
# FOSDEM21

Software defined storage devroom

# LINSTOR

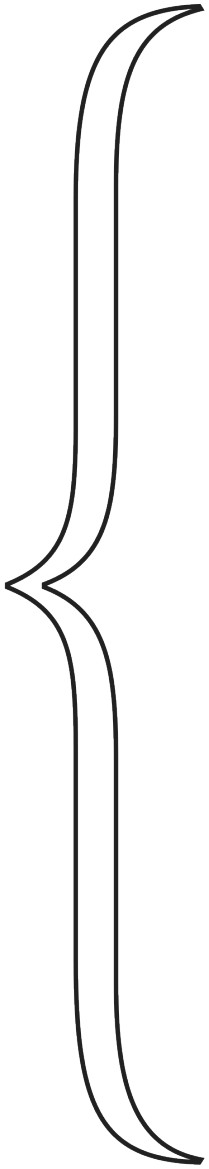
LINBIT

- Storage management automation
- Provides storage for various consumers on a cluster of nodes
- Manages backend storage
  (LVM, ZFS, fat- & thin-provisioning, NVMe-oF, etc.)
- Manages various optional layers
  (encryption, deduplication, caching, ...)
- Manages replication with DRBD
- Manages cluster-consistent snapshots
- Automates allocation of resources
  (Minor numbers, TCP/IP port numbers, etc.)

