Avoiding Downtime Using Linux High Availability

Jeremy Rust

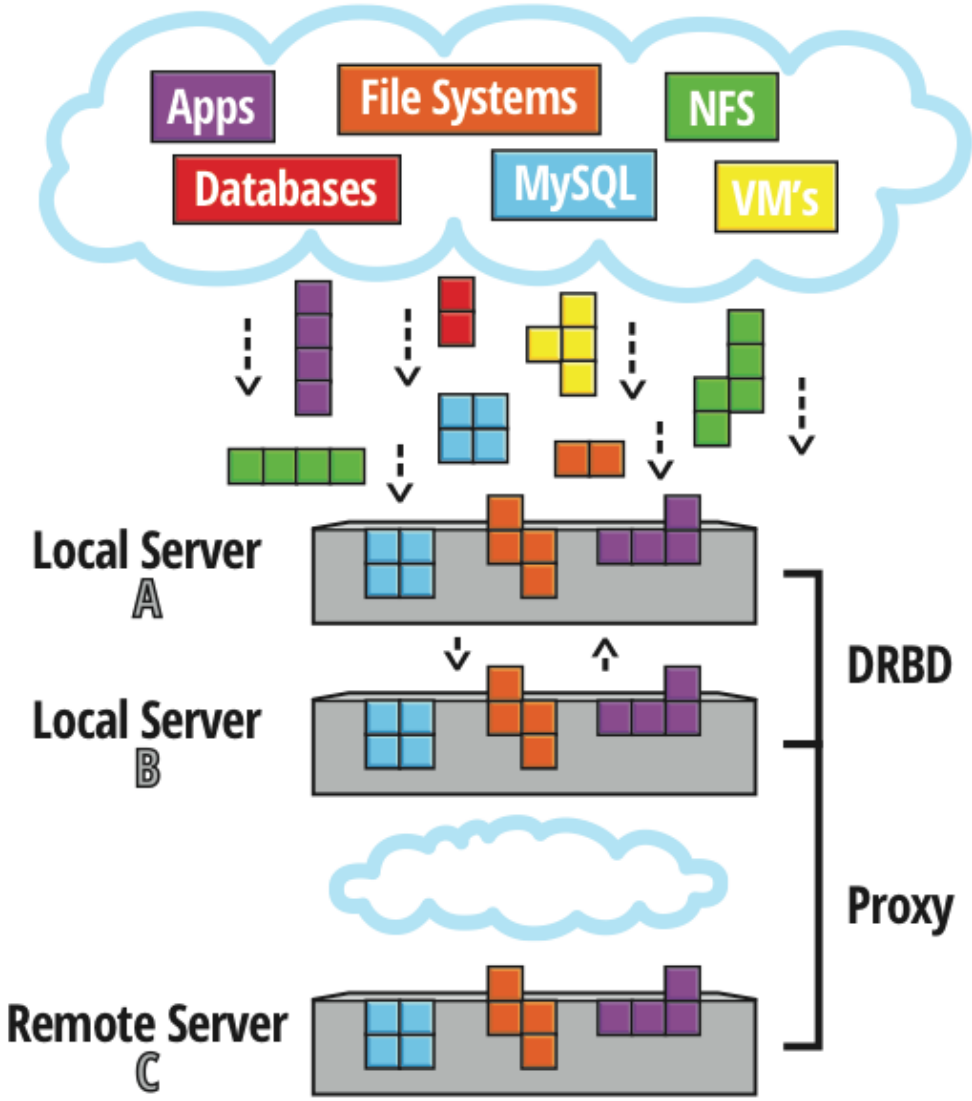
Jeremy@linbit.com
@linbit
@nerdhacker

Nagios® WORLD CONFERENCE 2014

Introduction & Agenda

- Downtime is not cheap
- What is High Availability = not a back up!
- Raid or Raid over the network (DRBD)
- SANs and clustered applications
- The Linux cluster stack
- Cluster management with Pacemaker
- Disaster Recovery / Linking sites
- DRBD and the Cloud

DRBD HA and DR



Downtime = $$$

- Lost revenue
- Lost reputation
- Almost every business these days has a critical database or file system that they could not do without.
- HP estimates $31,705 per hour 3.8 hours a year totaling $481,900/ year
- 40% internet traffic stops when Google goes down

Survey: Critical "IT-Systems in the medium-sized business", 2013 Techconsult, on behalf of HP Germany Basis: 300 medium-sized companies from Germany

Downtime = $$$

"YOU LOST THE DATABASE?!?!"

- "Ummm, can you ping ____?"
- "I can't seem to reach our inventory system."
- "Can you try pulling up this record?"

Devotion to Duty - xkcd

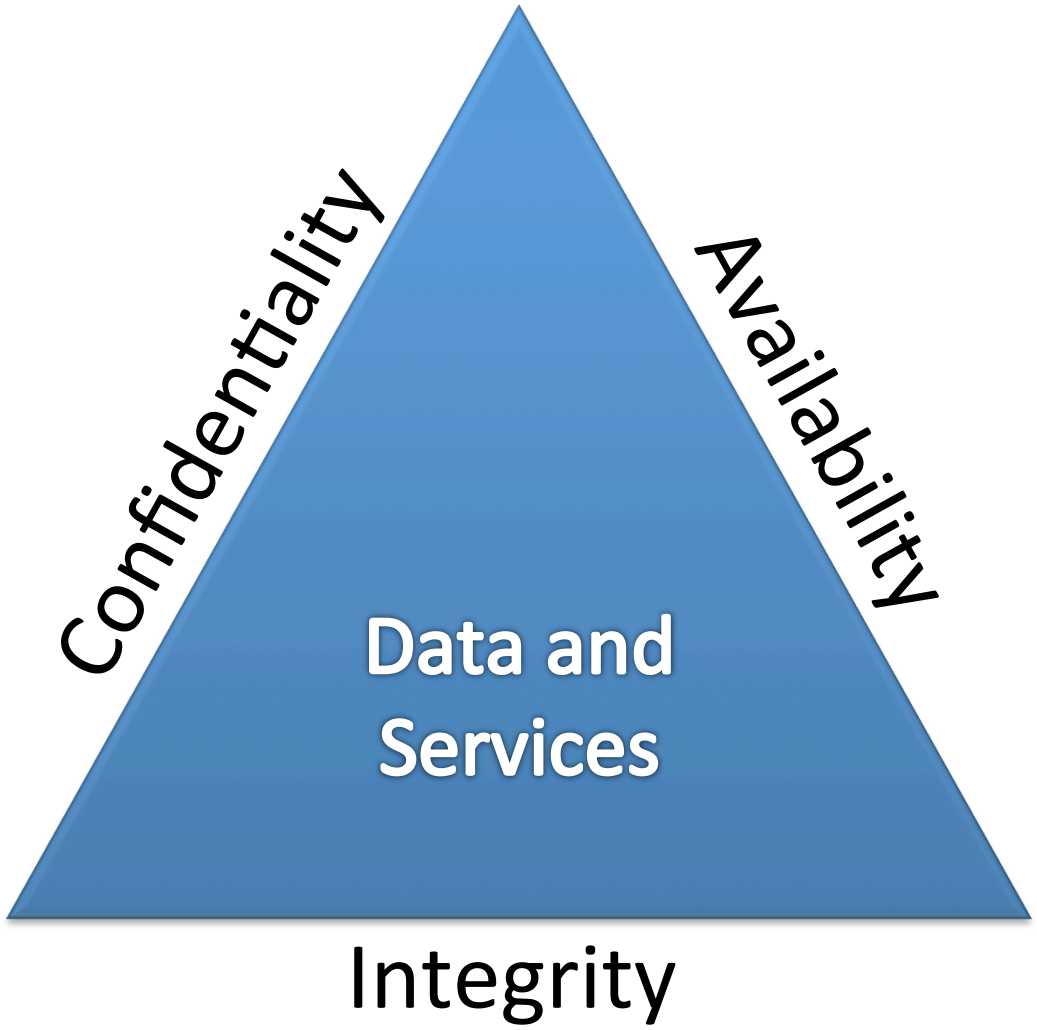
Why Monitor?

