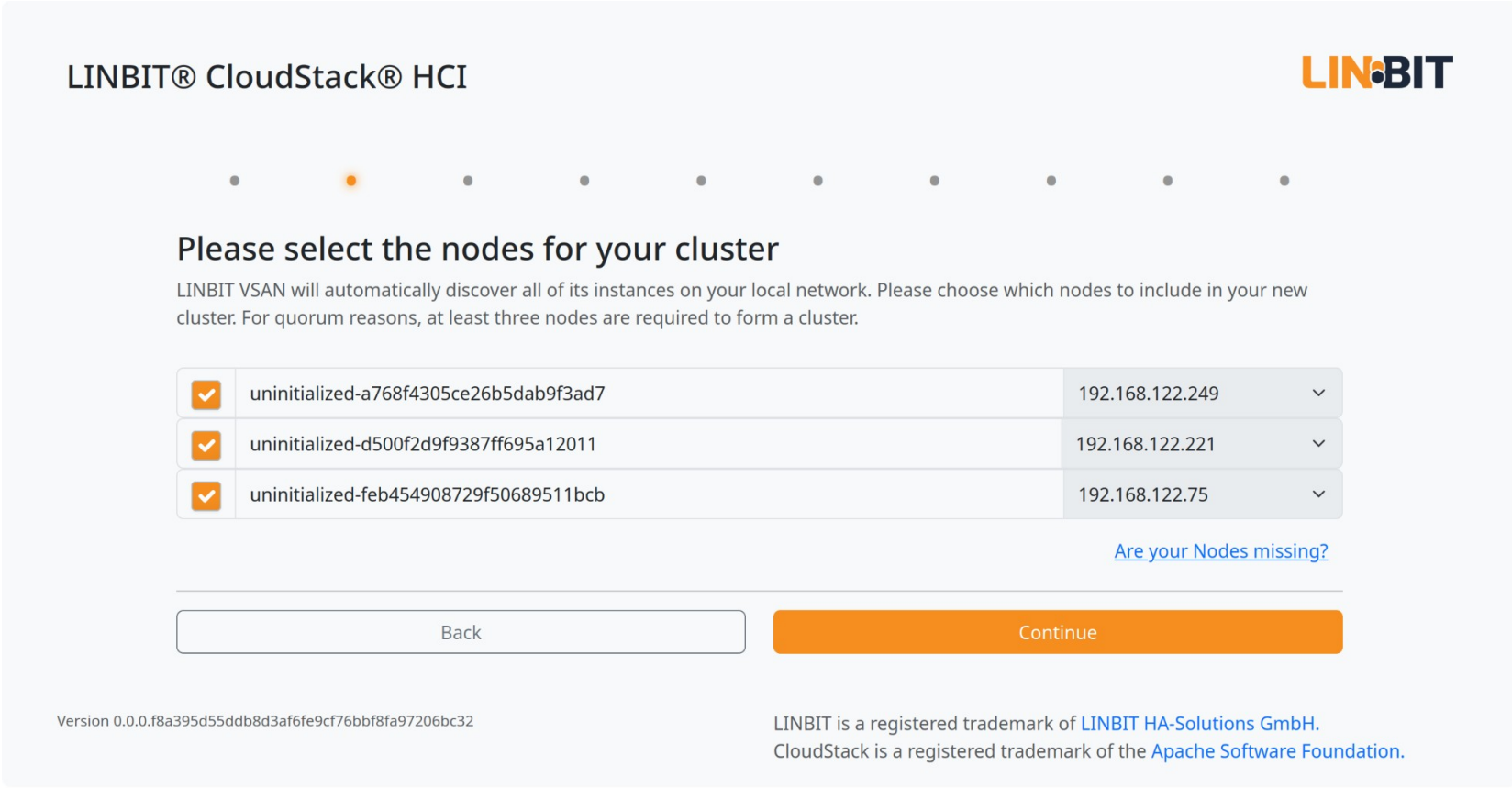
LINBIT

- Bootable ISO File
- About 20 minutes setup time
- Primary & Secondary storage configured
- Everything highly available
- 3 nodes minimum
- Tech preview at the moment
- Enterprise options will come in January 2025