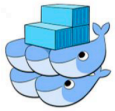
LINSTOR

LINBIT

Connectors



Kubernetes



Docker Swarm, Mesos

OpenStack Cinder


OpenNebula

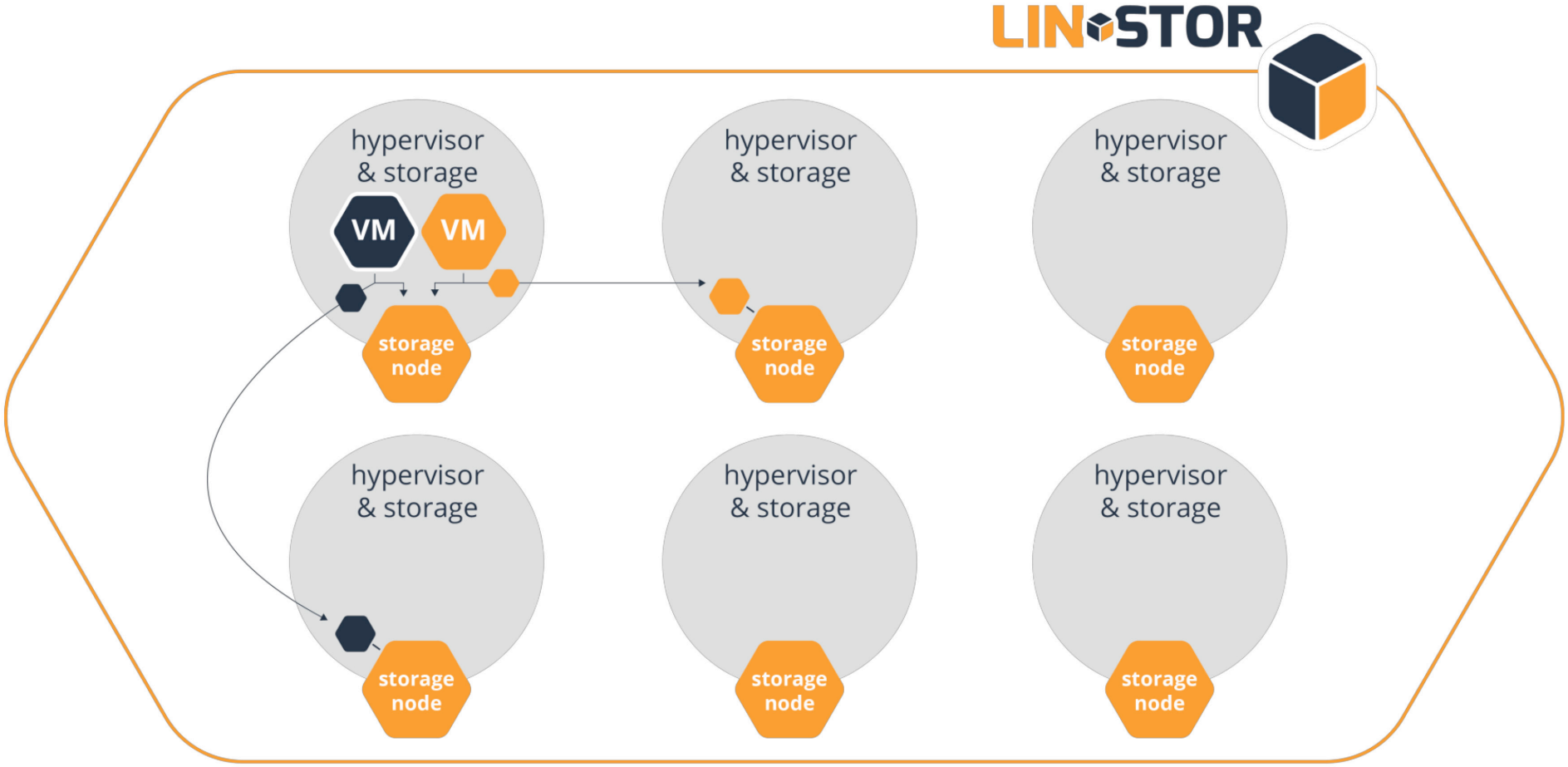
Proxmox VE


XenServer, XCP-ng