- Hardware dies
- DDOS attacks
- Set it and forget it mentality
- Internet connection
- Security programs

Hosting / XaaS

- Reliability
- Security
- Multi-tenant architecture
- Scalability
- Uptime

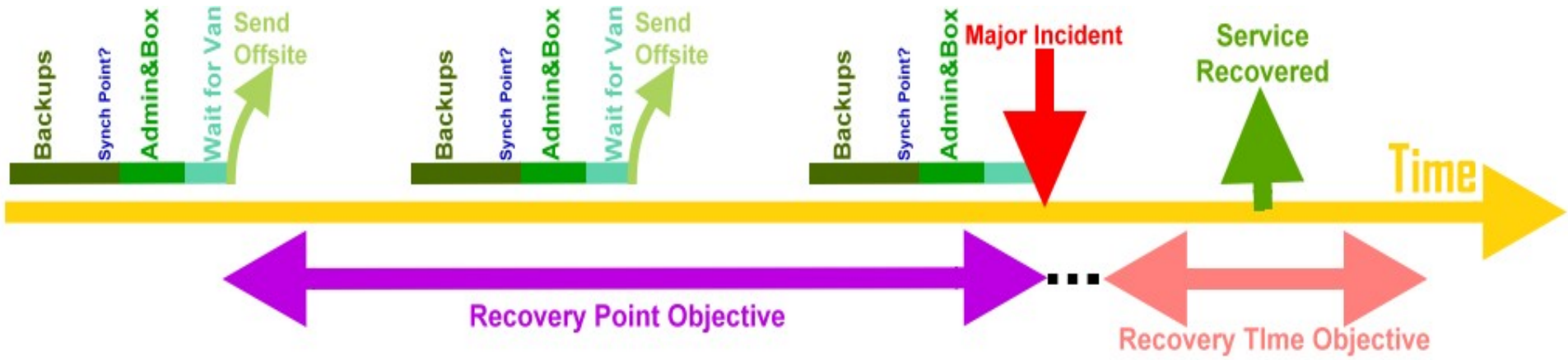
The Pillars of IT Security



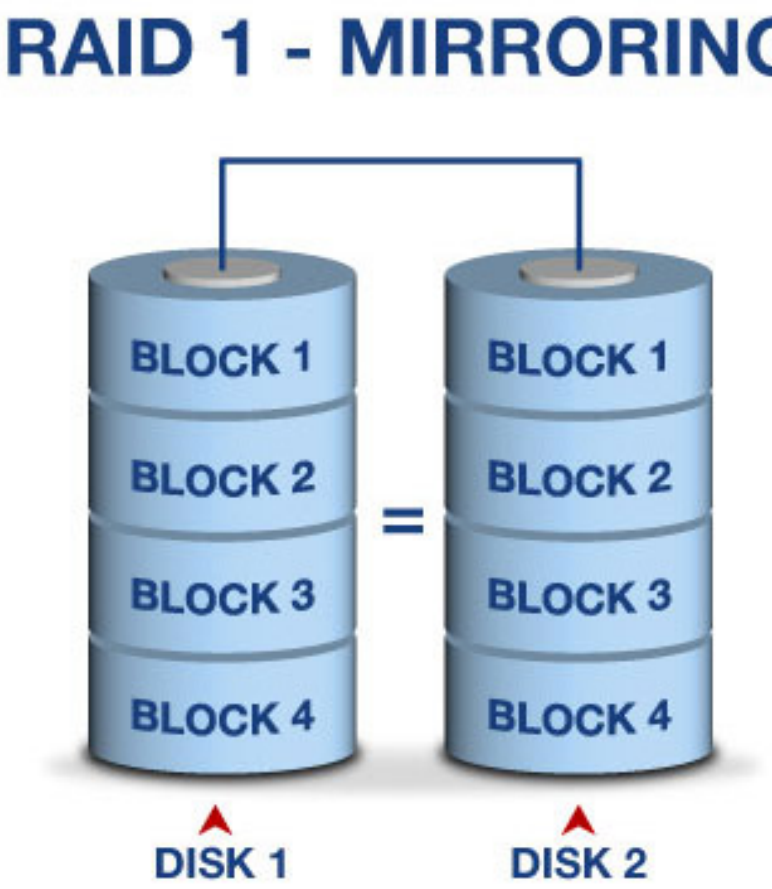
Types of Clustering Solutions

- Hardware redundancy
- SAN solutions
- NAS boxes
- External hard drives or JBODS
- **Software Solutions**

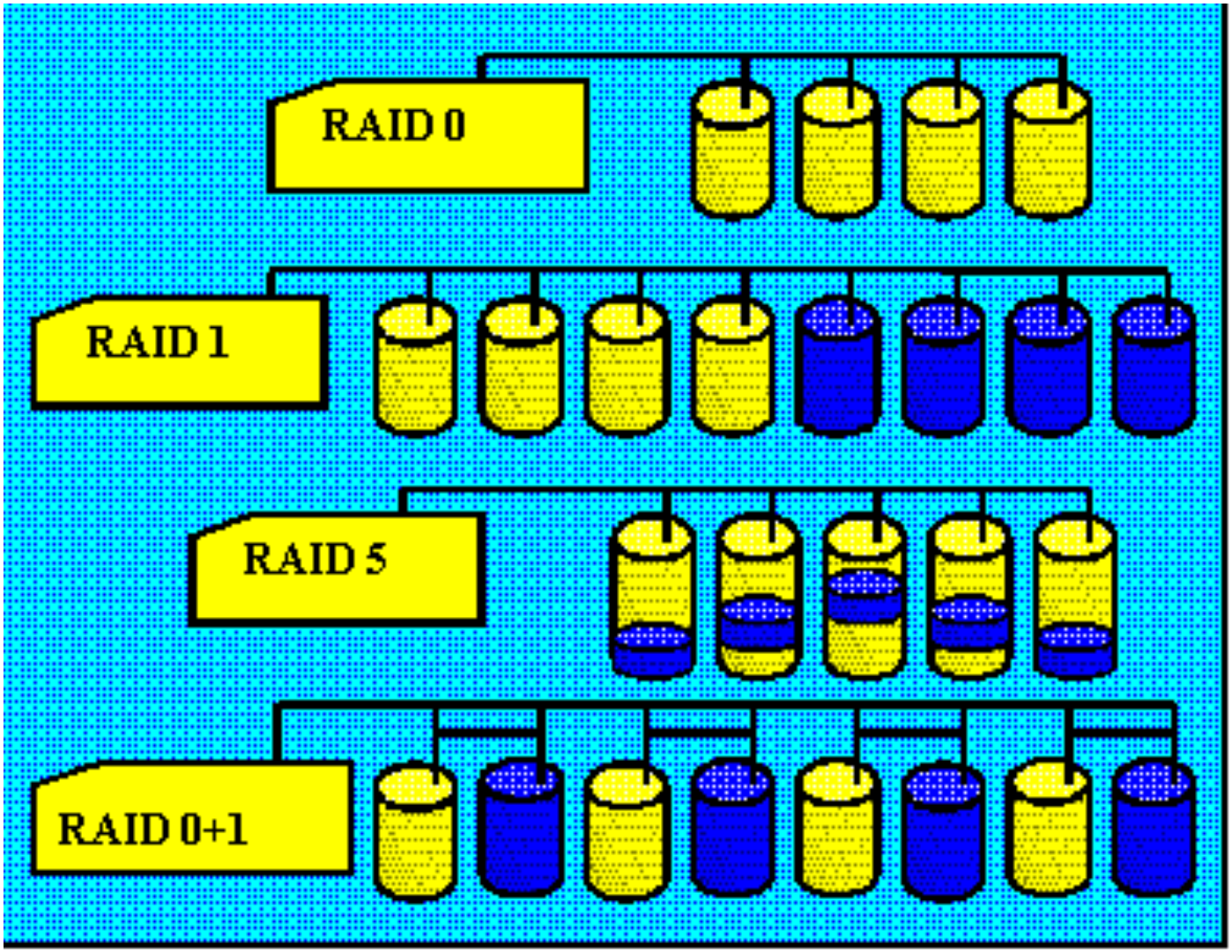
Recovery Time/Point Objectives



What is Raid? Is it enough?