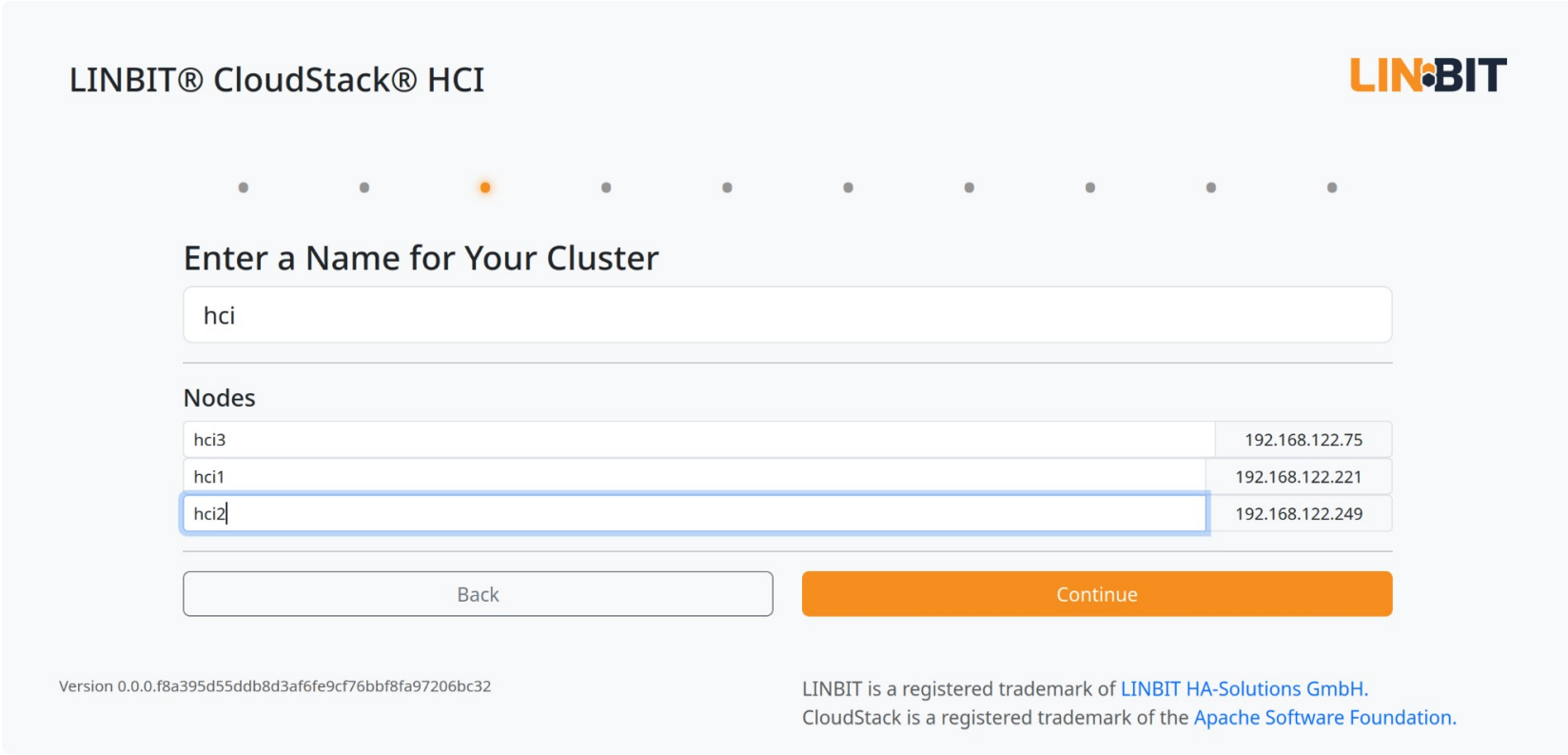
LINBIT CloudStack HCI Appliance



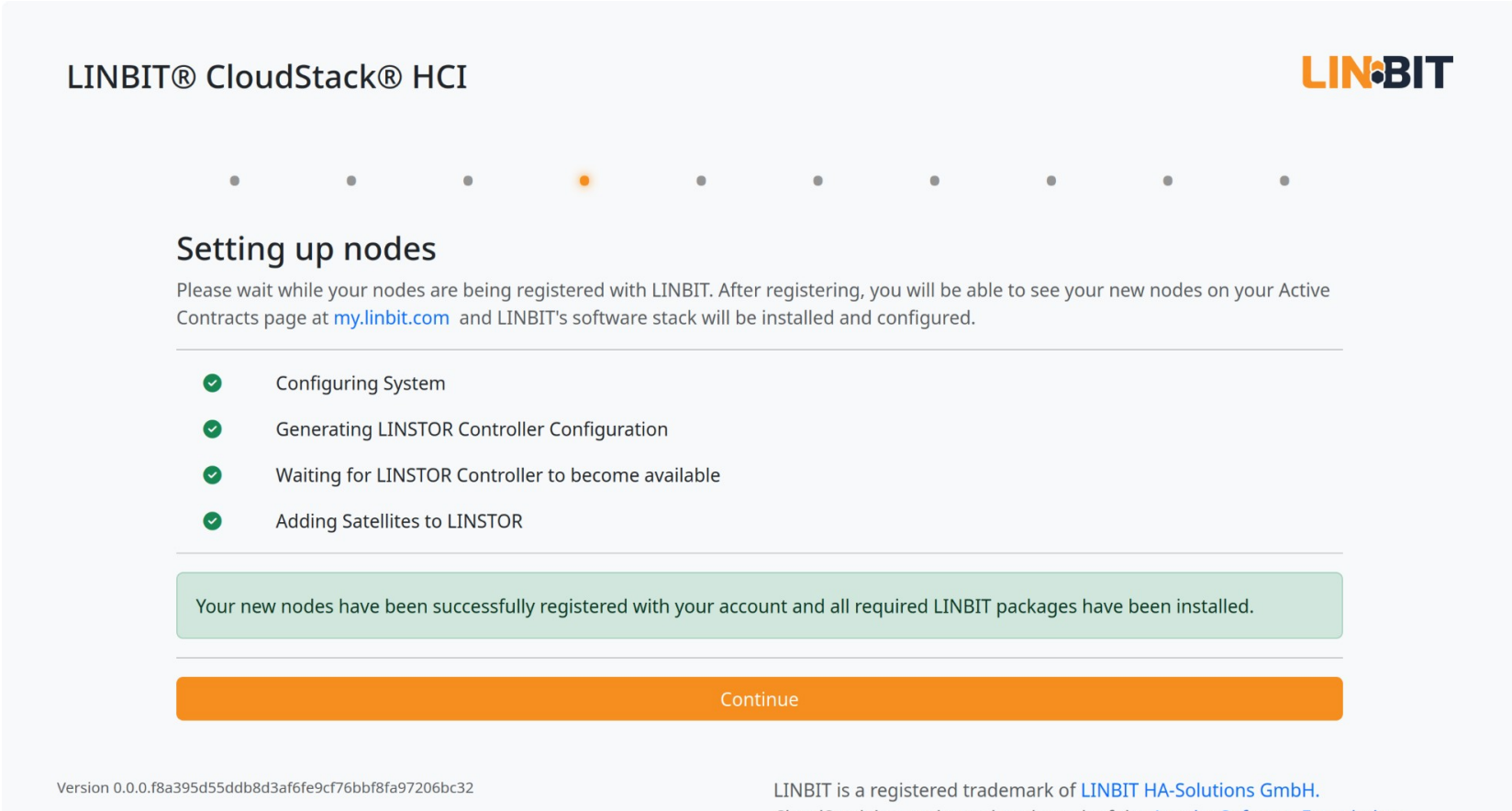
LINBIT CloudStack HCI Appliance



