

5th TF-Storage Meeting

Subject: DRBD

Author: C.L.N. den Besten (Chris)



Presentation contents

- Part 1 : Introduction on what DRBD is.
- Part 2 : Usage scenario's
- Part 3 : Performance tests
- Part 4 : Future

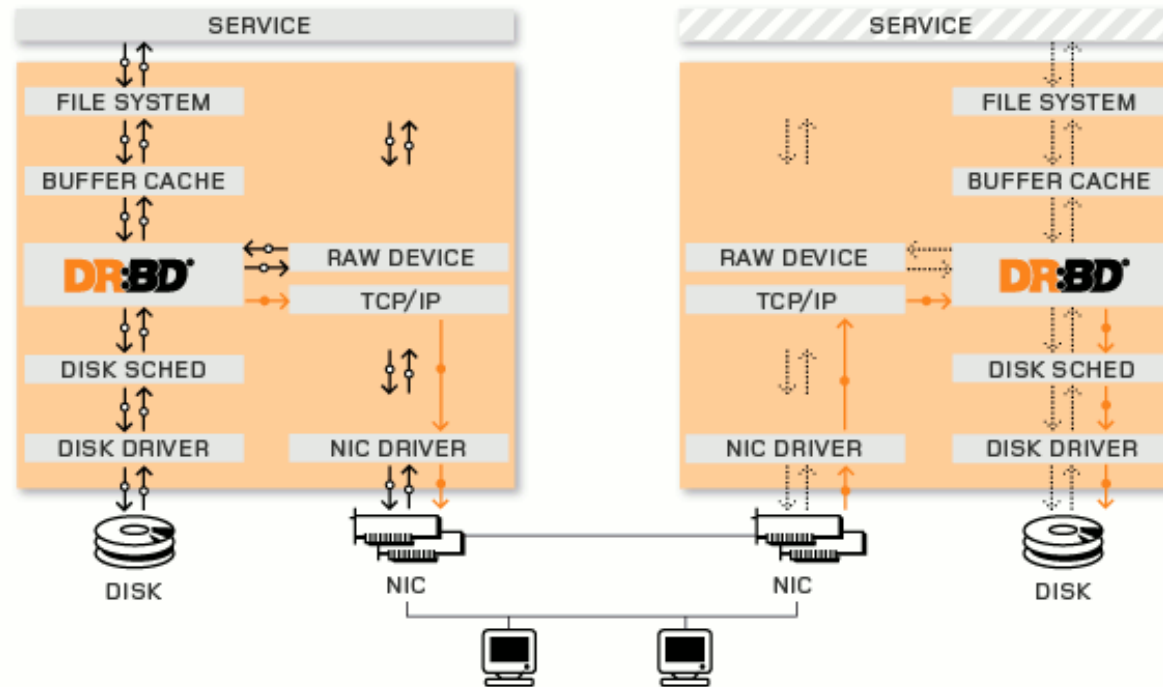


Part 1 : Introduction

“Distributed Replicated Block Device”



Part 1 : Introduction





Part 1 : Introduction

- Size limitation (16TiB per device, with 512 MiB meta)
- Replication modes
 - A = Async, written to local disk
 - B = Async, received by remote peer tcp-buffer
 - C = Synchronous, written on remote disk
- Throughput limitations (read / write)
 - 1 GiG Ethernet (or 2 GiG Ethernet in RR)
 - 10 GiG Ethernet
 - Dolphin SuperSockets



Part 2 : Usage scenario's

- HA filesystem for any application or service
- HA filesystem with both sides writable (GFS / OCFS2)
- Xen guest 'backend storage' (live migration)
- Offsite backup (optionally re-syncing once a day)
- Offsite backup with onsite mirror (drbd stacking)
- Round the clock



