

```
mysql\
apache2
```

Listing 14

links" ist der Name des ersten Rechners im Cluster. Es folgt die Service-IP-Adresse. Die nächsten Zeilen enthalten die Anweisung, welches DRBD-Gerät wo eingehängt werden soll und welches Dateisystem verwendet wird. Am Ende werden Startskripts für die Dienste gestartet. Jede Zeile repräsentiert ein Skript, das entweder in /etc/ha.d/resource.d/ oder in /etc/init.d zu finden ist, und mit einem Backslash abgeschlossen sein muss, außer der letzten.

Konfiguration der Dienste

Wie bereits erwähnt, verwaltet Heartbeat den Start und Stopp der Services. Normalerweise werden Dienste wie MySQL oder Apache beim Systemstart automatisch gestartet. Das muss durch entsprechende Konfiguration verhindert werden.

Mit auf RPM basierenden Distributionen, wie Opensuse oder Redhat, kann wie folgt vorgegangen werden:

```
##SEL
chkconfig mysql off
chkconfig apache2 off
```

Listing 15

Ensono muss sichergestellt werden, dass DRBD und Heartbeat bei einem Systemstart automatisch gestartet werden, um den Rechner selbstständig wieder in den Rechnerverbund einzufügen.

```
##SEL
chkconfig drbd on
chkconfig heartbeat on
```

Listing 16

Die Konfigurationsdateien der Dienste müssen auf beiden Rechnern synchron gehalten werden. Nach jeder Änderung der Apache- oder MySQL-Konfiguration müssen die entsprechenden Dateien auf dem zweiten Server kopiert werden. Da diese Dateien bei jeder Distribution anders heißen, kann an dieser Stelle nur auf das Handbuch der Distribution verwiesen werden.

Bei Opensuse befinden sich die Apache-Konfigurationsdateien in /etc/apache2/ bzw. /etc/my.c für MySQL. Für die Teilautomatisierung kann das Werkzeug "sync2" [4] gute Dienste leisten.

Testen des HA-Clusters

Bevor ein HA-Cluster in den produktiven Betrieb überführt wird, muss alles auf Herz und Nieren geprüft werden. Es ist wichtig, dass auf beiden Rechnern zuerst DRBD gestartet wird:

```
##SEL
/etc/init.d/drbd start
```

Listing 17

Um eventuell auftretende Fehler einfacher feststellen zu können, wird Heartbeat auf dem Rechner „rechts“ erst nach dem erfolgreichen manuellen Starten der Dienste auf „links“ gestartet:

```
##SEL
links--# /etc/init.d/heartbeat start
```

Listing 18

Nach etwa 120 Sekunden wird der Rechner „links“ feststellen, dass der Partnerrechner „rechts“ tot ist bzw. kein Heartbeat läuft und die Recovery-Maßnahmen einleiten.

Die von Heartbeat ausgeführten Aktionen können mit tail beobachtet werden:

```
##SEL
rechts--# tail -f /var/log/ha-debug
heartbeat11587f1:2006/12/29_16:57:31 info: local resource transition completed.
```

Listing 19

Wenn die Meldung „local resource transition completed“ erscheint, waren der Start von Heartbeat erfolgreich. Sie sollten jetzt in der Lage sein, mit Ihrem Webbrowser auf die Service-IP-Adresse zuzugreifen.

Im Erfolgsfall kann nun auf dem zweiten Rechner „rechts“ Heartbeat ebenfalls gestartet werden (siehe oben). Auf dem Rechner „links“ erscheint in der Logdatei der entsprechende Eintrag, dass „rechts“ als zweiter Rechner zur Verfügung steht.

Als nächster Test steht ein manuelles „Switchover“ an:

```
##SEL
/etc/init.d/heartbeat standby
```

Listing 20

Mit diesem Befehl teilen Sie dem aktuell aktiven Rechner mit, dass dieser nun in den Bereitschaftsmodus wechseln soll. Der sekundäre Rechner übernimmt nun sämtliche Dienste des primären Rechners. Ein Blick in die Logdatei /var/log/ha-debug zum Beobachten schadet nicht. Die Übernahme der Dienste dauert bei einer typischen TYP03-Installation etwa drei bis vier Sekunden.

Als letzter Test kommt nun der „Steckerrest“ zum Zuge. Simulieren Sie einen Hardwareausfall durch das Trennen des aktuell aktiven Rechners vom Stromnetz. Der sekundäre Rechner übernimmt in diesem Fall die Dienste innerhalb der festgelegten 60 Sekunden.

Der schlimmste anzunehmende Fall – „split-brain“

Wenn die beiden Knoten im Rechnerverbund nicht mehr miteinander kommunizieren können, sei es wegen Ausfall der (redundanten) Netzwerkmitteln oder durch eine Softwarestörung, dann entsteht eine „split-brain“-Situation. Das heißt beide Knoten nehmen an, der jeweils andere ist nicht mehr funktionsfähig und beide starten die Services. Da die Service-IP-Adressen ebenfalls auf beiden Knoten konfiguriert werden, sind die Services in der Regel nicht mehr verfügbar. Im schlimmsten Fall haben die beiden Knoten nun unterschiedliche Datenbestände.