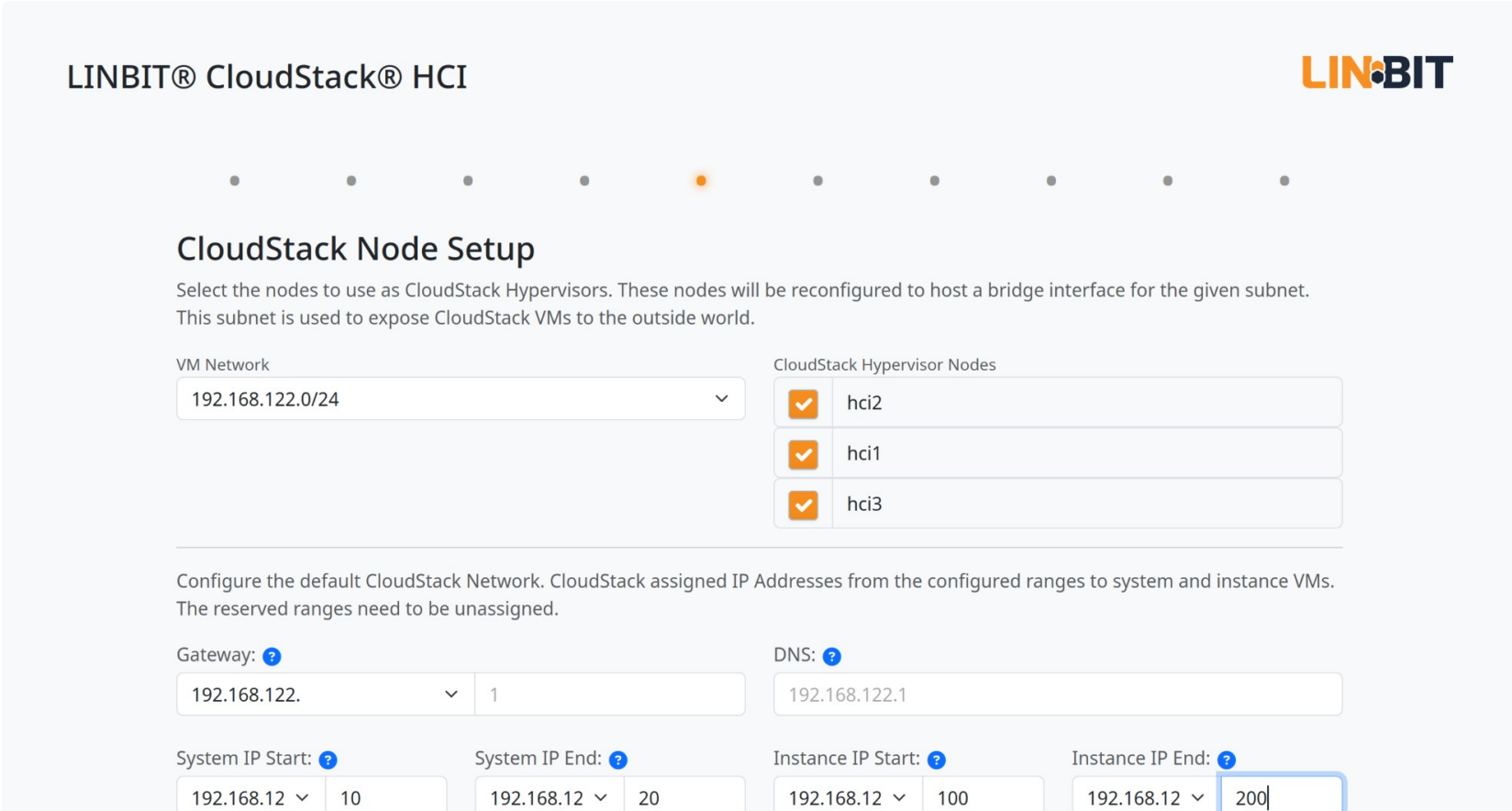
LINBIT CloudStack HCI Appliance



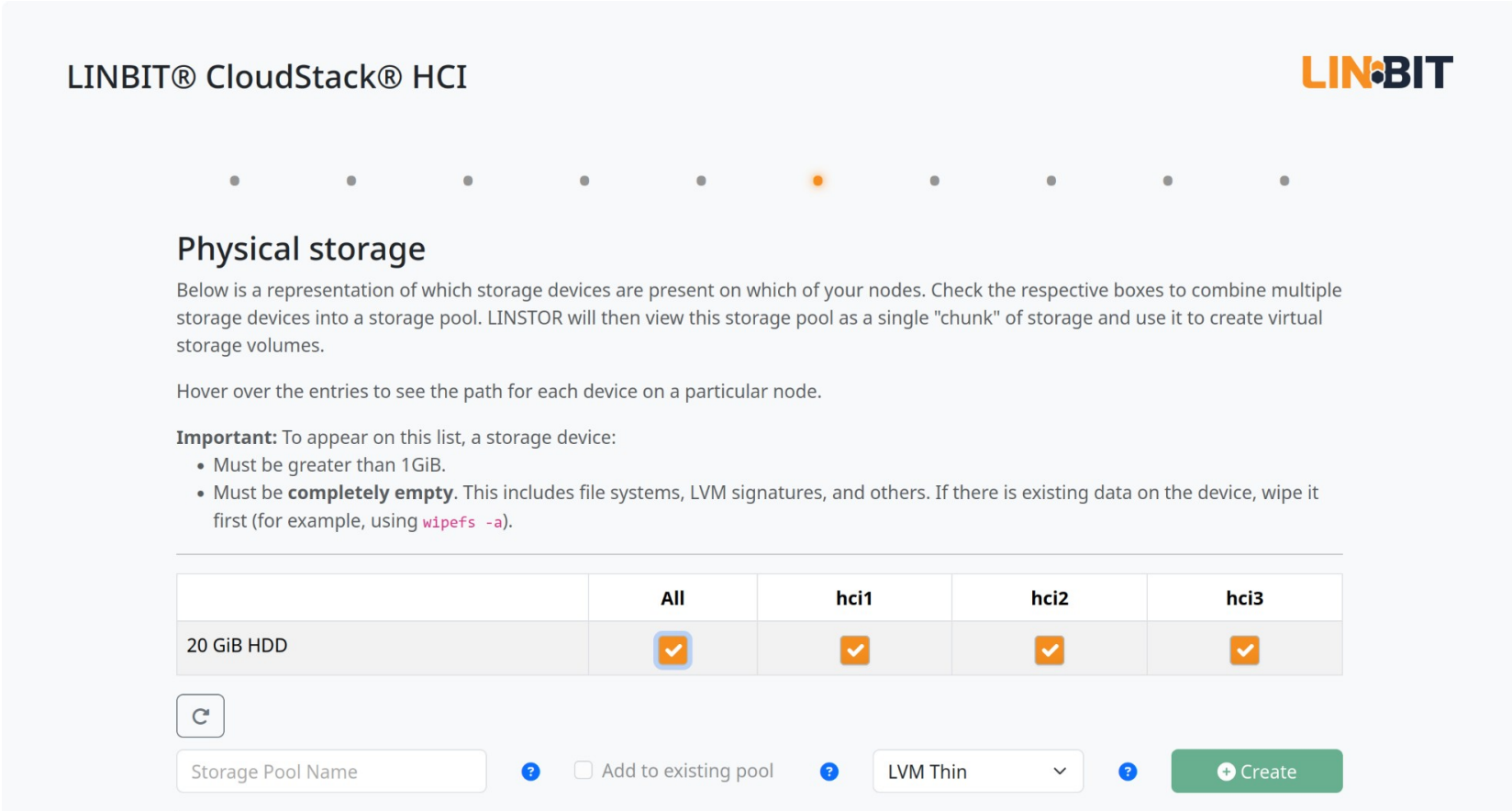