RAID



Microsoft Library http://msdn.microsoft.com/en-us/library/aa226166(v=sql.70).aspx#sql7perftune_moreinfo

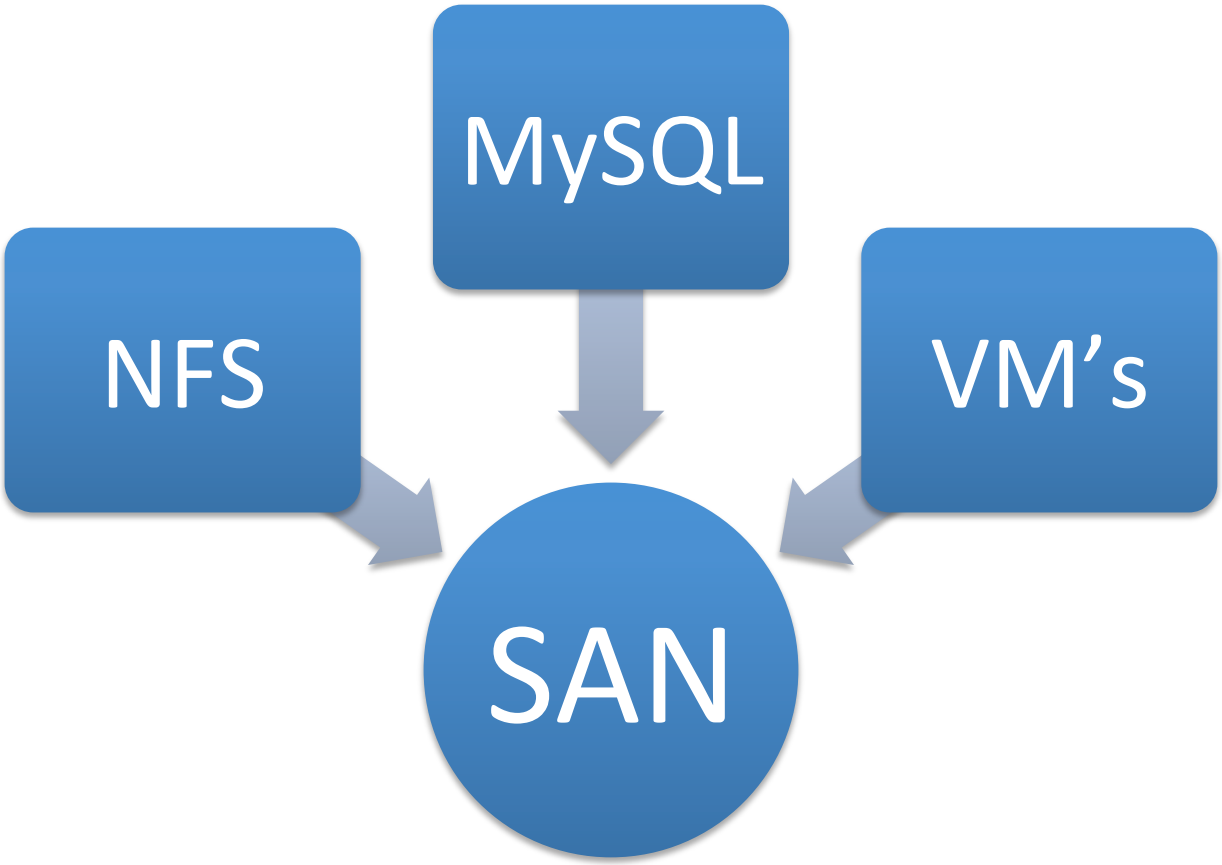
What Could Go Wrong

- Your shiny new hardware will fail
- Single points of failures are dangerous
- Dropped alerts
- Internet outage
- Power outage

SAN/NAS

- Easy to implement - high cost per TB
- Large SLAs - quality of technicians
- Management via GUI
- Scalable - with the right packages
- SAN maintenance - learning curve
- Off site replication is expensive

Single Point of Failure



Pitfalls

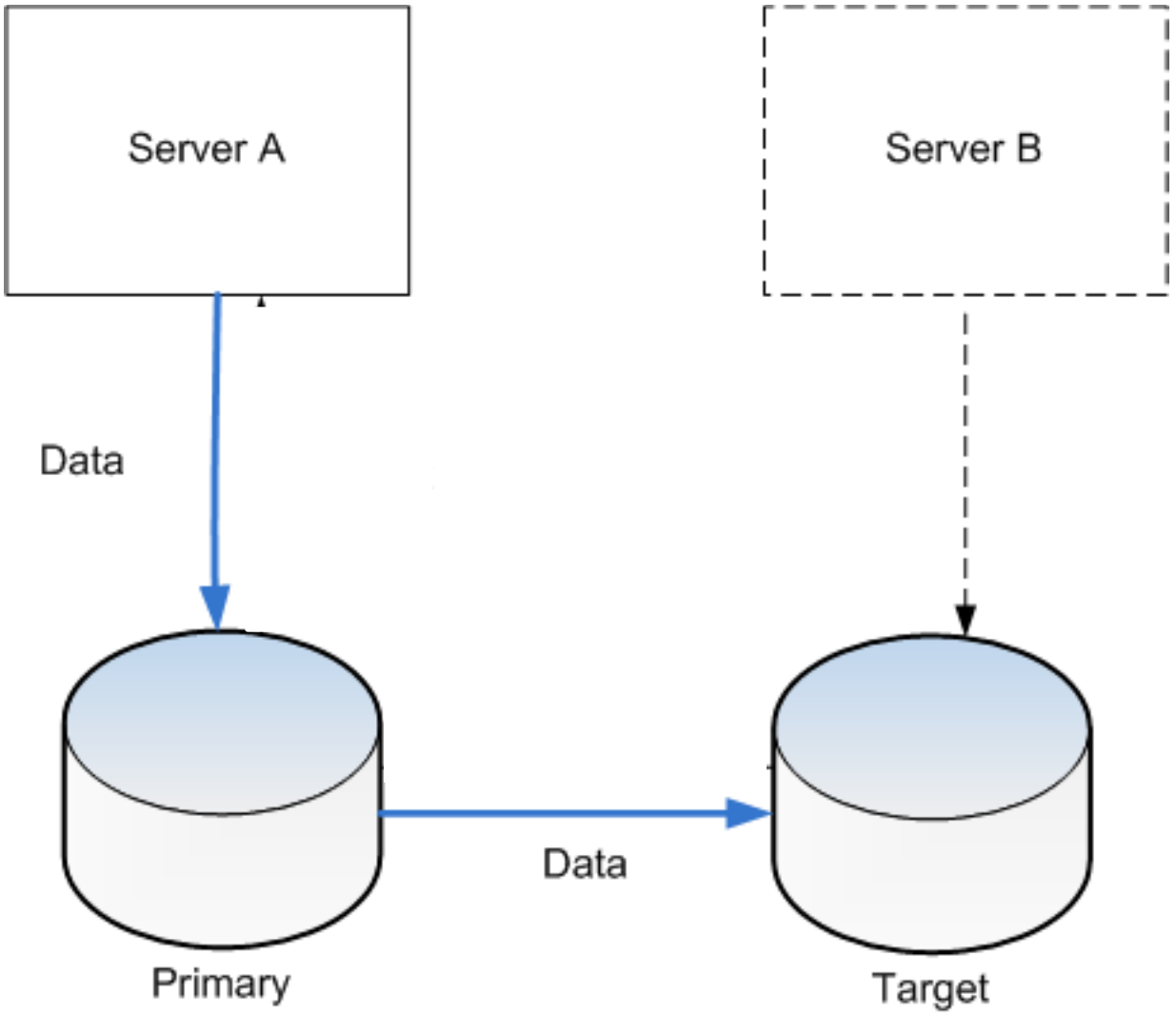
- High initial and ongoing costs
- Vender lock in is required
- Ongoing worry of voiding the warrantee
- Maintenance is tricky and ongoing
- It is a black box, typically Solaris based
- Cannot add or remove features
- It is still a single point of failure

