

File-Server-Cluster mit Open Source

Ursprünglich ein interner Dienstleister der Advance Bank, tritt die X-Cellent GmbH seit vier Jahren als unabhängiger Systemintegrator auf. Kunden aus dem Verlags- und Automobilbereich untermauern die Neupositionierung.

© ComputerPartner, Heft 19/2004, S. 30 - 32

Von ComputerPartner-Redakteur **DR. RONALD WILTSHECK**

Mit der Performance des eigenen File-Servers war Winfried Zittrell, der IT-Verantwortliche bei der Druckerei C. H. Beck, nicht mehr zufrieden. Bei einer Datenmenge von 2 TB in mehreren Millionen einzelnen Dateien hatte der Windows-2000-Server wirklich schwer zu kämpfen. Die starke Datenfragmentierung führte zu Performance-Problemen, die etwa alle zwei Wochen einen kompletten Ausfall des Systems zur Folge hatten.

optimieren. Hier installierte X-Cellent einfach ein System, das die Anfragen auf zwei Webserver aufteilte.

In den darauf folgenden Verhandlungen während der Monate Oktober und November 2003 wurde man sich einig, dass zur Speicherung der Daten auf keinen Fall Enterprise-Storage-Systeme von EMC, Network Appliance oder Hitachi Data Systems in Frage kämen. „Die

Damit im Katastrophenfall wichtige Daten der Druckerei nicht unwiederbringlich verloren gehen, beauftragte C. H. Beck ferner den Dienstleister mit der Implementierung von zwei voneinander unabhängig agierenden Rechenzentren, die zusätzlich auch noch räumlich getrennt sein sollten. So befinden sich die zwei Serverpaare immer noch im gleichen Gebäude, aber in zwei unterschiedlichen Brandabschnitten, wobei sie über eine 2 GBit/s schnelle Leitung miteinander verbunden sind.

Falls wirklich mal eines dieser Rechenzentren komplett ausfallen sollte, schaltet das System auf das andere Serverpaar um, sodass die Anwender innerhalb von etwa 20 Sekunden wieder Zugriff auf ihre Daten bekommen. Diese Aufgabe hat der Cluster-Manager „Heartbeat 1.0.4“ zu erfüllen. Diese Open-Source-Hochverfügbarkeitslösung wird nur in einem solchen Katastrophenfall aktiv, ansonsten werkelt sie ruhig im Hintergrund und misst die System-Performance.

Bei den eigentlichen File-Servern fiel die Wahl auf das Modell HP Proliant DL380G3. Insgesamt bestellte der Kunde vier derartige Blade-Server – je zwei für jedes Rechenzentrum. Hinzu kamen zwei 2-GB-Fibre-Channel-Storage-Systeme vom Typ „HP Storage Works MSA 1000“ (Modular Smart Array), die jeweils zwei Terabyte an Daten fassen, gleichzeitig aber auf eine Kapazität von bis zu 6 TB ausbaufähig sind. „Dies war Wunsch des Kunden, so sichert er seine Investitionen ab“, begründet Majer die Wahl dieses Systems. Beide „Plattensammlungen“ wurden jede für sich über einen 16-Port-Fibre-Channel-

Fortsetzung auf Seite 32 →



HPs „Plattensammlungen“ des Typs MSA 100 fassen bis zu 6 TB an Daten.

Die Master File Table, also der Platz auf der Platte, wo die Informationen über den Ort aller Dateien auf den Harddisks sowie über deren Größe und Zugriffsrechte abgespeichert sind, hatte eine kritische Größe überschritten. Deshalb musste das NTFS-Dateisystem so lange nach den angeforderten Dokumenten suchen, bis die Clients oftmals aufgaben und dem Anwender suggerierten, die gewünschte Dateien seien nicht mehr vorhanden. Daraufhin fiel auch das Windows-Cluster aus, und die redundant angelegten Datensätze konnten nicht mehr synchronisiert werden. Außerdem waren sämtliche Daten bei der Druckerei C. H. Beck in einem Raum untergebracht, wodurch ein Katastrophenfall wie Brand oder Explosion den totalen Verlust sämtlicher Daten verursacht hätte.

Keine teuren Enterprise-Lösungen

Angesichts derartiger Probleme kontaktierte Zittrell die X-Cellent GmbH. Mit dem Münchener Dienstleister hatte die Druckerei bereits gute Erfahrungen in Sachen Linux, Hochverfügbarkeit und Lastverteilung gemacht, als es darum ging, die Zugriffszeiten auf die eigene Website zu

waren dem Kunden schlicht und ergreifend zu teuer“, erinnert sich Stefan Majer, Projektleiter bei der X-Cellent GmbH.