LINBIT CloudStack HCI Appliance



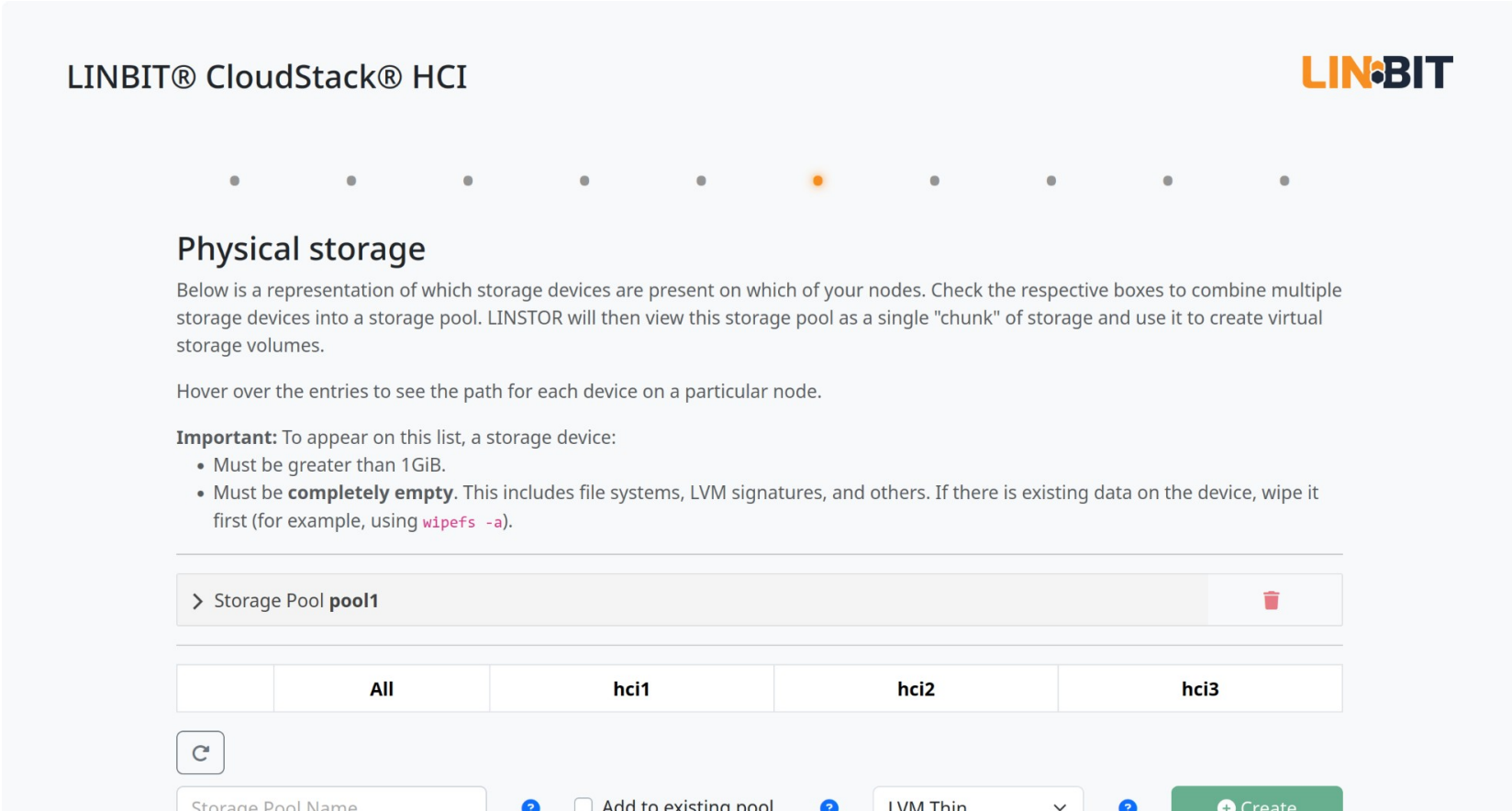
LINBIT CloudStack HCI Appliance



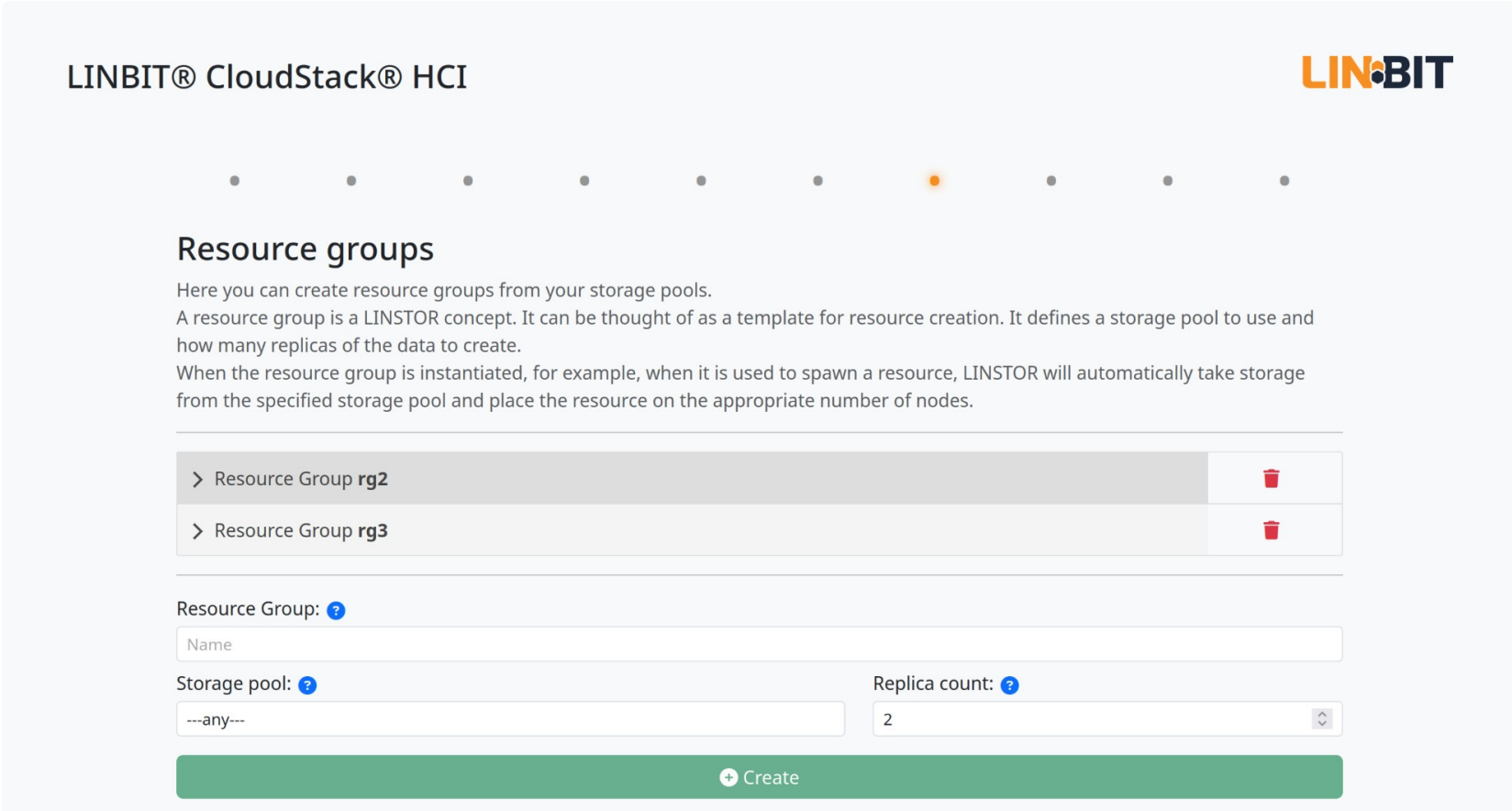
