

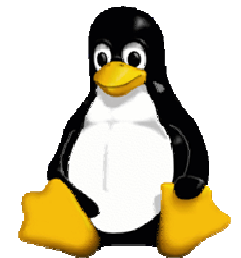
Linux High Availability

mit Heartbeat und DRBD

Werner.Fischer@wefi.net

Christoph.Mitasch@in4.at

IBM Linuxkongress, FH Hagenberg, 26.06.2003



Vorstellung

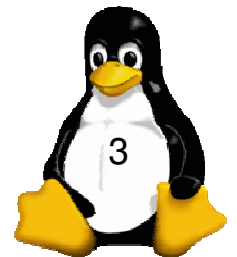
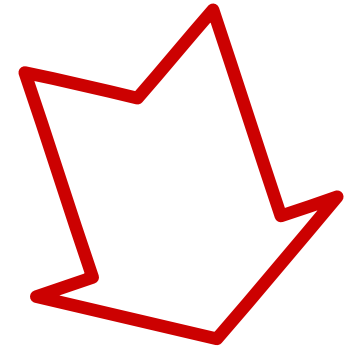


- Werner Fischer, 6.Semester
- Christoph Mitasch, 6.Semester
- Projekt im Wintersemester 2002
- Studiengang
Computer und Mediensicherheit
an der Fachhochschule Hagenberg



Ziele

- Serverdienste unter Linux hoch verfügbar machen
- Tatsächliche Implementierung eines Linux HA Clusters
- Einsatz von OpenSource Software
- kostengünstige Hardware



Was ist High Availability?