Software-Lizenzkosten sollten ebenfalls vermieden werden. So entschieden sich die Partner schließlich für eine preisgünstigere Hardware von Hewlett-Packard und setzten obendrauf nur auf quelloffene Software. Als Betriebssystem wählte X-Cellent Suse Linux 9 aus, wobei der darin eingepackte Kernel 2.4 durch die Version 2.6 ersetzt wurde. Grund hierfür war der LVM 2 (Logical Volume Manager), der im Gegensatz zum LVM 1 des Kernel 2.4 besser mit größeren Datenmengen umgehen kann. Diese Kernel-Komponente unterteilt eben die gesamte Datenmenge von zwei Terabyte in kleinere, leichter handhabbare Einheiten und passt dies auch während des laufenden Betriebs dynamisch so an, dass der File-Server stets performant bleibt.

Den Windows-XP-Clients in der Druckerei gaukelt das neu installierte System dennoch vor, es sei ein Windows-File-Server. Diese Aufgabe erfüllt normalerweise die quelloffene Software Samba. Im Projekt bei C. H. Beck entschieden sich die Spezialisten von X-Cellent für die Version 3.0 von Samba, die während der Projektlaufzeit gerade frisch herauskam.

„Mit Windows wäre es auf keinen Fall einfacher gewesen.“



Manager Internet-Systems bei X-Cellent, Stefan Majer

Switch von Brocade jeweils transparent mit den zwei Blade-Servern von HP in jedem Rechenzentrum extra verbunden (siehe auch schematische Skizze: Netzwerk-Infrastruktur bei C. H. Beck).

Für die Redundanz des gesamten Datenbestandes bei Beck sorgt ein Open-Source-Paket namens DRBD 0.6.10 (Distributed Replicated Block Device). Es handelt sich hierbei um eine hoch verfügbare quelloffene Cluster-Lösung des österreichischen Systemhauses Linbit. Damit lassen sich Cluster unter Linux realisieren. Im Falle der Druckerei C. H. Beck spiegelt DRBD fortwährend die Festplatten des einen Rechenzentrums in den zweiten Brandabschnitt. Das heißt, jedes Mal, wenn ein ganzer Block auf der Festplatte der einen „Plattensammlung“ abgespeichert wird, tritt DRBD sofort in Aktion und überträgt die geänderten Daten auf den zweiten File-Server im anderen Rechenzentrum. Dies erfolgt über die schon erwähnte 2-GB-Leitung so performant, dass der User nichts davon mitbekommt.

Im Gegensatz zu Heartbeat arbeitet DRBD ständig im Vordergrund, ist also permanent aktiv.

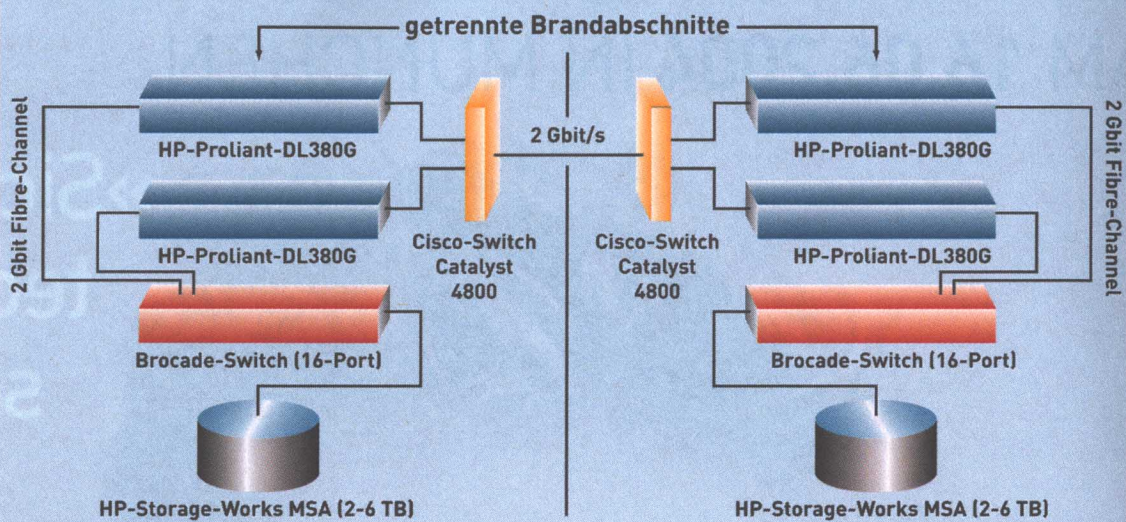
Während des bei C. H. Beck laufenden Projekts hat Linbit seine Lösung auch an die Funktionen von LVM 2 aus dem Linux-Kernel 2.6 optimiert und bietet diese Features nun standardmäßig an. Majer hält DRBD für das derzeit beste Open-Source-Produkt im Umfeld Hochverfügbarkeit.