Um diesem Szenario vorzubeugen, überwachen sich die beiden Knoten dreifach redundant: über zwei Netzwerkschnittstellen und über die serielle Schnittstelle mit einem Nullmodemkabel.

Sollte der unwahrscheinliche Fall trotzdem eintreten, ist dennoch nicht alles verloren, die Daten können meist vollständig resynchronisiert werden. Dazu muss DRBD lediglich mitgeveilt werden, welcher Knoten die aktuelleren Daten vorhält.

Performance

Die Performance wird durch zwei Faktoren beeinflusst: Die Geschwindigkeit des eingesetzten Disk Subsystems und des internen Netzwerkes. Wenn nicht ein RAID5 Disk System eingesetzt wird, sind der Flaschenhals die Festplatten. DRBD generiert einen gewissen Overhead, daher muss mit einem Performanceverlust beim Schreiben von etwa drei Prozent gerechnet werden.

Für den Katastrophenfall vorsorgen

Optional kann für „Disaster Recovery“ DRBD+ - die kommerzielle Variante von DRBD - eingesetzt werden, um die Daten asynchron auf einen dritten Server an einem zweiten Standort zu replizieren. Dadurch ist die Aufnahme des produktiven Betriebs bei Verlust des kompletten Rechenzentrums durch Brand, Erdbeben oder ähnlichen Vorkommnissen am zweiten Standort rasch wieder möglich. Außerdem unterstützt DRBD+ bis zu 16TB pro logischer Disk, in DRBD ist die Größe einer einzelnen logischen Disk auf 4TB beschränkt.

Ausblick

Beim Schreiben dieses Tutorials im Januar 2007 ist von DRBD die Version 8 als „Release Candidate“ eins verfügbar. Die Version 8 wird einige neue Funktionen mitbringen, arbeitet aber im Prinzip gleich.

Die neue Konfiguration des Verhaltens bei verschiedenen „split-brain“-Situationen erhöht die Zuverlässigkeit im Problemfall. Die größte Neuerung ist die Unterstützung von Aktiv/Aktiv-Clustern. Dies erlaubt es, mit speziellen Cluster-Dateisystemen wie OCFS2 (Oracle Cluster File System) oder GFS (Global File System) die DRBD-Geräte auf beiden Rechnern gleichzeitig einzuhängen und lesend wie auch schreibend darauf zuzugreifen. Da die Komplexität eines Aktiv/Aktiv-Clusters ungleich höher ist, müssen vor dessen Einsatz die Vor- und Nachteile genau abgewägt werden.

Fazit

Durch den Einsatz von rein auf Software basierenden HA-Cluster-Technologien können Unternehmen aller Größen Kostengünstig eine Infrastruktur aufbauen, welche den heutigen Anforderungen von global handelnden Unternehmen gerecht wird. Die Funktionalität entspricht dabei logisch einer klassischen Shared-SCSI- oder Fibre-Channel-Umgebung, ohne jedoch die bekannten Probleme aufzutreten.

Ein Auszug aus der Konferenzschrift der TYP03-Konferenz 2006, auf der der Autor einen Vortrag zu diesem Thema gehalten hat, bietet zusätzliche Informationen [5].

Links und Literatur

- [1] DRBD-Website: <http://www.drbd.org>
- [2] Linux-HA-Website: <http://www.linux-ha.org>
- [3] Downloadseite DRBD: <http://oss.linbit.com/drbd/>
- [4] Cluster-Synchronisierungstool sync2: <http://oss.linbit.com/sync2/>
- [5] Vortrag über HA-Cluster TCON06: <http://www.bit-heads.ch/weber-uns/veroeffentlichungen>

© **Sorkink 1435**



DEBATOR

Luc de Louw ist Linux-Enthusiast seit seinen ersten Versuchen 1994. Seine speziellen Kenntnisse sind neben HA-Technologien und Linux alles rund um Mail- und Webserver. Die ersten „mission critical“ HA-Systeme installierte er als Unix System Engineer im Jahr 2000 in der Airline-Industrie. Im November 2003 gründete er zusammen mit Geschäftspartnern die bit-heads GmbH, ein auf hochverfügbare Webapplikationen und TYP03-basiertes Websitespezialisiertes Unternehmen, und ist dort zuständig für die Server-Infrastruktur und deren Weiterentwicklung. Die bit-heads GmbH ist seit 2006 Partner der LINBIT, dem Hersteller von DRBD.

Wir, Fortrabbit & Control B, suchen ab sofort für unsere Bürogemeinschaft in Berlin einen Freelancer:

Wir bieten:
Büroplatz in netter Gemeinschaft, Berlin Kreuzberg 36, Tisch, Stuhl, Strom, Internet, Red Bull ...

Wir suchen:
einen/e Webdesigner/in, Webentwickler/in der/die uns langfristig projektbezogen unterstützen kann.

Bitte mitbringen:
XHTML, PHP (oder PERL, ROR), AJAX, CSS, Grafikskills ...

Kontakt:
info@fortrabbitt.de

control b
fortrabbitt.de