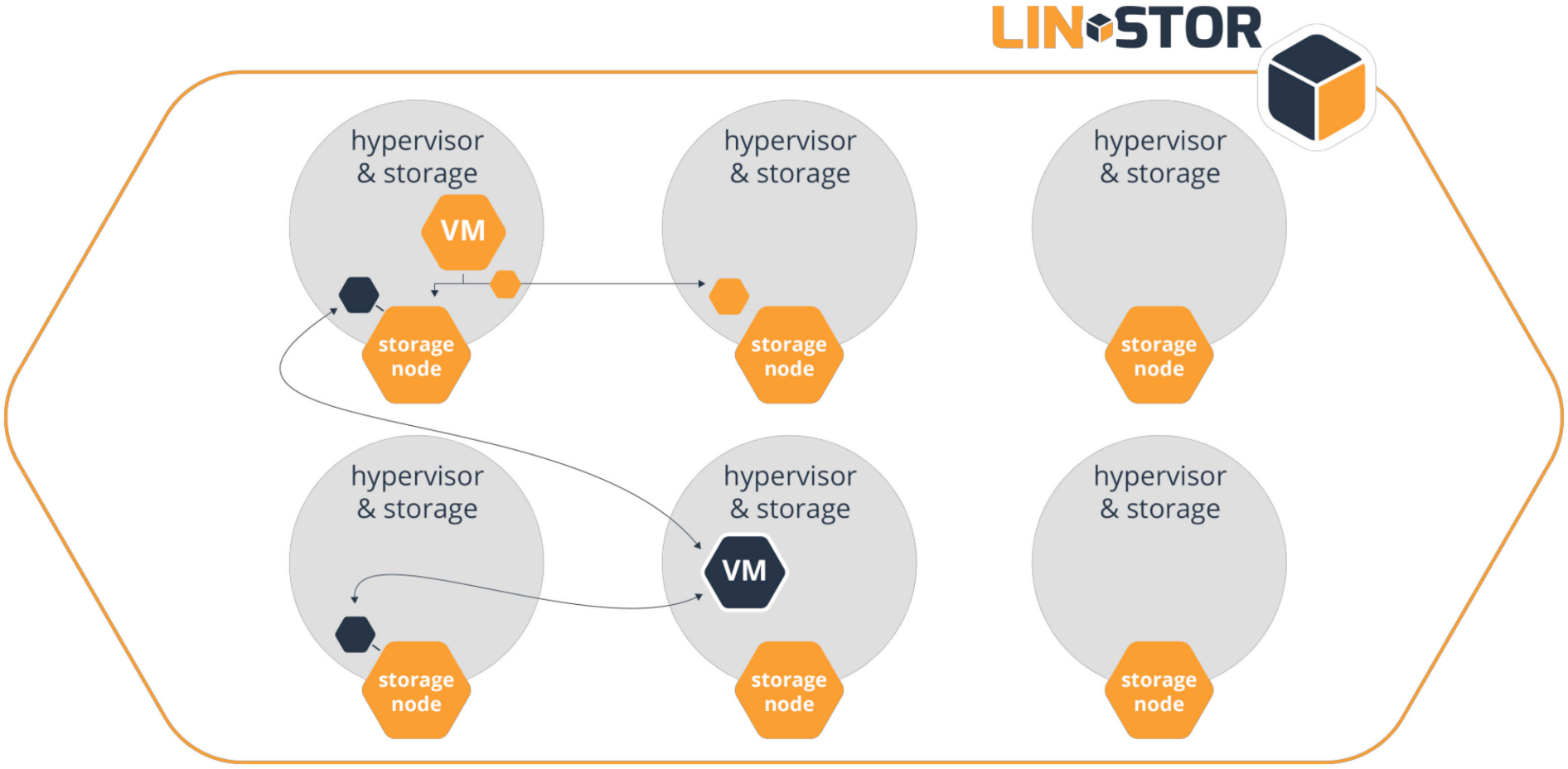

# Hyperconverged

Node-local storage

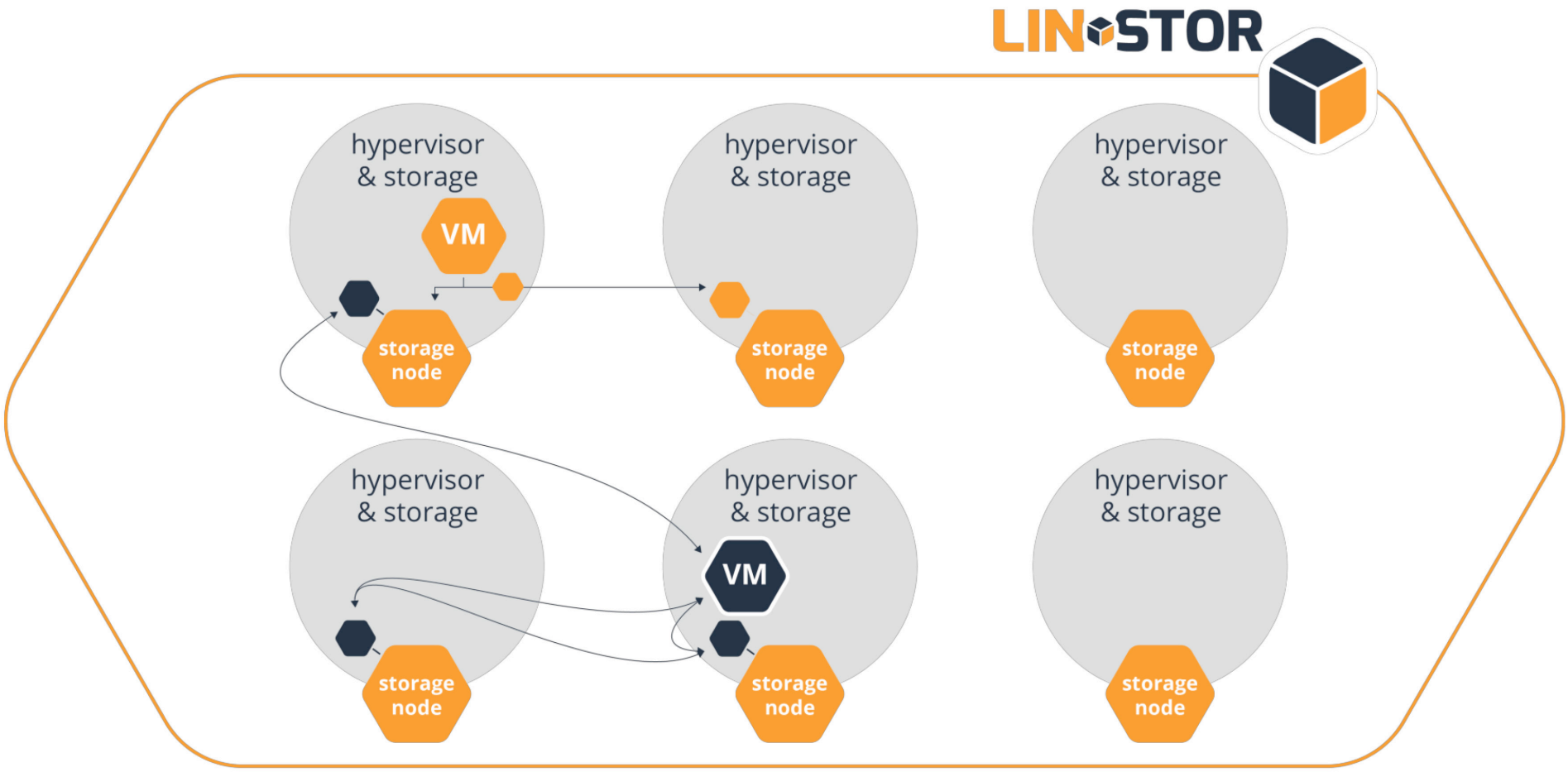
# LINSTOR - Hyperconverged



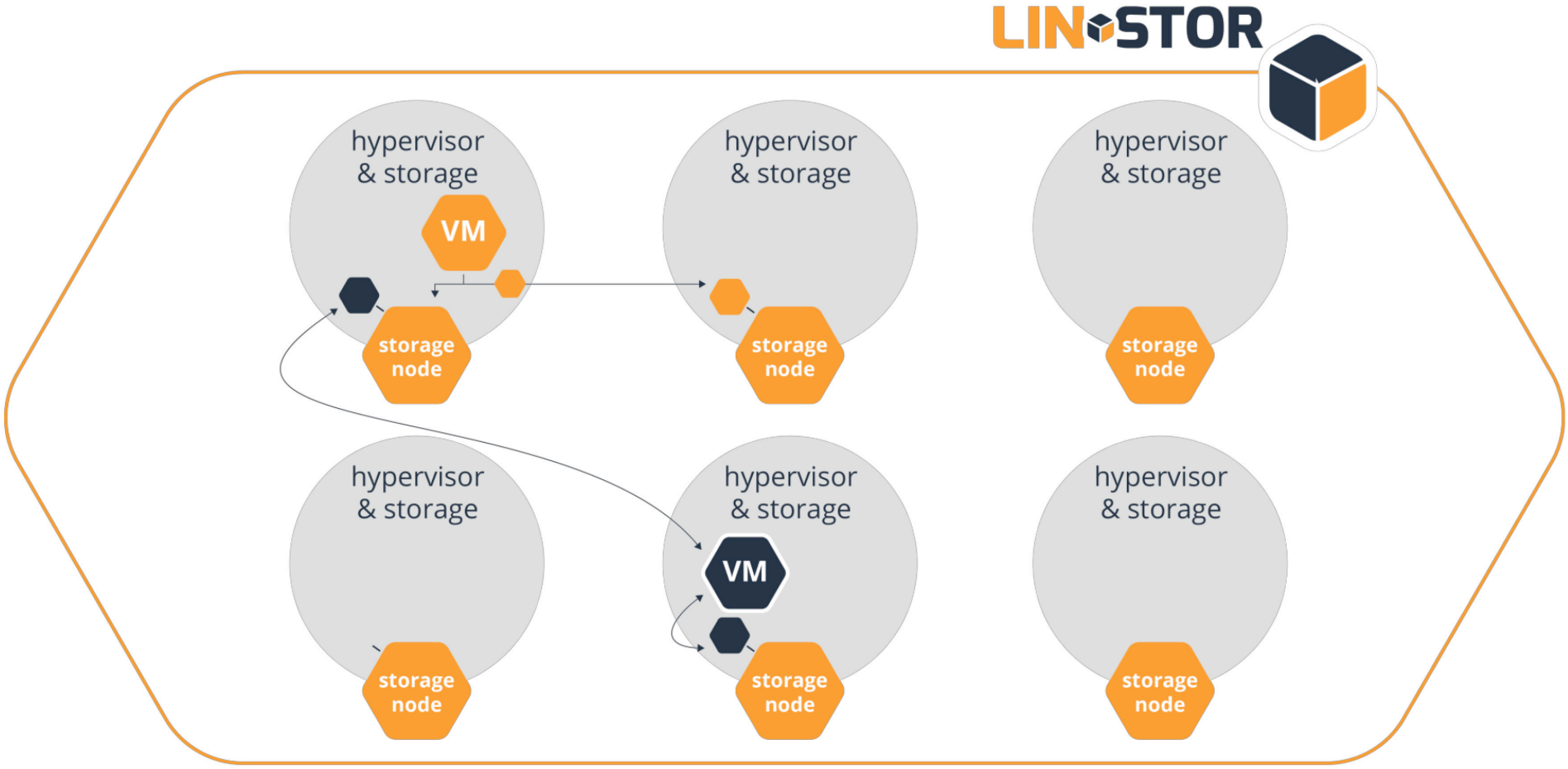
# LINSTOR - VM migrated



# LINSTOR - add local