LINBIT CloudStack HCI Appliance



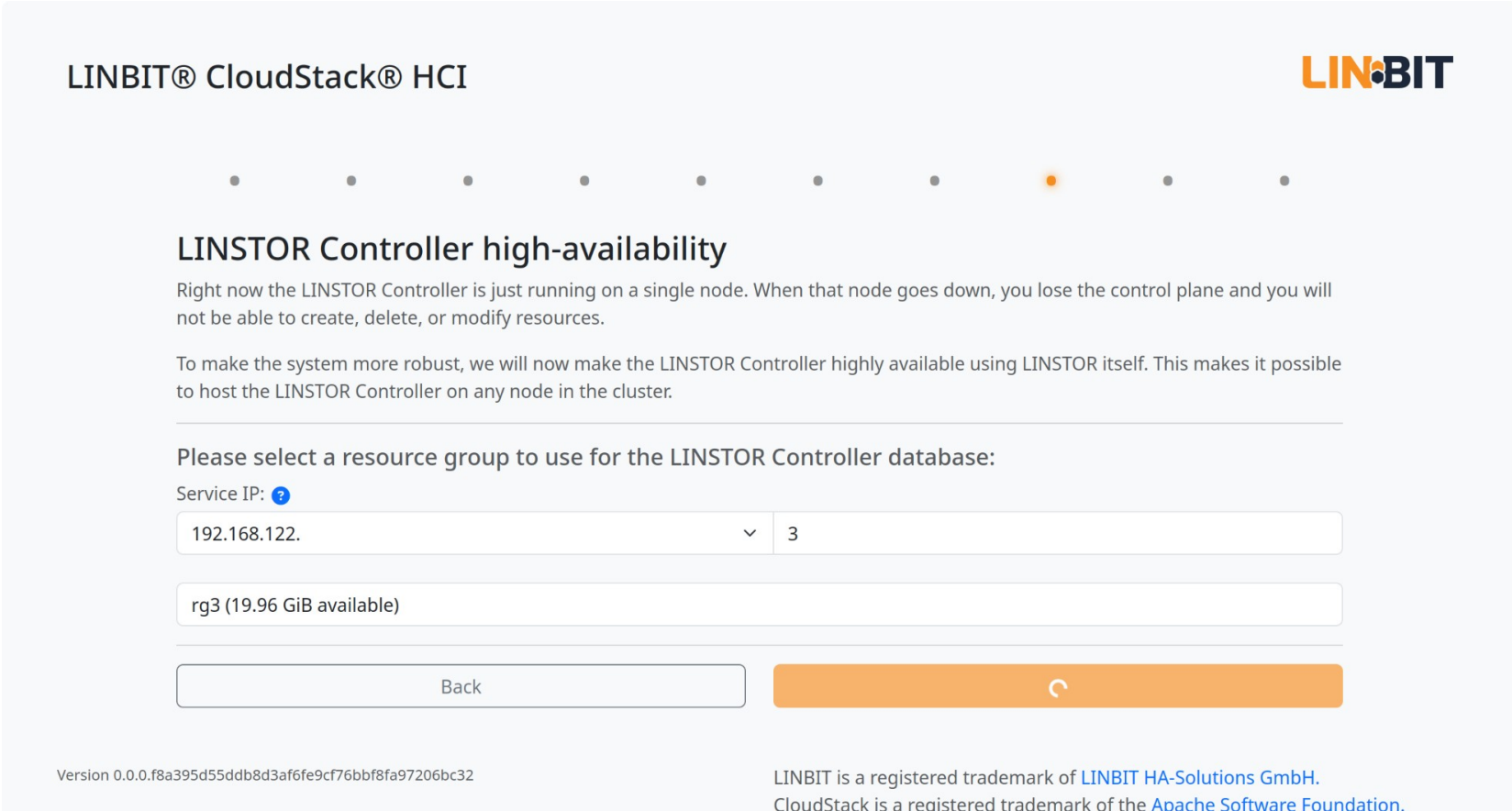
LINBIT CloudStack HCI Appliance



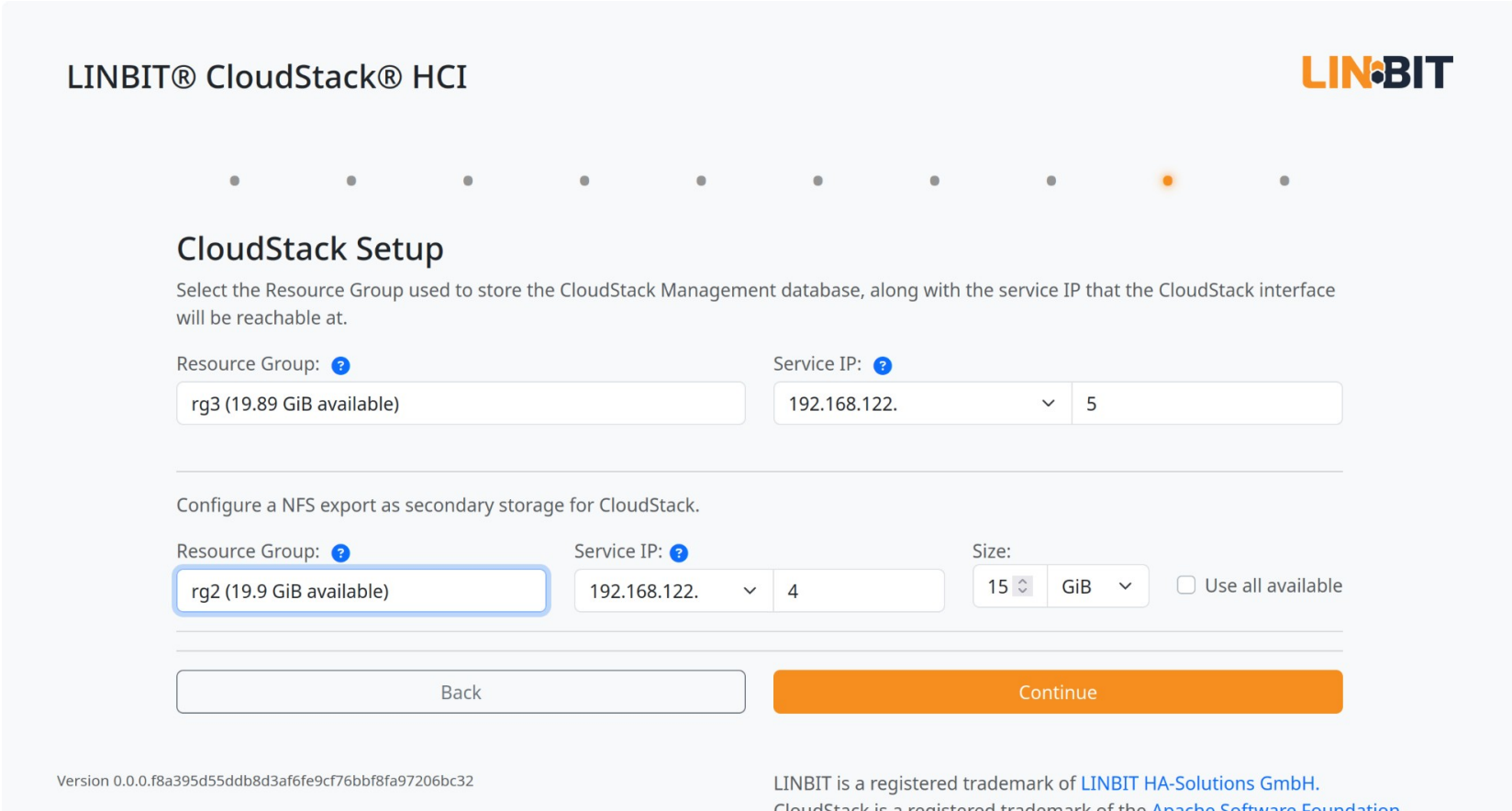