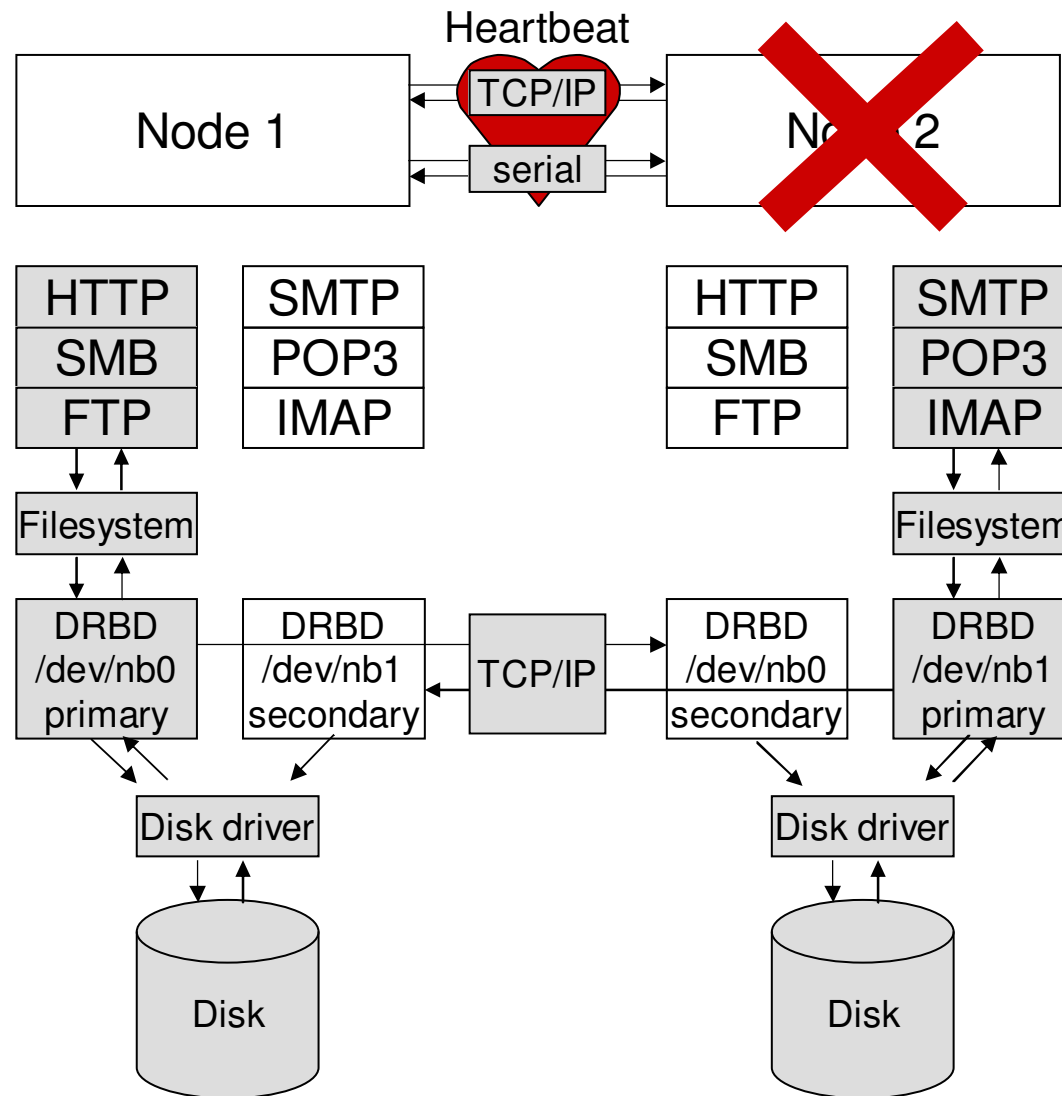


- Hochverfügbarkeit

| Verfügbarkeit | erlaubte Ausfallszeit pro Jahr |
|---------------|--------------------------------|
| 99 % | 88 h |
| 99,9 % | 9 h |
| 99,99 % | 52 min |
| 99,999 % | 5 min |
| 99,9999 % | 31 sec |

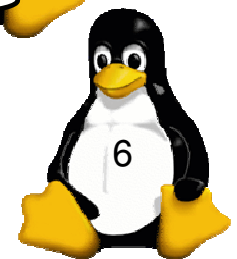
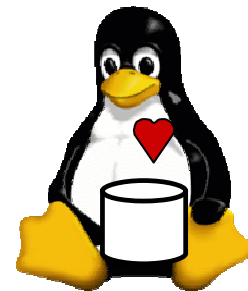
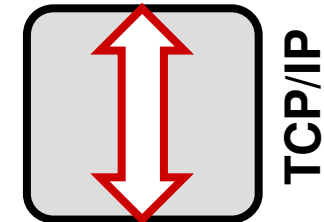


Funktionsweise



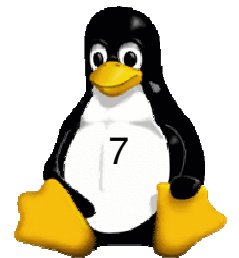
DRBD

- **Distributed Replicated Block Device**
- von Philipp Reisner (Österreich) entwickelt
- zur Zeit 0.6.4 aktuell
- GPL
- Kernelmodul
- Block Device