replica



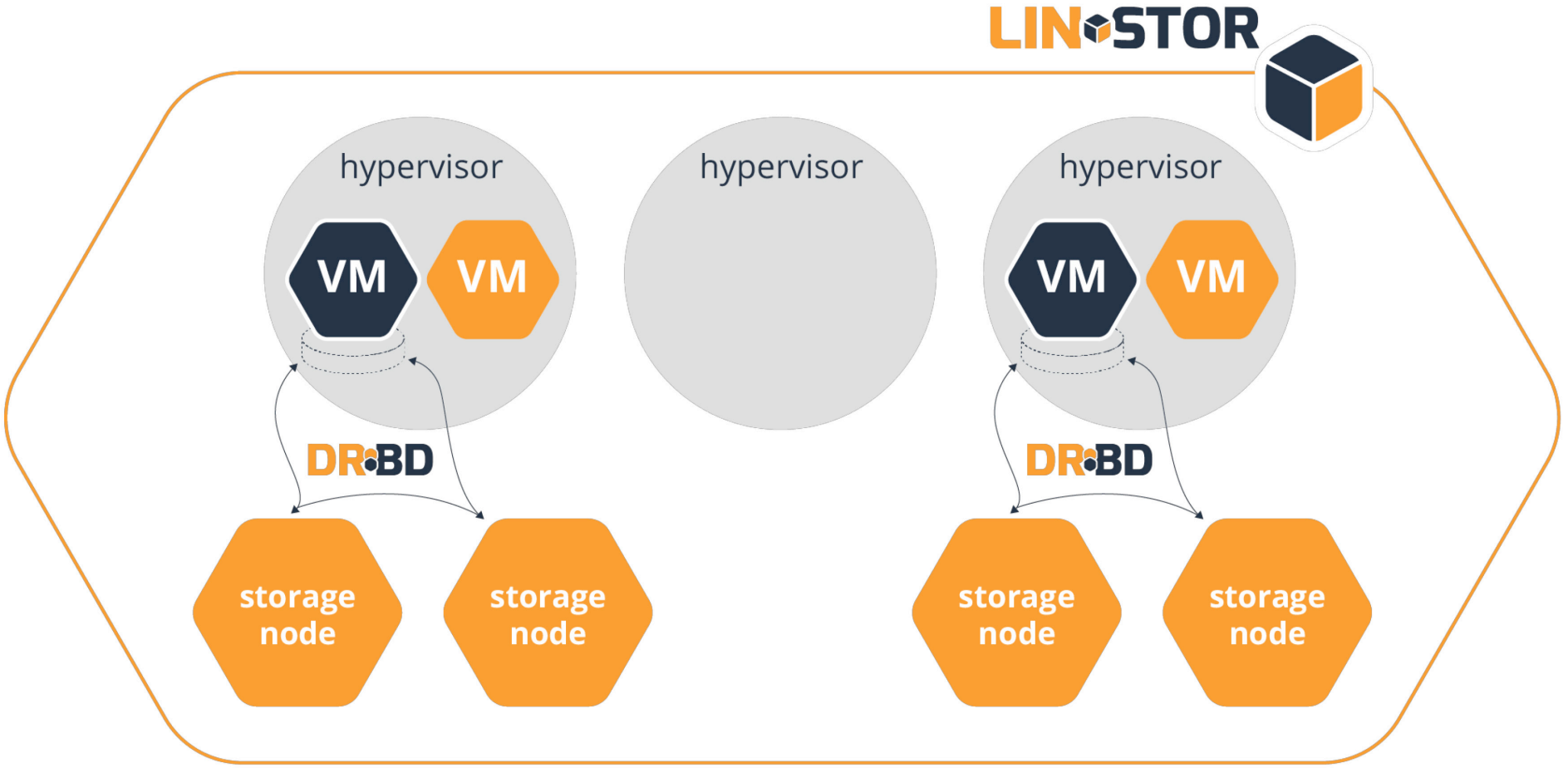
# LINSTOR - remove 3rd copy



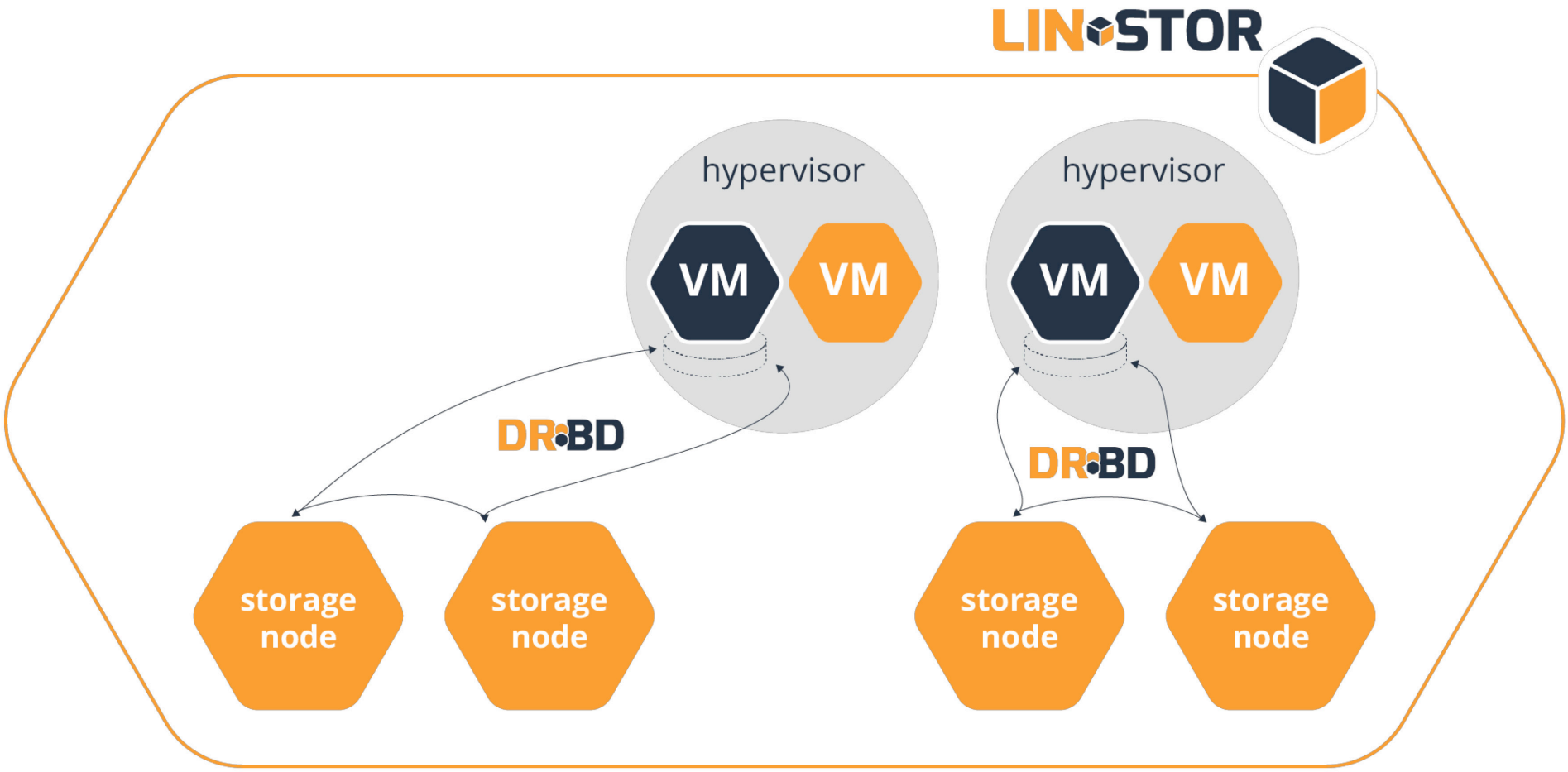

# Separate storage nodes

Remote storage

# LINSTOR – disaggregated stack

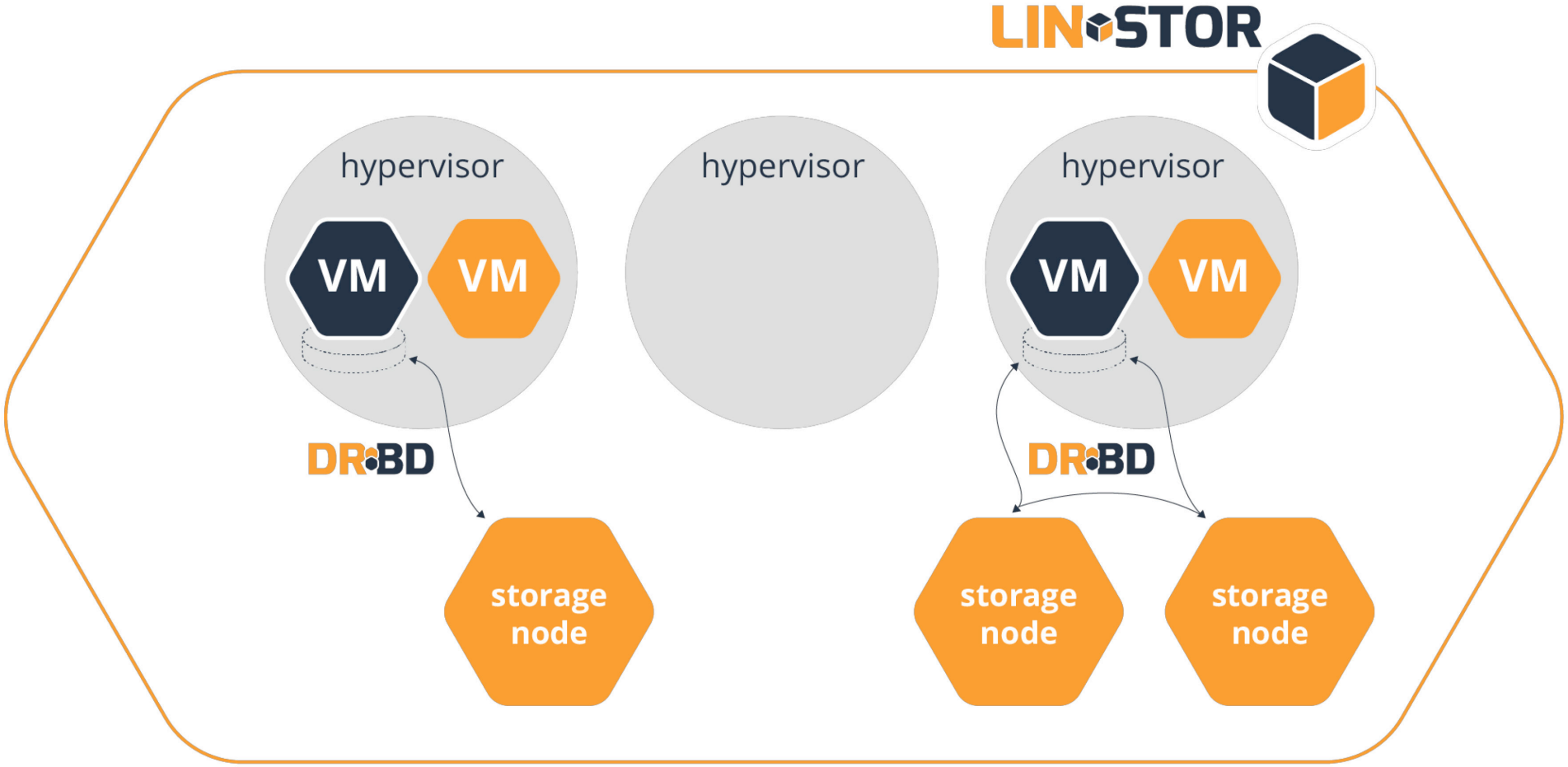