Software Only Solutions

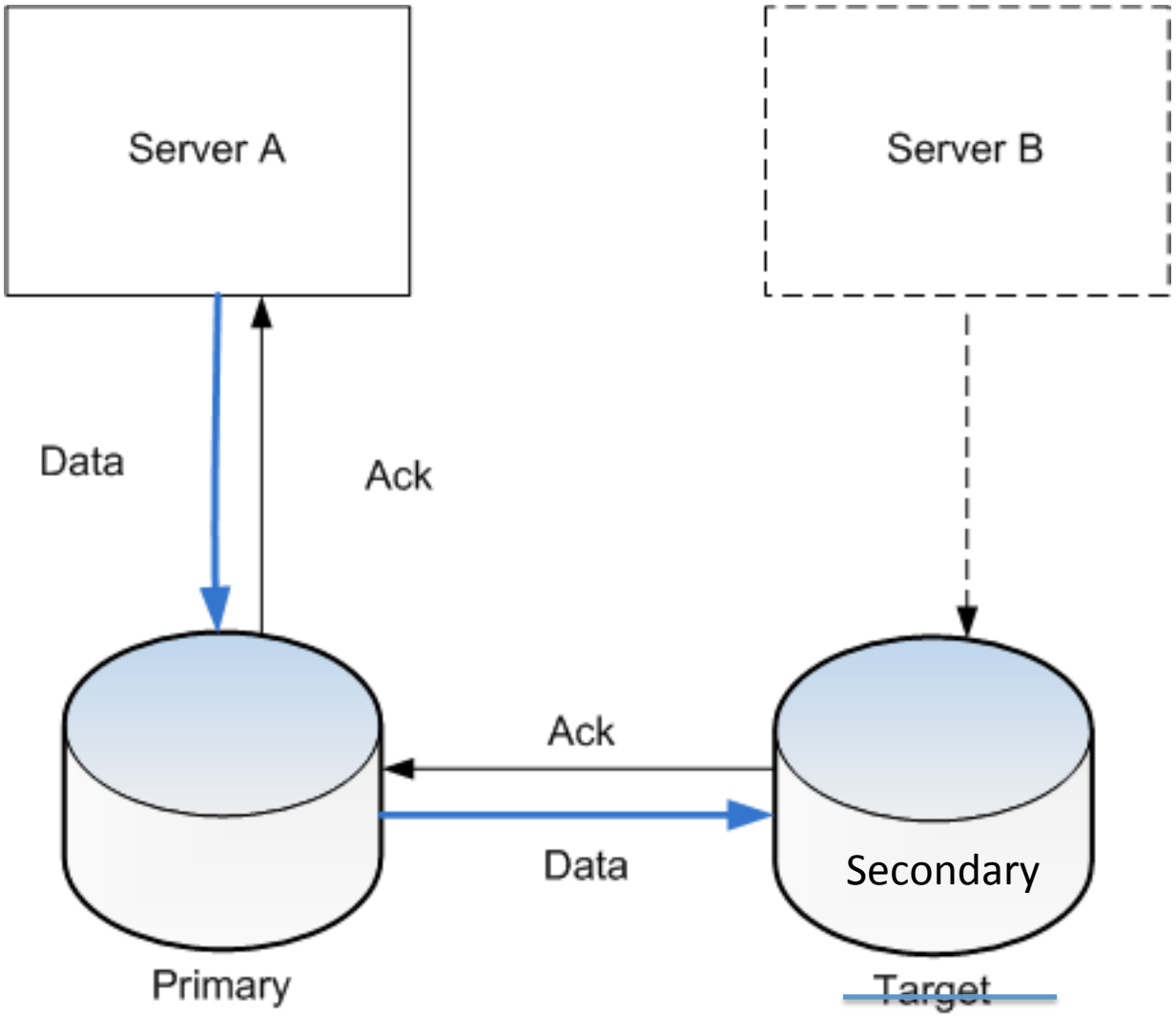
Things to look for:

- Synchronous or Asynchronous replication
- Stability / maturity
- Time to recovery
- Chance of data loss
- Onsite / offsite
- Is it real time (live) or snap shots

Asynchronous Architecture



Synchronous Architecture



Layer Cake of Replication

- Virtualization
- Application
- File system
- Object store
- Block layer

DRBD

http://images.pinkcakebox.com/cake696.jpg

Cluster Cake Fail

Common Issues / Pitfalls

- File locking
- Network congestion
- Data consistency / data corruption
- High overhead and/or additional CPU cycles
- Asynchronous or even back up based
- Require ongoing licensing and royalties

DRBD

- Completely hardware and application agnostic
- German engineering
- In development since 2001
- Created by LINBIT founder and CEO Phillip Reisner
- DRBD built into the native Linux kernel as of 2.6.33
- Ships in all major Linux distributions
- Does not void RHEL or Oracle support