DRBD

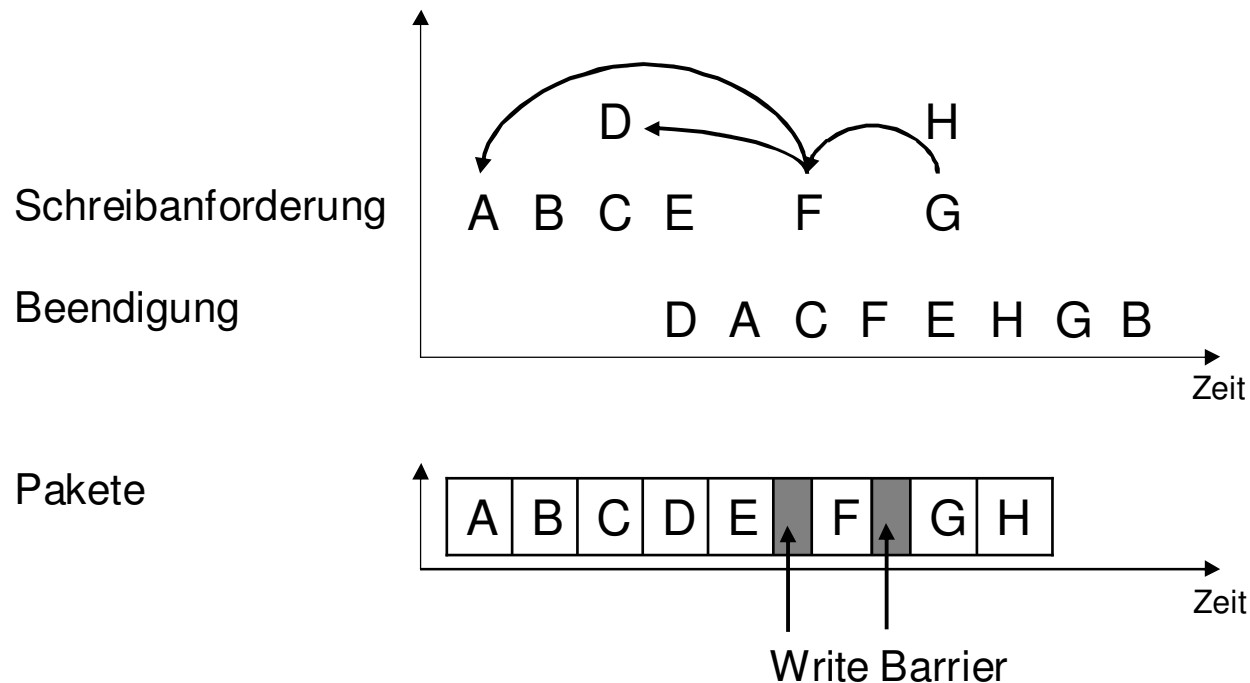
- 2 Zustände: primary, secondary
- Read/Write Zugriff nur für Primary
- 3 Protokolle:
 - **A**: OK, wenn Block auf lokaler Disk und zu zweitem Node versandt
 - **B**: OK, wenn Block auf lokaler Disk und Empfangsbestätigung von zweitem Node
 - **C**: OK, wenn Block auf lokaler Disk und Block auf Disk vom zweitem Node



DRBD



- Parallelisierung von Schreiboperationen über Netzwerk
- Write Barriers



DRBD

- Metadaten: Counter um den Node mit den aktuellen Daten zu finden
- Synchronisation
 - Quick Sync
 - Full Sync
- Datendurchsatz
 - Limitierung lediglich durch Netzerk- bzw. Diskperformance (seit 0.6.1)
 - Unsere Tests: ca. 10 MB/s
 - Laut drbd-devel: bis zu 40 MB/s erreicht
- /proc/drbd

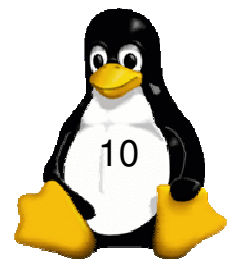
```
version: 0.6.4 (api:61/proto:62)
0: cs:Connected st:Primary/Secondary ns:12036 nr:2860 dw:1527 dr:7401
1: cs:Connected st:Secondary/Primary ns:0 nr:784 dw:784 dr:0 pe:0 ua:0
```



Heartbeat



- von Alan Robertson
- GPL
- Cluster Manager
- überwacht die beiden Nodes über Heartbeat Pakete
- initiiert Failover und Failback
- verwaltet die hochverfügbaren Dienste



Heartbeat Arbeitsweise

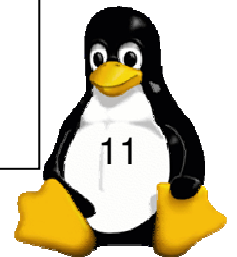


- Heartbeat-Pakete werden ausgetauscht
 - über UDP-Broadcast (Port 694)
 - über seriellen Link

```
>>>
t=status
st=active
src=node1
seq=1cb
hg=1
ts=3ef4a01d
ld=0.33 0.35 0.36 1/45 2363
ttl=3
auth=2 3b264e63
<<<<
```