# LINSTOR / failed Hypervisor



# LINSTOR / failed storage node

LINBIT




# Architecture and external components

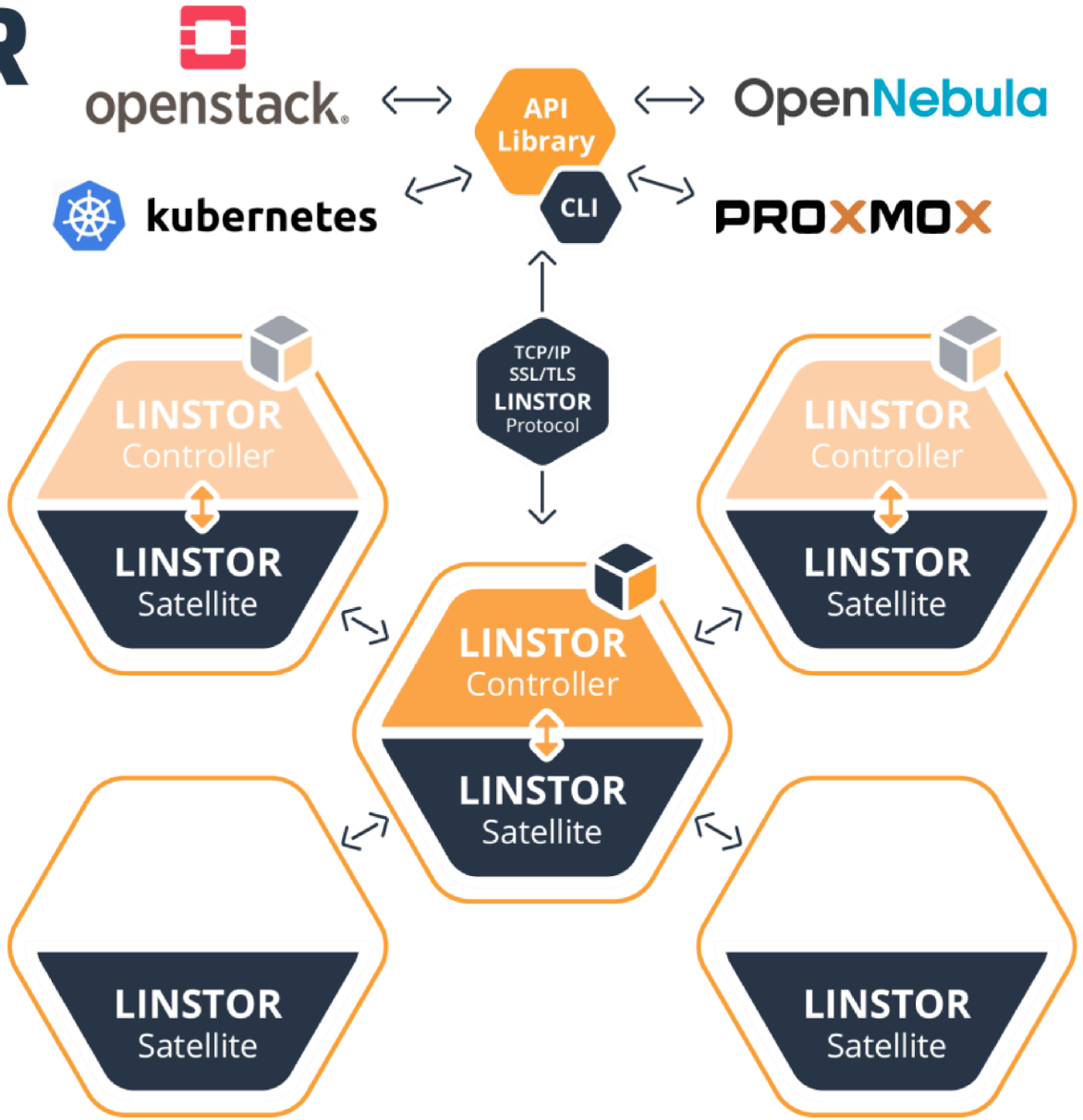
LINSTOR LINBIT

- Controller with transaction-safe persistence
- Stateless Satellites
- Platform independent (Java VM)
- Control-plane / data-plane separation
- Optional Controller high availability with Pacemaker
- HA Persistence: Replicated storage, shared storage, DB replication