DRBD Users

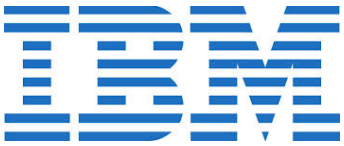

Google


cisco


ING


PORSCHE
STUTTGART


FUSION-IO


IBM


adidas


SIEMENS


Yale
UNIVERSITY


UNITED STATES
NAVY


Mercedes-Benz


DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA


ORACLE


FUJITSU


POLICE.UK


PURDUE
UNIVERSITY

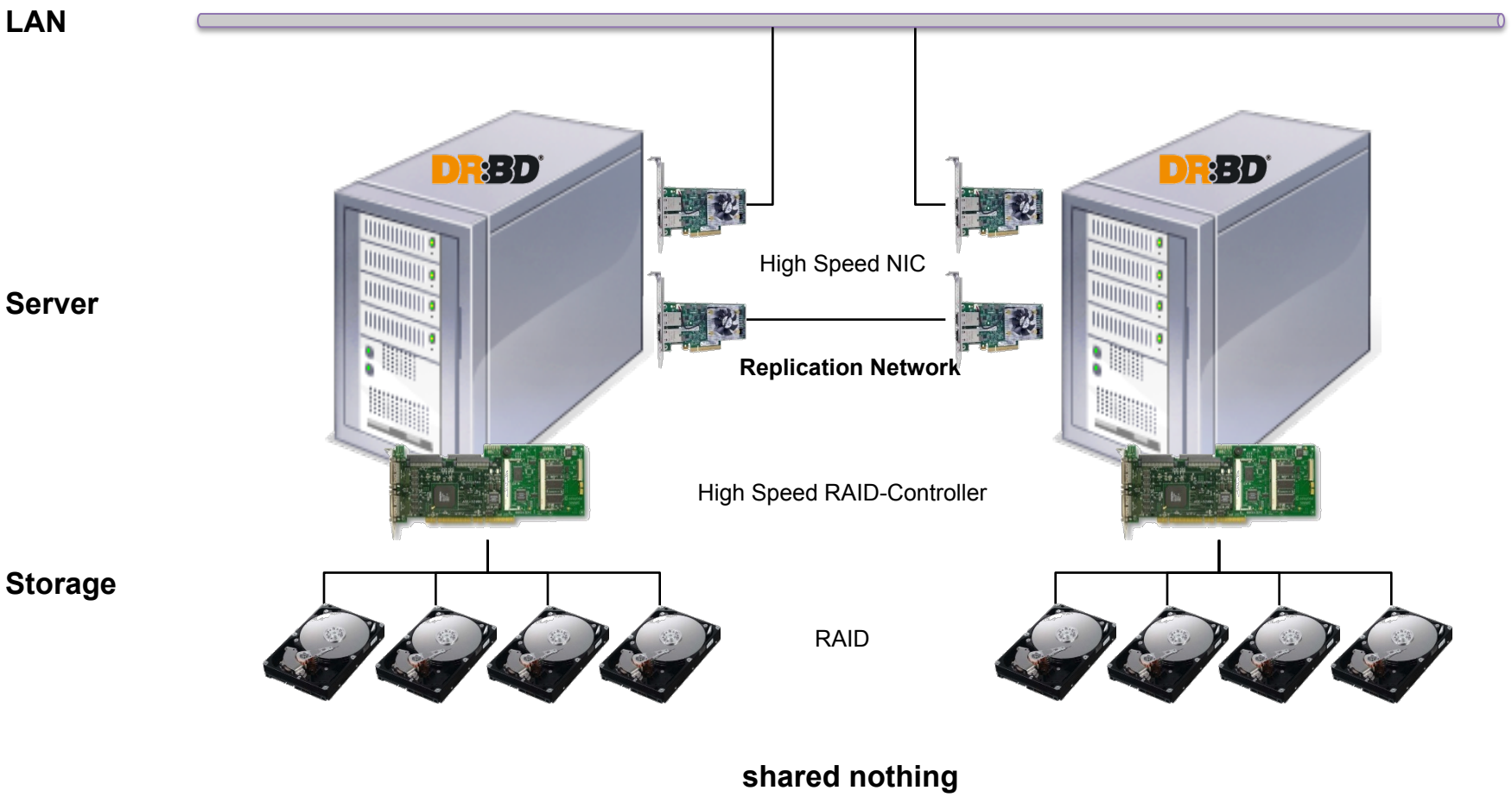

EVERNOTE


T · · Mobile ·

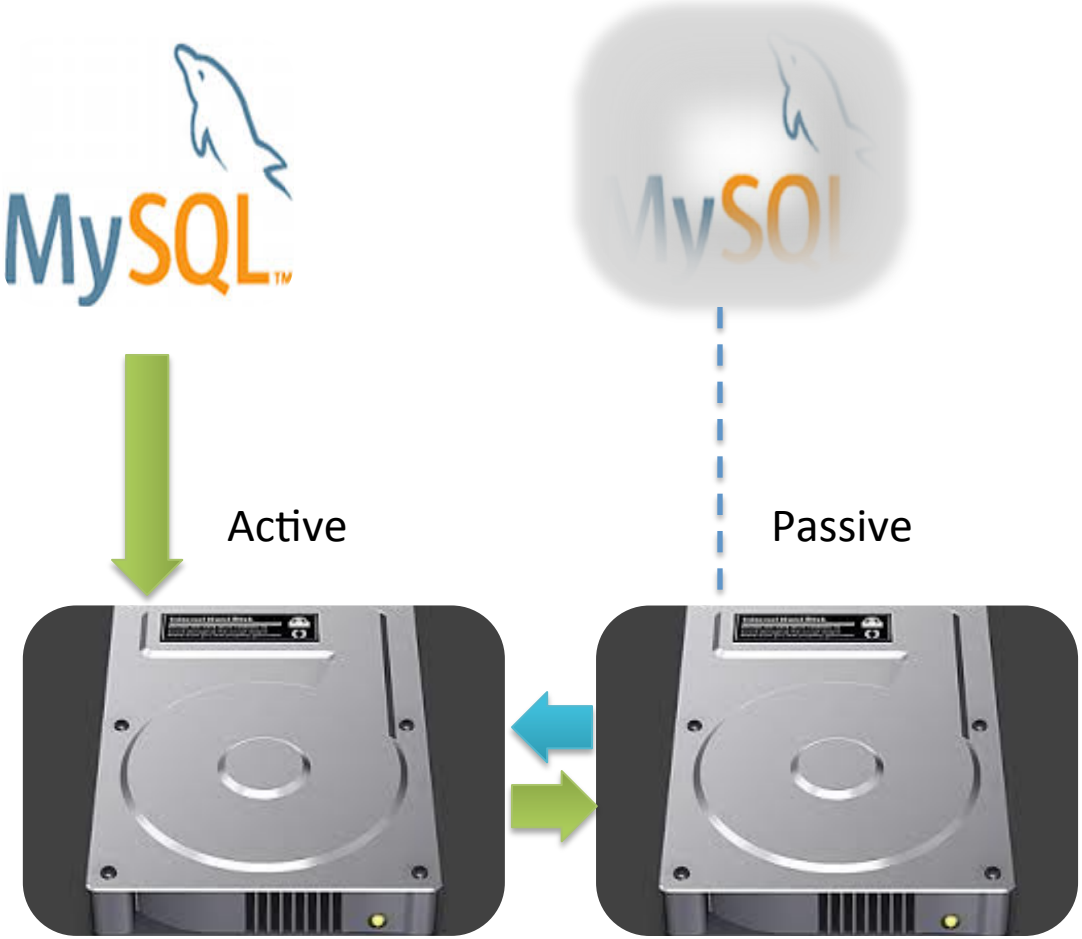

Coca-Cola


KNAPP

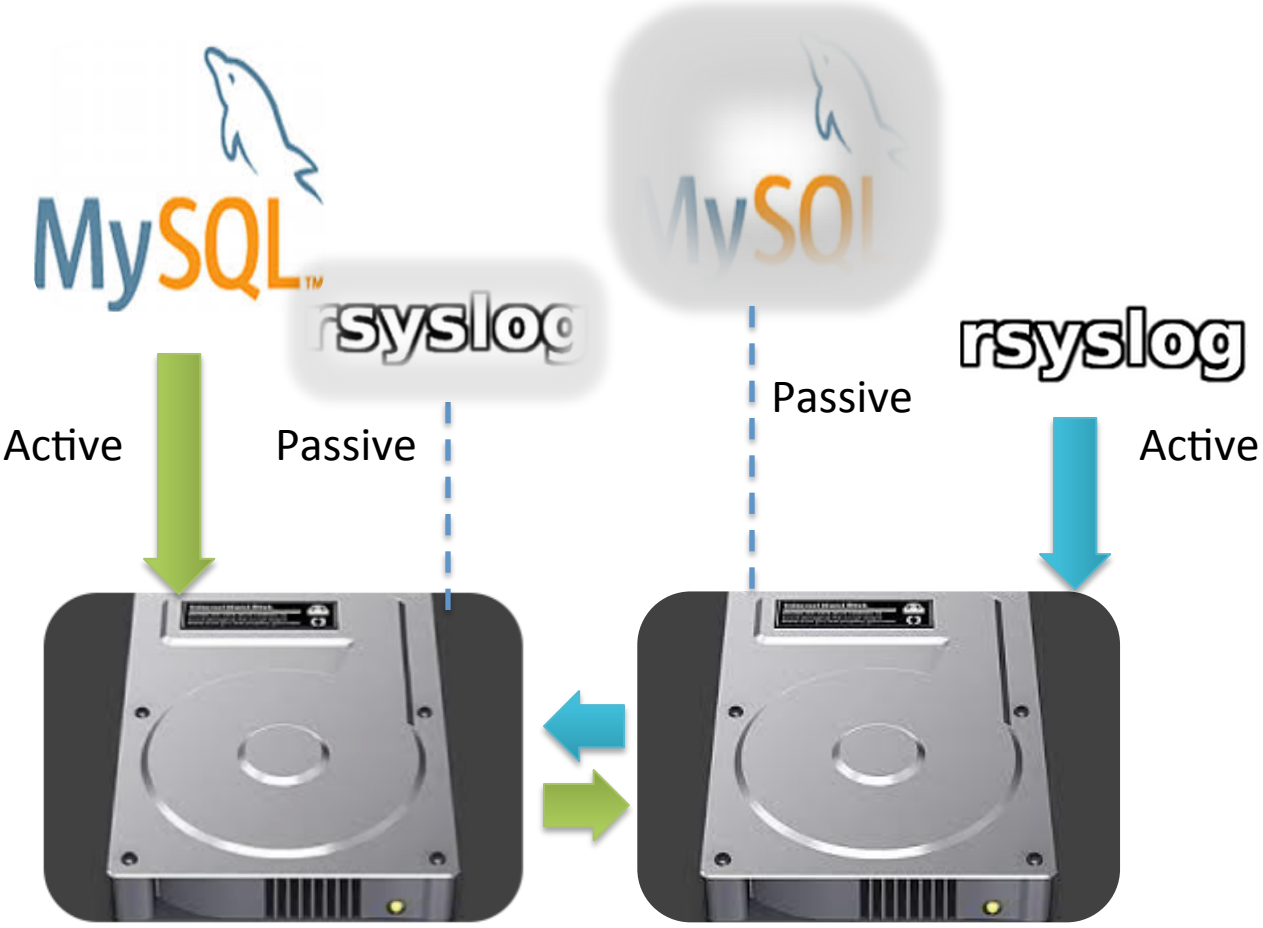
A DRBD Cluster Stack



Fully Redundant System



Fully Redundant System



MySQL.com Ausweb.com

Heartbeat/Corosync: The Comm Layer

- These are the communication tools of the cluster
- “Are you dead?”
- “Are you alive?”
- Heartbeat is seasoned and stable (reliability = HA)
- Corosync is newer and under development