Bedeutung:

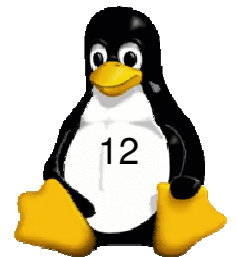
- Message type*
- New status*
- Real originator*
- Sequence number*
- Heartbeat generation no*
- Timestamp*
- Load average*
- Time to live*
- Authentication string*



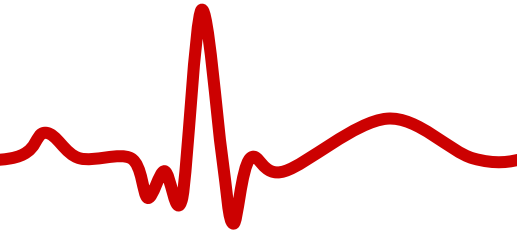
Fehlerfälle



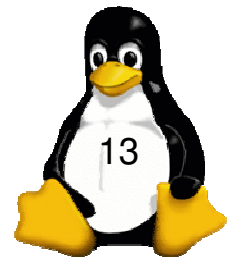
- gleichzeitiger Ausfall beider Nodes
 - starten beide Nodes wieder gleichzeitig, werden die Daten über FullSync aktualisiert
 - startet nur ein Node wieder, ist ein manueller Eingriff notwendig (beim Booten fragt der verbliebene Node, ob er auch primary anstelle des zweiten Nodes werden soll)



Fehlerfälle



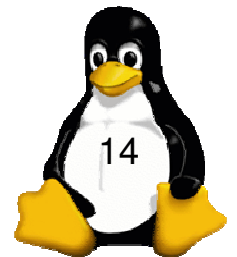
- Ausfall eines Nodes
 - verbleibender Node übernimmt die Dienste
 - beim Starten des ausgefallenen Nodes jeweils ein FullSync und ein QuickSync
- Herunterfahren eines Nodes
 - ein Node wird sauber heruntergefahren (z.B. für Wartungsarbeiten)
 - verbleibender Node übernimmt die Dienste
 - kein FullSync beim Starten (nur zwei QuickSync)



Einschränkungen

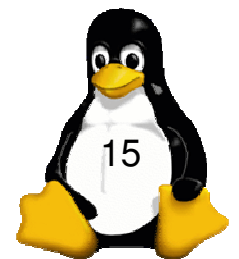


- Lösung unterstützt max. zwei Nodes
- Heartbeat überwacht Hosts, nicht Services
 - dafür ist Network-Monitoring notwendig
- DRBD erlaubt (noch) nicht das gleichzeitige Mounten eines Devices auf beiden Nodes
 - wäre vor allem mit einem GFS von Vorteil



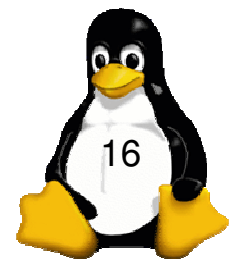
Vorteile

- nur Standard-Hardware
 - Hardware schnell/leicht verfügbar
 - keine zusätzliche Komplexität
- Geringe Kosten
 - eingesparte Lizenzen
 - neue Anwendungsgebiete für Cluster



Vorteile

- beliebige Ausdehnung
- große Verbreitung
 - in über 100 Unternehmen im Einsatz
 - in einem der größten Rechenzentren Japans auf 4000 Rechnern im Einsatz



Zusammenfassung



- Linux-Hochverfügbarkeitssysteme können mit OpenSource-Tools aufgebaut werden
- Einsatz von DRBD erspart shared disk
- DRBD erhöht auch Datensicherheit bei HDD-Ausfällen

„Hochverfügbarkeit muss
nicht teuer sein“