Im gesamten Projekt bei der Druckerei C. H. Beck wurde kein einziges lizenzpflichtiges Produkt verwendet, die gesamte File-Server-Infrastruktur wird nun von quelloffener Software gesteuert – bis auf den Backup-Client von Veritas, der beim Kunden gesetzt war.

Die Linux-Alternative – Migration auf Windows 2003 – wäre laut Majer nur mit Zusatzprodukten möglich, die Daten auf zwei Partitionen auslagern können. Dann müssten aber nach Ansicht des X-Cellent-Projektleiters teure Storage-Lösungen à la EMC oder HDS zum Einsatz kommen. Da der Dienstleister aber über genügend Linux-Know-how verfügt, rückte man serverseitig von der Microsoft-Plattform ab.

Eine der größten Herausforderungen des Projekts bei C. H. Beck war die initiale Spiegelung des gesamten Datenvolumens von 2 TB

Netzwerk-Infrastruktur bei C. H. Beck



Bei C.H.Beck sind die File-Server doppelt redundant ausgelegt. Sogar nach Komplettausfall eines ganzen Brandabschnitts dauert es weniger als 20 Sekunden, bis das System wieder zur Verfügung steht.

vom ersten Rechenzentrum auf das Backup-System am zweiten Standort. Dieser Vorgang nahm insgesamt etwa zwölf Stunden in Anspruch. Das war aber nicht weiter tragisch, da X-Cellent den Cluster auf der „grünen Wiese“ neu aufbaute; das heißt, das Produktivsystem der Druckerei lief – noch unter Windows 2000 – parallel weiter.

Länger als ursprünglich geplant dauerte auch die Neueinrichtung der Benutzerberechtigungen auf dem Linux-Server. Dies lag zum einen an den bei C. H. Beck gewachsenen Strukturen und zum anderen an der Version 3.0 der Windows-File-Server-Komponente Samba. Damit war eine Eins-zu-eins-Übernahme der vorhandenen Rechte nicht so ohne weiteres möglich. Zuerst mussten die Experten von X-Cellent die Samba-Konfiguration neu erstellen und im Nachhinein die Vergabe von Rechten nachprüfen. Diese Arbeit hat sich aber gelohnt, denn nun befinden sich die Benutzerberechtigungen bei C. H. Beck auf dem neuesten Stand.

Größere Probleme gab es auch bei der Installation der Dual-Gigabit-Adapterkarten: Sie verhielten sich anfangs nicht so, wie X-Cellent das erwartet hatte. Die verwendeten Intel-Boards waren zu diesem Zeitpunkt erst zwei Wochen auf dem Markt, und bis dato gab es noch keinerlei Erfahrungen mit deren Einbau in reinen Linux-Netzwerken. Da konnte der Techniker vom Kartenlieferanten HP auch nicht weiterhelfen.

Man wusste anfänglich nicht, ob die Hardware defekt oder ob die Linux-Treiber nicht auf dem neuesten Stand waren. Nach diversen Tests fand schließlich der Dienstleister heraus, dass es am Netzwerkdesign selbst lag. Nachdem die Spezialisten von X-Cellent dieses entsprechend angepasst hatten, arbeiteten die Dualport-Gigabit-Ethernet-Interfaces einwandfrei. Dabei entdeckte der Dienstleister auch, dass die brandneuen Karten so „intelligent“ waren, Hochverfügbarkeitsmerkmale bereitzustellen, ohne dass sie das System überhaupt angefordert hätte. Auch dies war eine Erfahrung, von der X-Cellent noch heute profitiert.

Performance-Probleme gab es ferner mit dem Fiber-Channel-Sto-

rage: Die Festplatten wurden erst nicht über die gesamte Bandbreite der Verbindung zum Controller angesprochen. Erst eine Rekonfiguration ermöglichte hier die erwarteten hohen Datentransferraten.