Pacemaker

The Linux Cluster Resource Manager

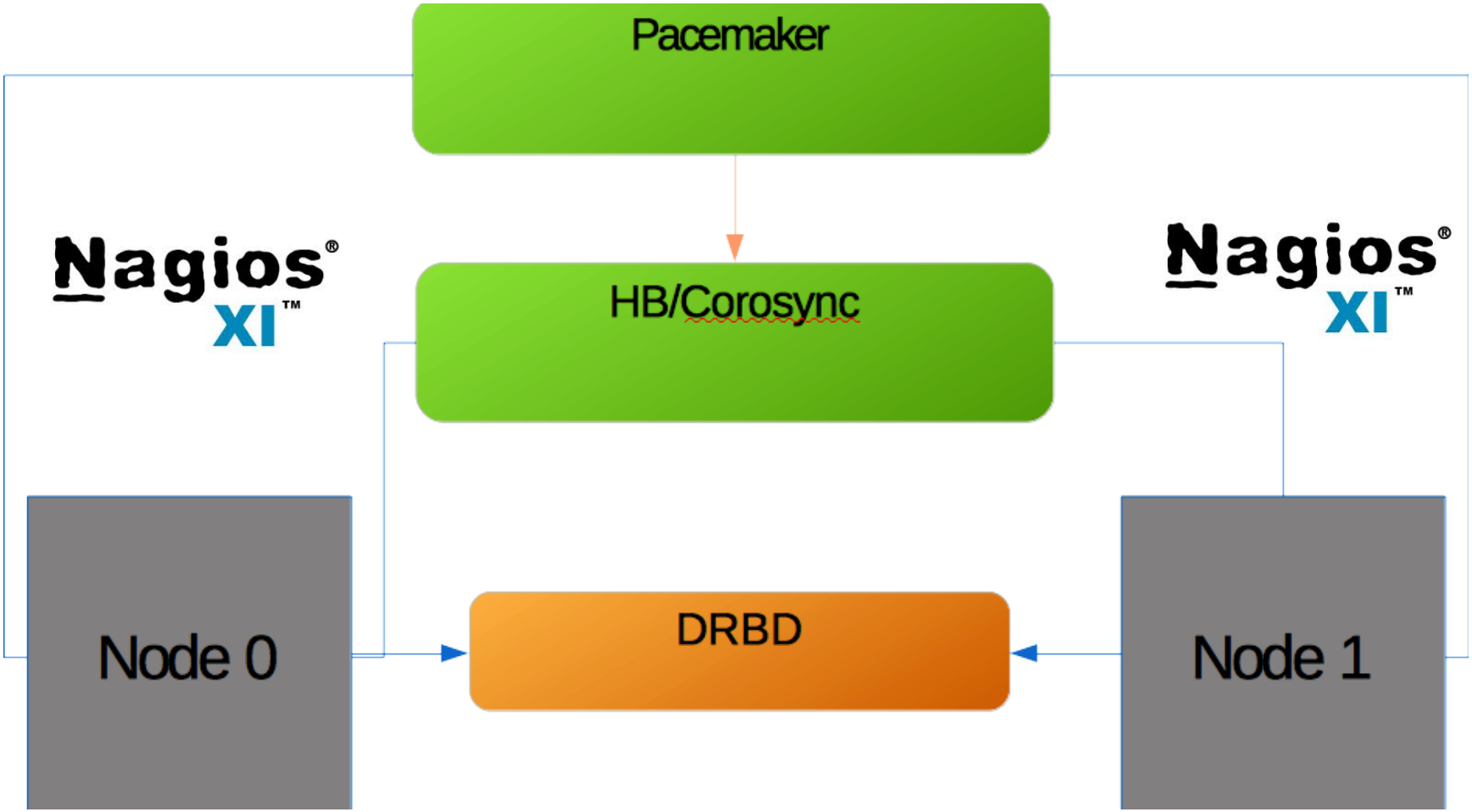
- The powerful and bossy cluster manager
- Manages all aspects of system
- Decides who is alive and primary
- Well known
- Widely deployed
- Does not require applications have specific plugins

Pacemaker : Sleep All Night

- It lets you sleep though the night even if there's a failure.
- Highly Configurable
- Used with a number of clustering tools / File Systems
- Very powerful if done well

  Disastrous if done wrong

Linux HA Stack



Disaster Recovery / Offsite Replication

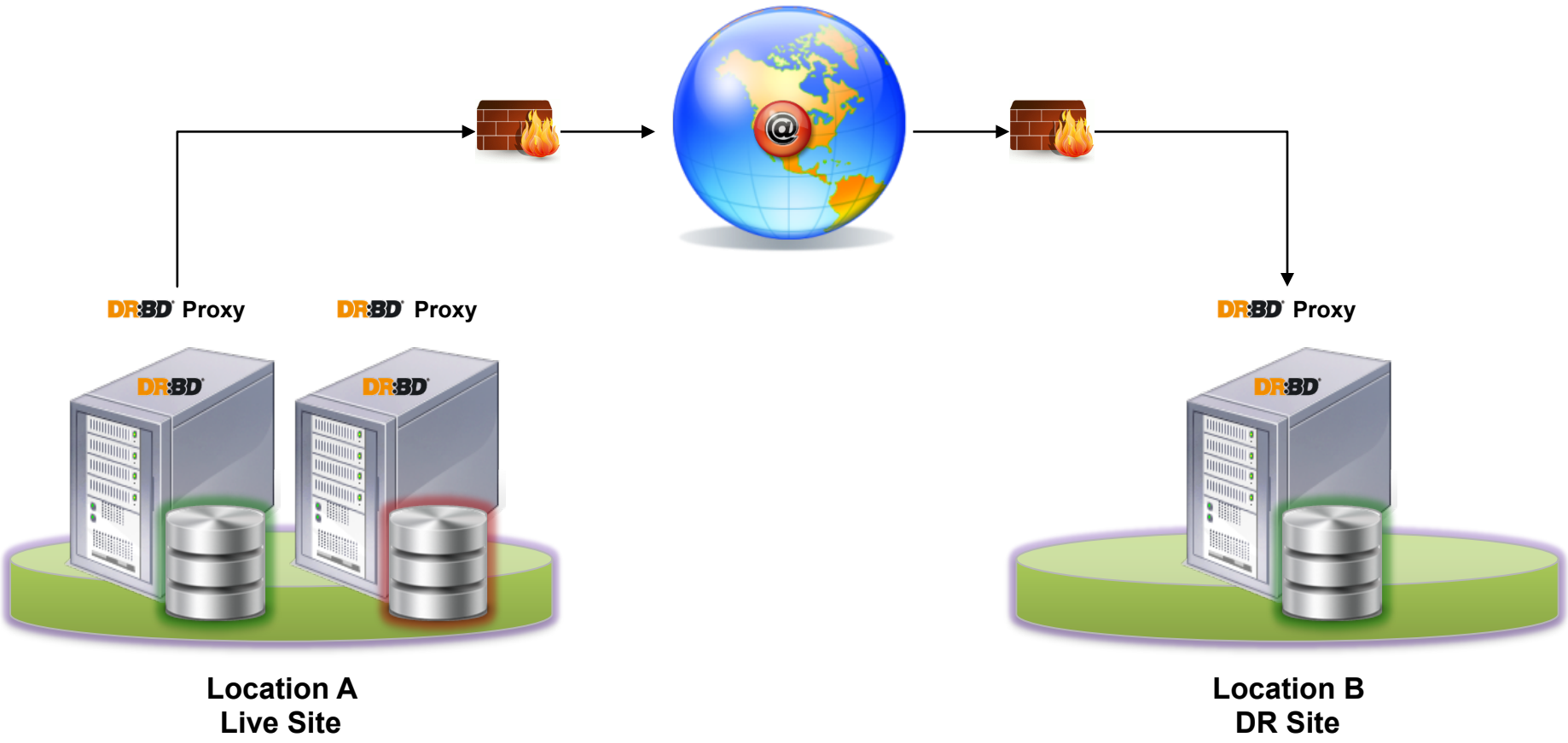
- True Disaster Recovery happens live
- Interval based snapshots no longer meet todays SLA requirements
- DRBD does real-time replication on-site and off-site
- DRBD Proxy tool mitigates throughput constraints and latency- highly configurable