LINSTOR

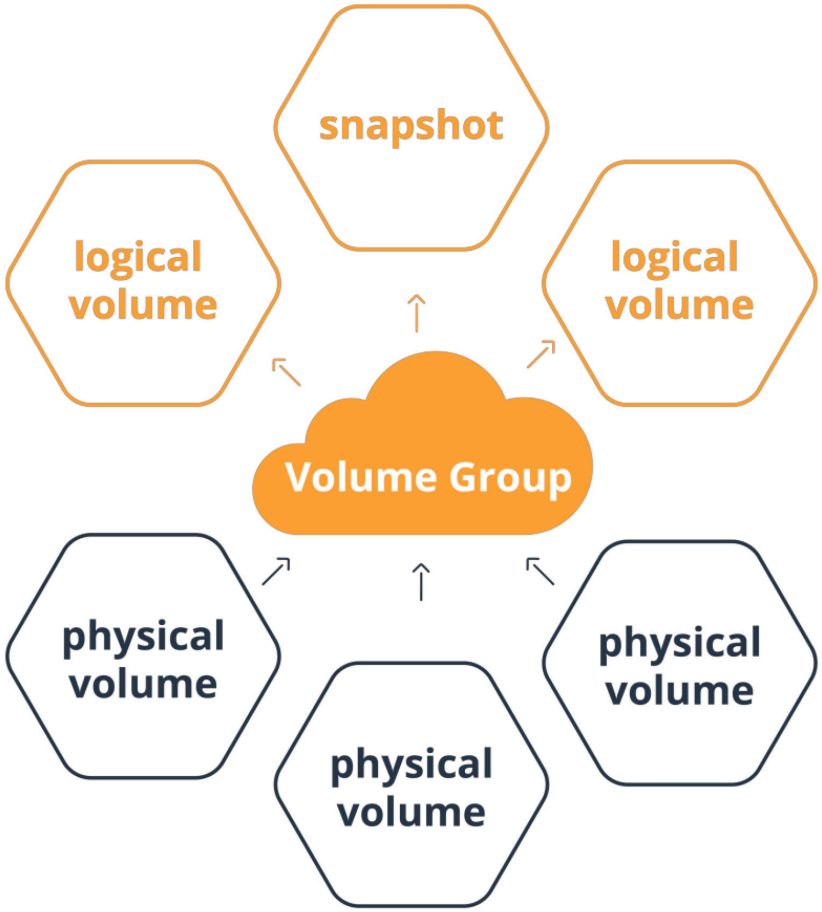
# Architecture

LINBIT

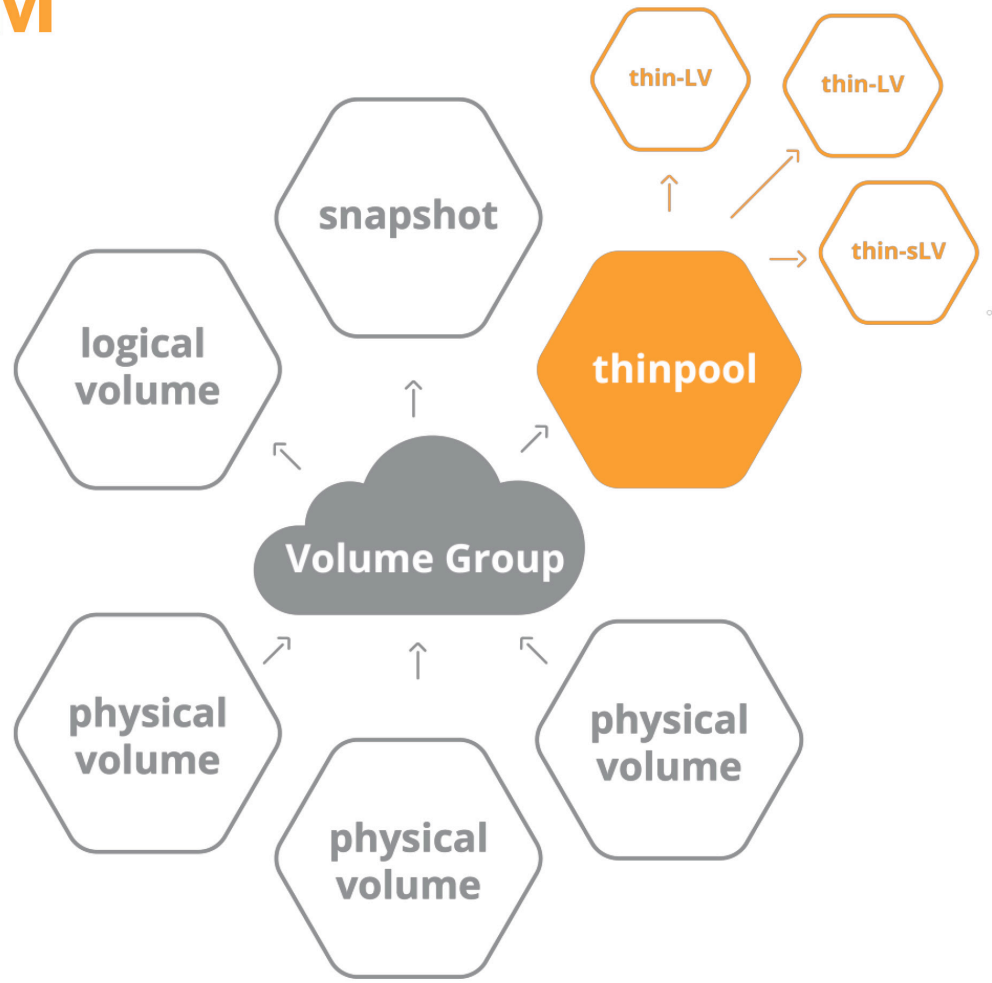


# Linux's LVM

LINBIT



# Linux's LVM


thin-LV
thin-LV
thin-sLV
snapshot
logical volume
thinpool
Volume Group
physical volume
physical volume
physical volume


# Data replication with DRBD

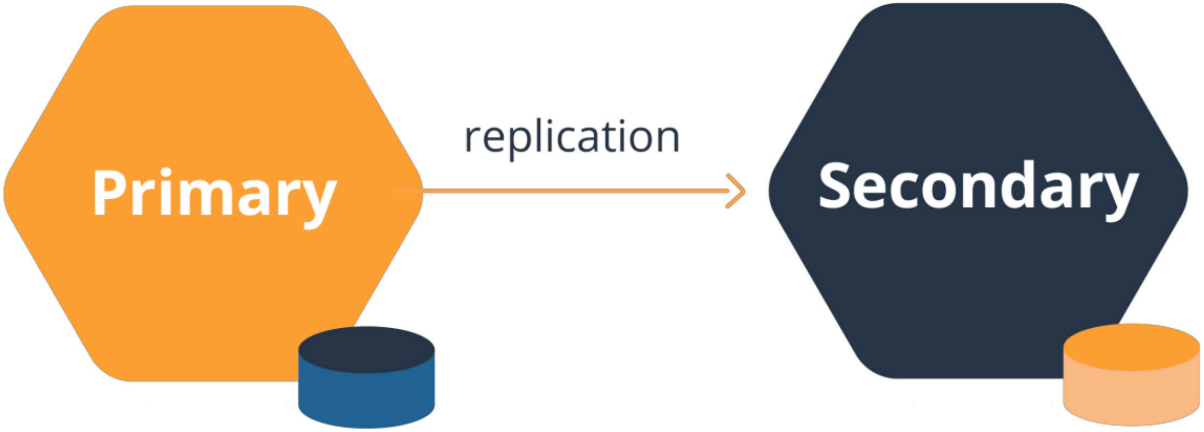
# DRBD - Main features

LINBIT

- Linux kernel module, GPL
- Synchronous or asynchronous replication
- Replicates any (random access) block device
- Consistency groups: Multiple volumes per resource
- Each resource available on up to 32 nodes
- Simple TCP/IP replication link
- Multi-pathing: Replication link fail-over
- Can be integrated with cluster resource managers
- Fencing & Quorum