Trotz dieser Hindernisse konnte X-Cellent die ursprünglich auf 60 Tage angesetzte Projektdauer deutlich unterschreiten. Nach bereits 45 Tagen war die Linux-basierte Clusterlösung betriebsbereit. Seit Ende Januar 2004 verrichtet sie zur vollen Zufriedenheit des Kunden die an sie gestellten Aufgaben.

Für X-Cellent war dies das erste Storage-Projekt mit dem Open-Source-Produkt DRBD und einem derart großen Datenaufkommen. Obwohl im wahrsten Sinne des Wortes ein „Pilot“, war der Kunde nach acht Wochen Verhandlung von der angestrebten Lösung überzeugt. Hier spielte sicherlich auch die Tatsache eine Rolle, dass man in der Druckerei C. H. Beck für Linux offen war und dort auch einige Kenntnisse auf diesem Gebiet schlummerten. Der Kunde zeigte sich von Anfang an zuversichtlich, was den Ausgang des Projekts betraf. Zwar waren hierzu einige Tests notwendig, doch sie verliefen schlussendlich erfolgreich, und die mutige Entscheidung für reine Open-Source-Software wurde schließlich belohnt.

Meinung des Redakteurs

Das vorliegende Projekt beweist, dass man auch unternehmenskritische Anwendungen quelloffener Software überlassen kann. Die hier geschilderten Problem wären in ähnlicher oder sogar in krasserer Form auch bei einer Windows-Installation vorgekommen. Dies sollte auch der Stadt München Mut machen, die bei ihrer Linux-Migration ebenfalls auf Schwierigkeiten stößt.

Solution Snapshot

Kunde	Druckerei C. H. Beck, www.becksche.de
Problemstellung	Zentraler Windows-2000-File-Server mit 2 TB Daten hatte massive Stabilitätsprobleme, daraus resultierten Datenverluste; Datenfragmentierung führte zu Performance-Problemen; das System fiel des Öfteren komplett aus; keine Katastrophenfall-Auslagerung vorhanden
Lösung	Doppelt ausgelegter Linux-basierender Fileserver-Cluster: Hardware: 4 HP-DL380G3-Blade-Server; 2 HP-MSA1000-Storage-Systeme, 2 Brocade-FC-Switches Software: Suse Linux 9 (Basis-Betriebssystem), Samba 3.0 (Windows-File-Server, de.samba.org), Heartbeat 1.0.4 (Cluster-Manager, www.linux-ha.org), DRBD 0.6.10 (Plattenspiegelung, www.drbd.org)
Dienstleister	X-Cellent Technologies GmbH, www.x-cellent.com
Technologie-Lieferant	Suse (www.suse.de), Linbit (www.linbit.com), HP (www.hp.com/storage)
Kontaktaufnahme	seit zwei Jahren bestehendes Kundenverhältnis
Verhandlungsdauer	zwei Monate (Oktober bis November 2003)
größte Herausforderung	initiale Spiegelung der Daten in das Backup-Rechenzentrum
unerwartete Schwierigkeiten	Dualport-Gigabit-Adapterkarten: ungewöhnliches Netzwerkverhalten unter Linux-Fibre-Channel-Storage: starker Einfluss der Konfiguration auf die Performance
länger gedauert hat	Samba-Implementierung (Rechtekonfiguration)
Implementierungsdauer	zwei Monate (Dezember 2003 bis Januar 2004)
Arbeitsaufwand	45 Personentage
Kostenumfang	130.000 Euro
Projektaufteilung	Hardware: 65 Prozent; Dienstleistung; 35 Prozent
Service-Verträge	Hardware-Wartungsvertrag mit HP; Softwarepflege durch X-Cellent
Schulung	nicht notwendig, nur die Administratoren mussten kurz eingewiesen werden
Benefit für den Kunden	keine Softwarelizenzen (Open Source); geringe Hardwarekosten durch Verwendung kostengünstiger SAN-Infrastruktur mit Softwarespiegelung anstelle teurer Lösungen aus dem Enterprise-Bereich (Hitachi, EMC, Netapp)
Benefit für den Dienstleister	Neuer Referenzkunde im Verlagsumfeld; Erweiterung des Know-how im Bereich Hochverfügbarkeit; Verbesserung von DRBD; ausgereifte und kostengünstige Storage-Infrastruktur-Lösung mit Funktionen aus der Enterprise-Welt interessant für mittelständische Kunden