Real-time Disaster Recovery

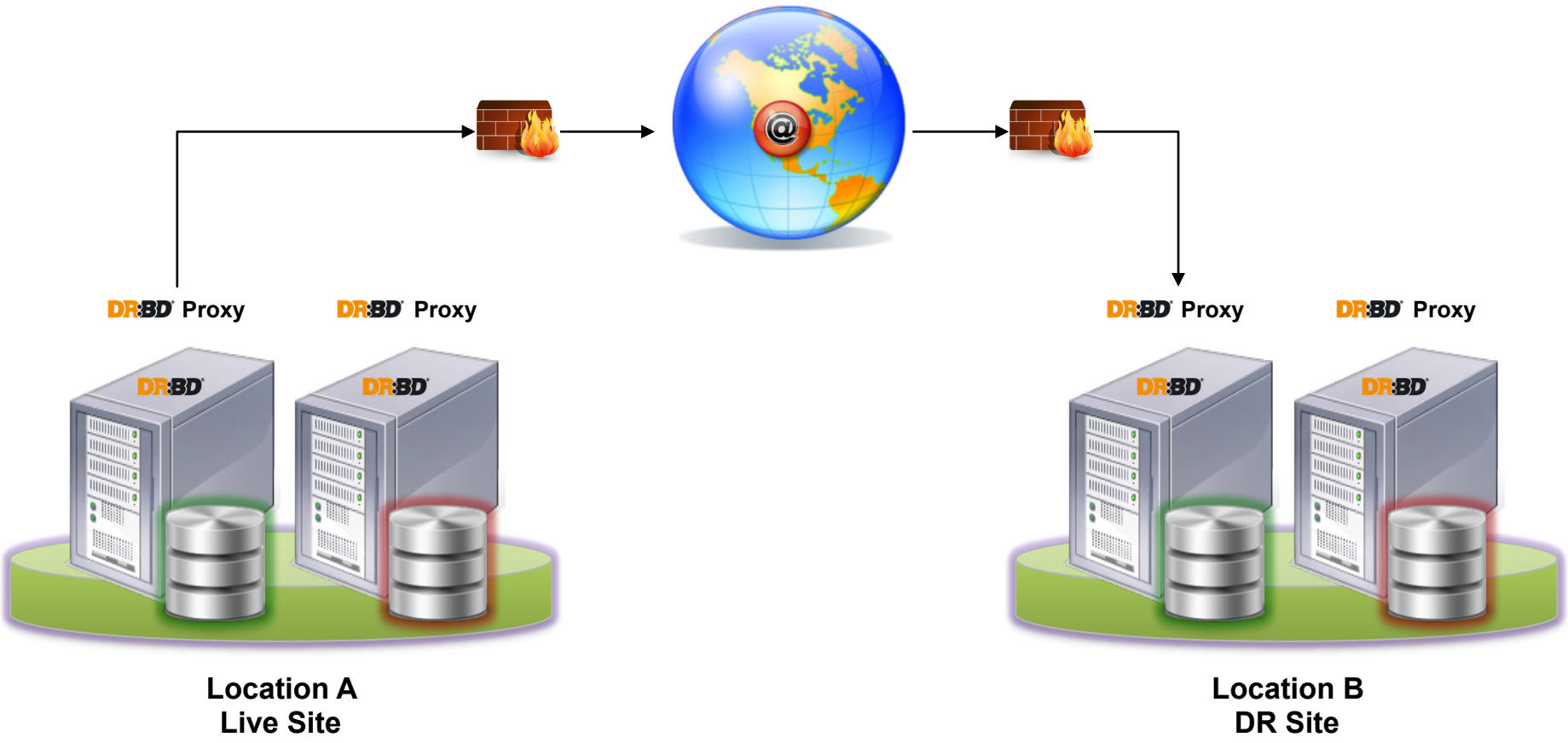
Scaling DRBD

- DRBD Proxy is typically done in 3 node configurations.
- Extremely configurable
- Proxy mitigates bandwidth constraints and latency
- Can replicate across 4 machines even across distances

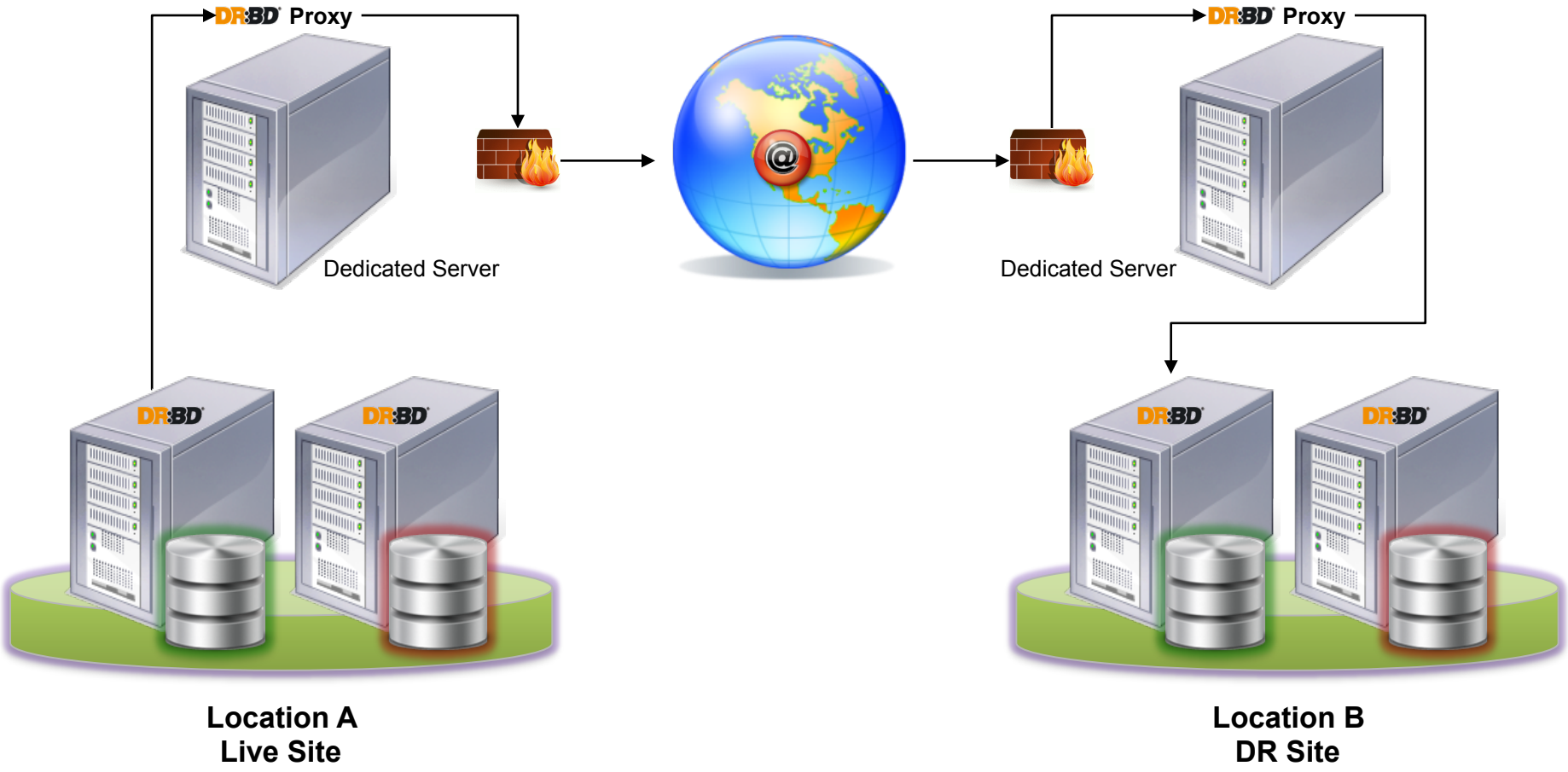
3 node HA / DR


DR:BD Proxy
DR:BD Proxy
DR:BD Proxy
DR:BD
DR:BD
DR:BD
Location A
Live Site
Location B
DR Site