LINBIT CloudStack HCI Appliance



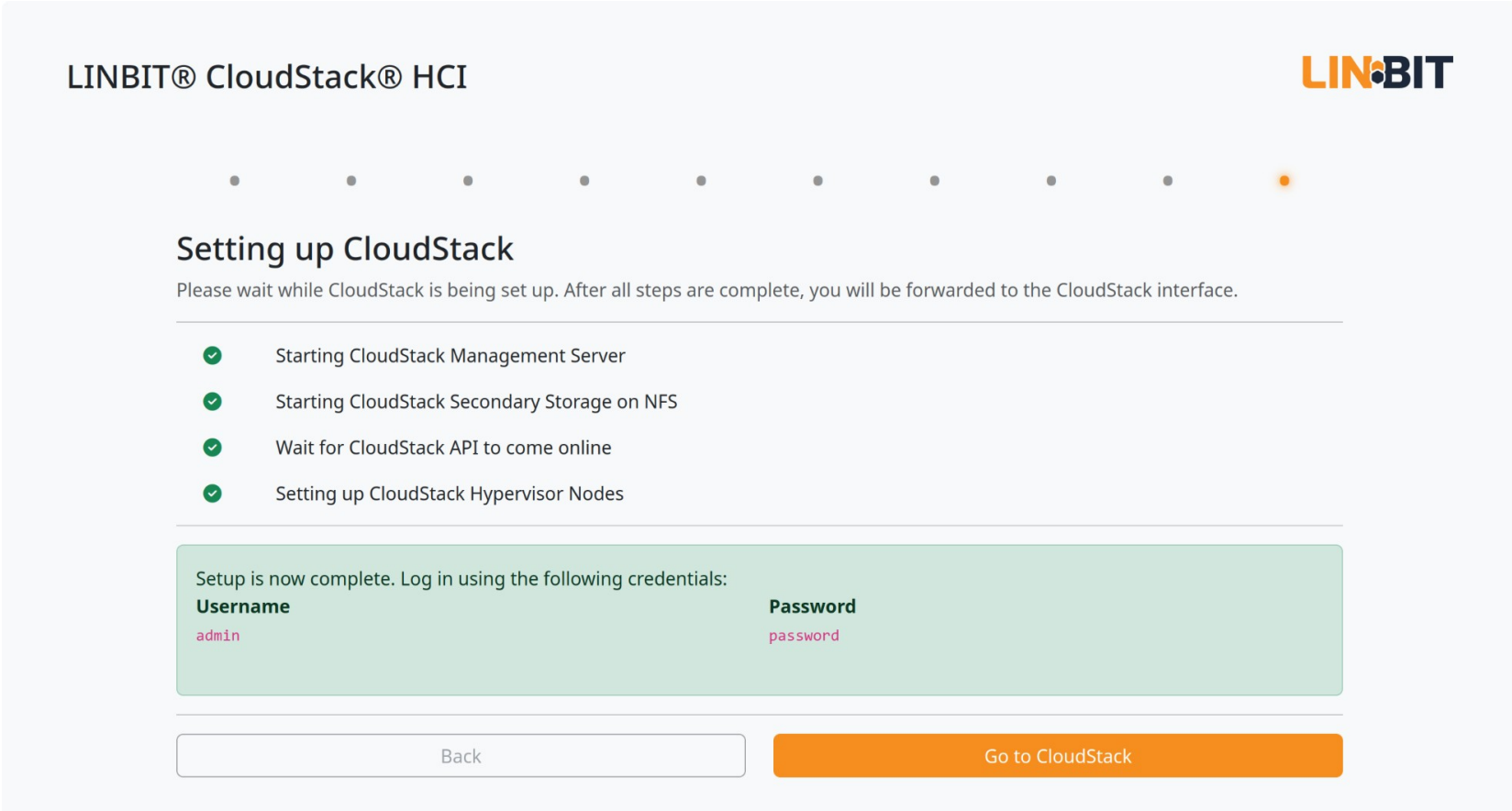
LINBIT CloudStack HCI Appliance