# DRBD Roles: Primary & Secondary

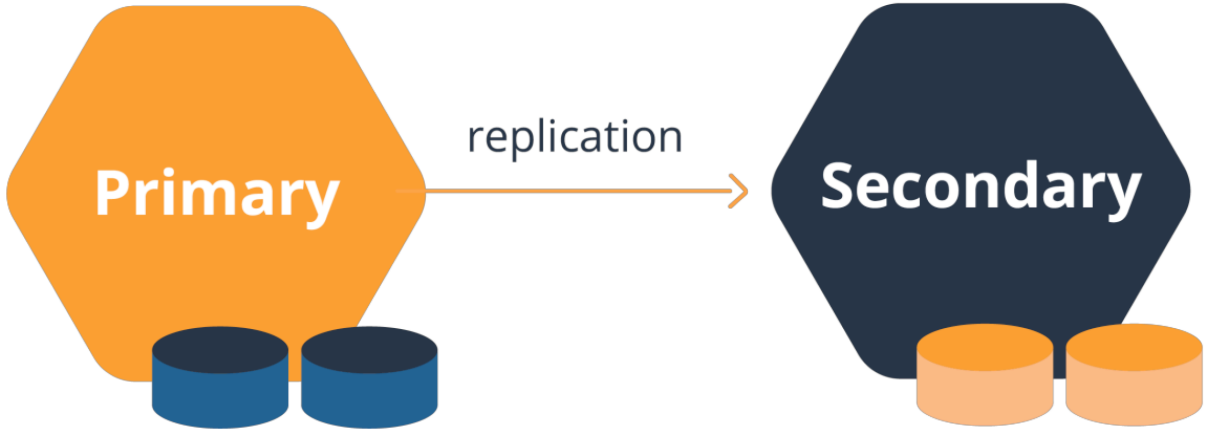
LINBIT



# DRBD – multiple Volumes

LINBIT

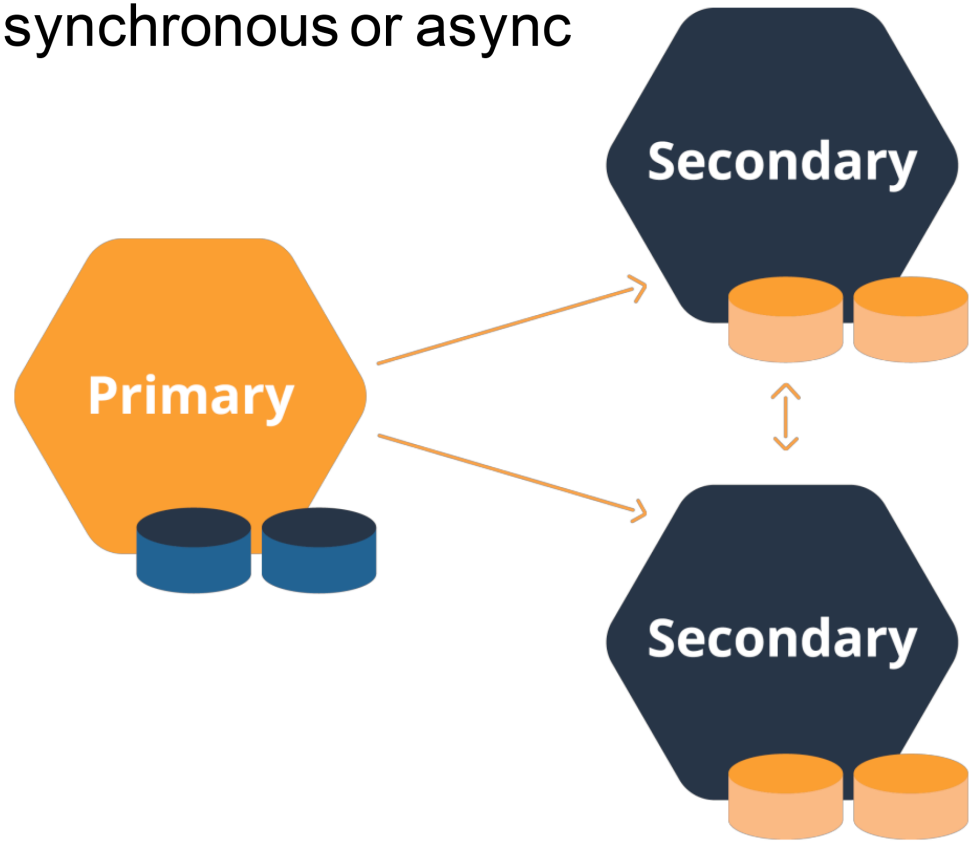
- consistency group



# DRBD – up to 32 replicas

LINBIT

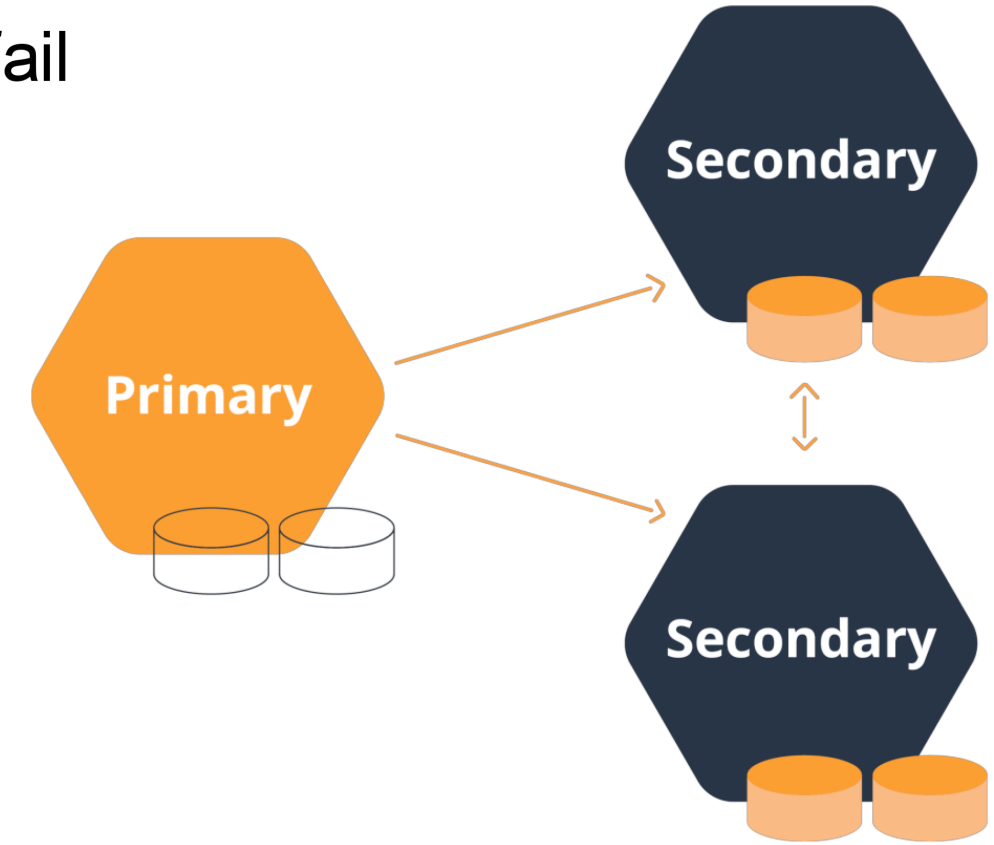
- each may be synchronous or async



# DRBD – Diskless nodes

LINBIT

- intentional diskless (no change tracking bitmap)
- disks can fail



# DRBD Quorum

LINBIT

- Quorum of the majority, all or a specified number of nodes
- Minimum redundancy
- Upon loss of quorum: suspends or fails I/O
- Minimum of three nodes

# DRBD Optimizations

LINBIT

- Infiniband, Dolphin Express
- RDMA (Remote Direct Memory Access):
  Infiniband/RDMA or RoCE (RDMA over converged Ethernet)
- Multi-path load-balancing with RDMA
- Meta data on PMEM / NVDIMM


# Dedicated storage systems

LINSTOR

## linstor-gateway

- Provides iSCSI and/or NFS resources
- Storage provided by LINSTOR
- High availability provided by Pacemaker/Corosync
- Integrates LINSTOR with Pacemaker


# More information

## LINSTOR at LINBIT

`https://www.linbit.com/linstor`

## LINSTOR on GitHub

`https://github.com/LINBIT/linstor-server`

## DRBD on GitHub

`https://github.com/LINBIT/drbd-9.0`