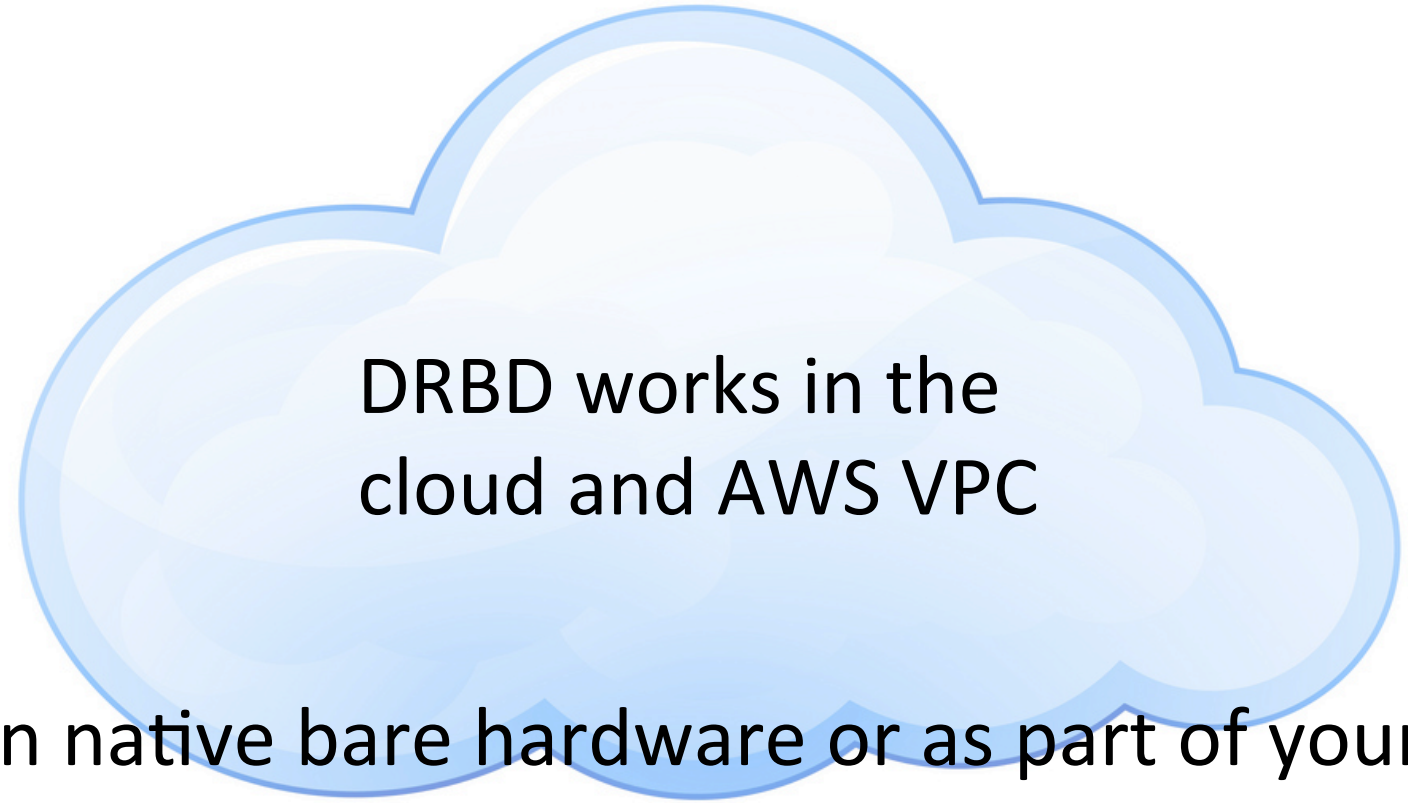
4 node DR + Active-Active HA



Dedicated Proxy-Many Resources



How to apply this in your cloud



On native bare hardware or as part of your hardware or software appliance

DRBD can be used as backing storage for ISCSI

http://www.gamesparks.com/wp-content/uploads/2013/07/the-cloud.jpg

HA with Nagios!

- Filesystem (which has many symlinks in it)
- MySQL
- PostgreSQL
- Crond
- Ndo2db
- The Nagios application itself
- A Virtual IP

Q+A

Jeremy Rust

Jeremy@linbit.com

@NerdHacker

877-DRBD247

www.linkedin.com/in/RustJeremy

DRBD.org

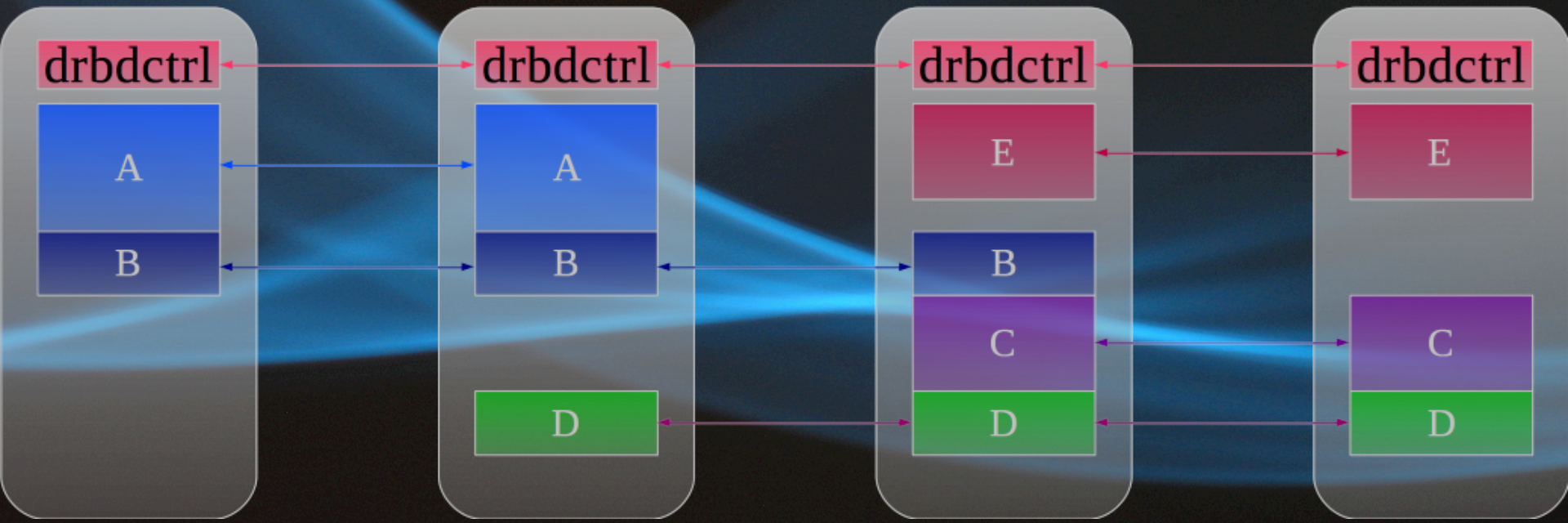
Linbit.com

Linux-HA.org

DRBD 9 the future

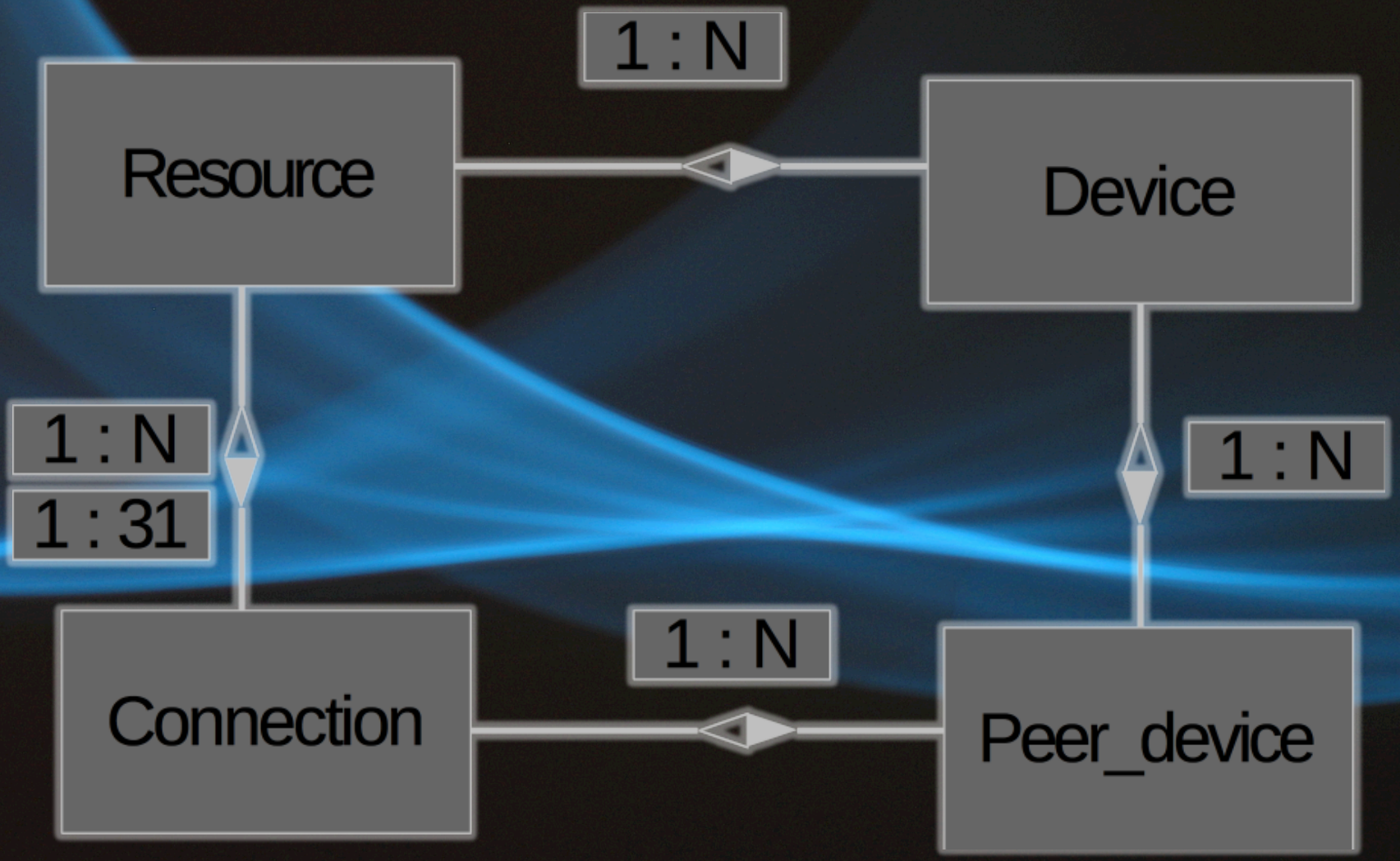
control volume (Manage) – replicated across all nodes

Colors are 5 different and separate automatically managed and replicated volumes



DRBD 8 Branch build structure

DRBD 9 Branch build structure



2 Full redundant systems