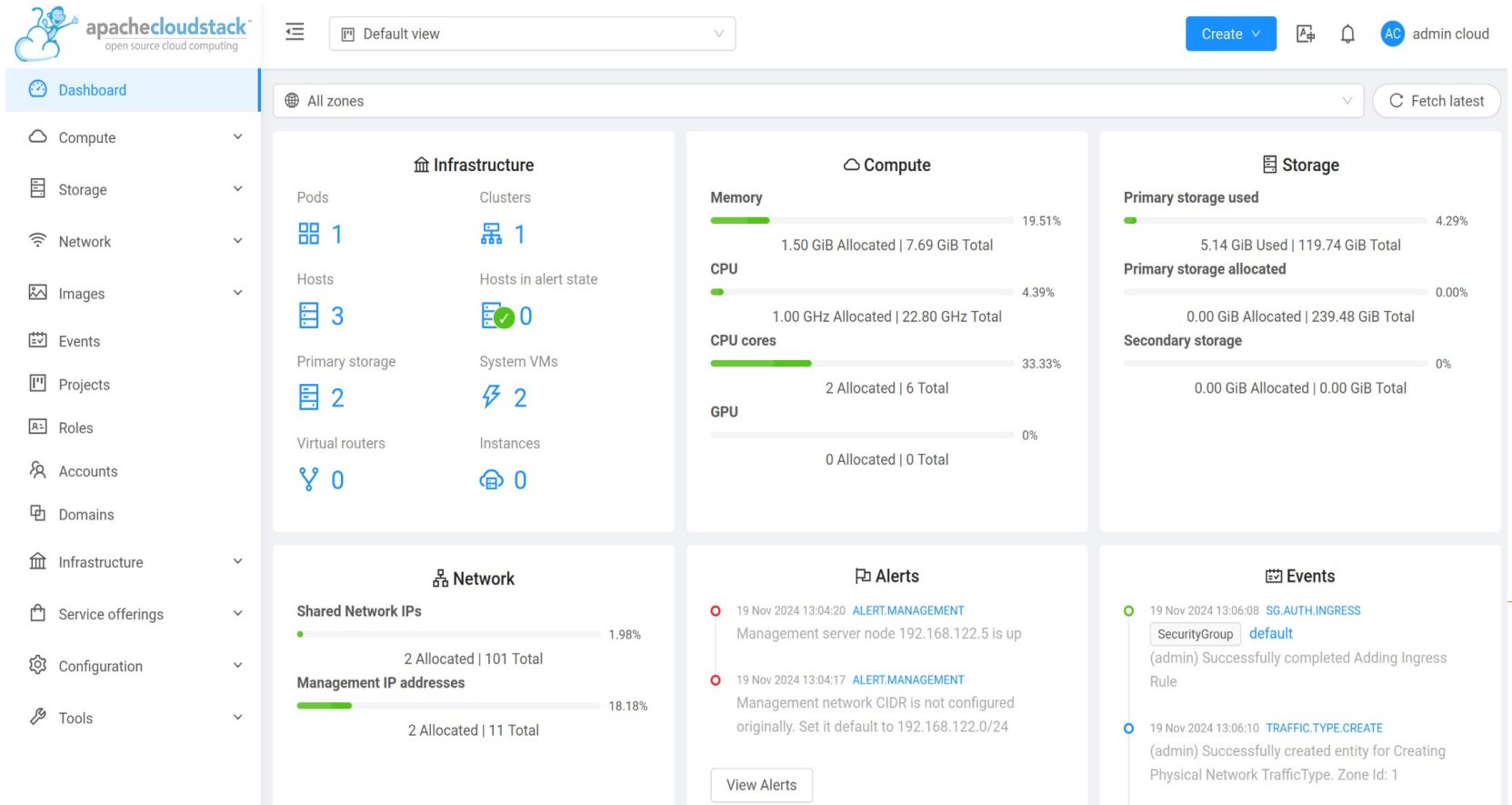
LINBIT CloudStack HCI Appliance



LINBIT CloudStack HCI Appliance



LINBIT CloudStack HCI Appliance



LINBIT CloudStack HCI Appliance

https://linbit.com/linbit-cloudstack-hci-appliance

Get it Today!

LINBIT

Thank you

https://www.linbit.com