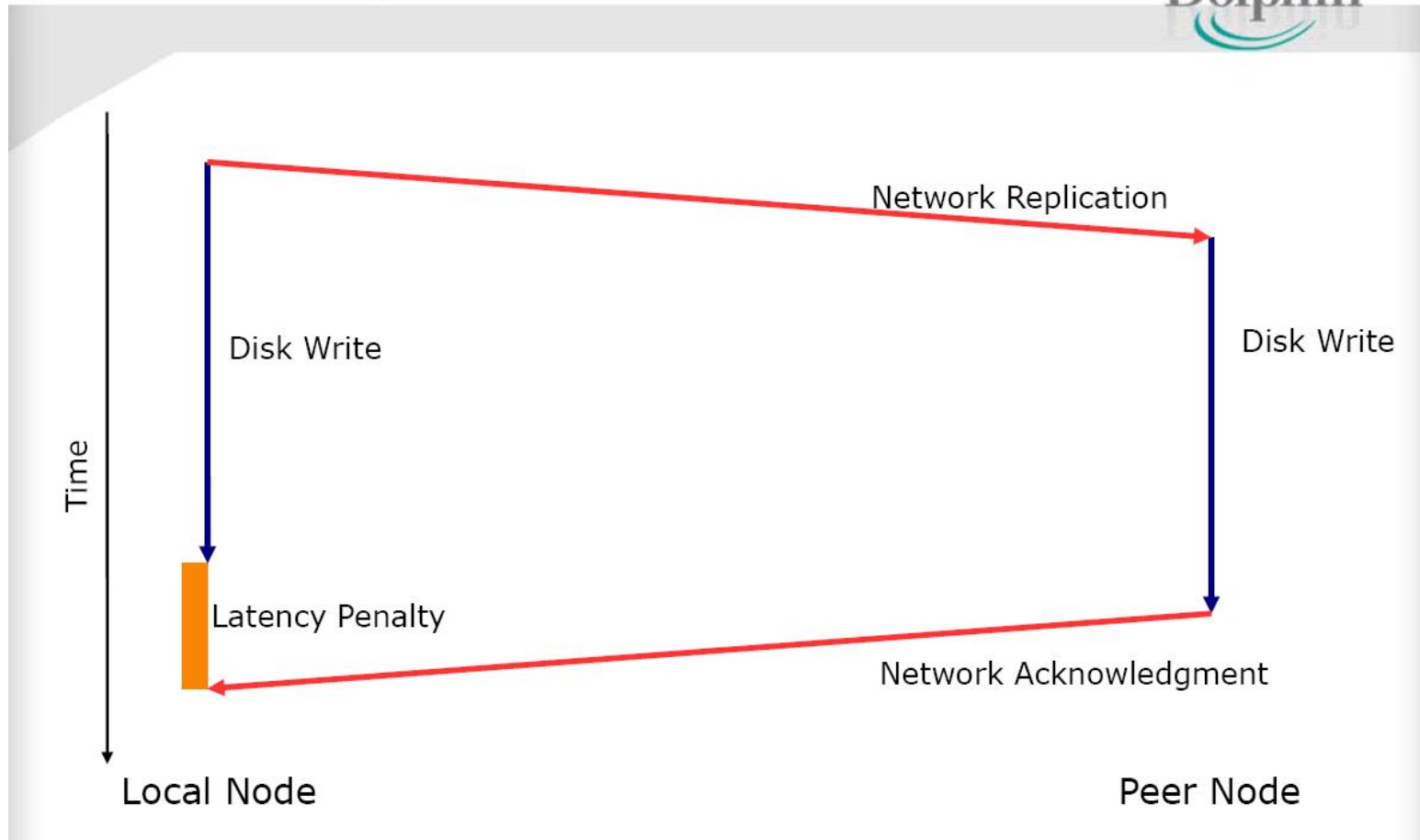
Part 3 : Performance tests

- 1Gb Ethernet (local)
- 20Gb Dolphin Super Sockets (local)
- 10Gb Ethernet -> (+/-5Gb) Optical Lightpath (remote)



Part 3 : Performance tests

Write Latency Considerations





Part 3 : Performance tests

- 1Gb Ethernet (local)

Native SAS (8 disk raid1+0): 350Mb/s

Navite SSD (2 disk raid0) : 165Mb/s

Write Sequentially (Protocol B: async)

SAS : 120Mb/s

SSD : 120Mb/s

Write Sequentially (Protocol C: synchronous)

SAS : 120Mb/s

SSD : 120Mb/s



Part 3 : Performance tests

- 10 Gb Dolphin SuperSockets (local)
Native SAS (8 disk raid1+0): 350Mb/s
Navite SSD (2 disk raid0) : 165Mb/s

Write Sequentially (Protocol B: async)

SAS : 200Mb/s

SSD : 165Mb/s

Write Sequentially (Protocol C: synchronous)

SAS : 256Mb/s (?)

SSD : 165Mb/s



Part 3 : Performance tests

- 10Gb Ethernet -> (+/-5Gb) Optical Lightpath (remote)
Native SAS (8 disk raid1+0): 350Mb/s
Navite SSD (2 disk raid0) : 165Mb/s

Write Sequentially (Protocol B: async)

SAS : 185Mb/s

SSD : 162Mb/s

Write Sequentially (Protocol C: synchronous)

SAS : 220Mb/s (?)

SSD : 162Mb/s



Part 4 : Future

- The future is now ! (since 2.6.33)
- Using “generic” linux dm-replicator
- Order aware replication (meta on raid1, data on raid5/6)
- Adding activity log (replicate to more then 1 node)
- More then two primary's
- For us : Test more tunable parameters ...



Questions ?
