

Fileservice am CeBiTec: gestern, heute, morgen

Dipl.-Chem. Rainer Orth
CeBiTec/BRF
Universität Bielefeld
ro@CeBiTec.Uni-Bielefeld.DE

Übersicht

- Situation
- Motivation
- Evaluation
- Installation
- Expansion
- Migration

Situation: Status Quo

- derzeit:
 - 5 Haupt-Fileserver (2× Sun Fire V880, V440, X4600, X4540)
 - mit Storage: 15× Sun StorEdge 3510, 1× 3511, 1× J4400
 - Gesamtkapazität: 12.5, 11.7, 2.7, 40.6, 20.3 TB $\Rightarrow$ 87.8 TB
 - genutzte Kapazität: 9.1, 4.4, 1.4, 17.8, 16.7 TB $\Rightarrow$ 49.4 TB
- Betrieb unter Solaris 9 und 10
- diverse Filesysteme: UFS, VxFS, ZFS
- teilweise VxVM
- Netzwerkanbindung: Gigabit-Ethernet, f0rd: 5× Gigabit-Ethernet (IPMP)
- NFSv3/v4 direkt, CIFS mittels Samba 3 auf 2 der Fileserver

Motivation: Performance, Maintenance, ...

- massive Performance-Probleme bei Zugriffen aus dem Cluster
- derzeitige Fileserver von 2002–2009, teilweise aus der Wartung
- komplexe Mischung verschiedener Hard- und Software-Komponenten

Evaluation: mögliche Lösungen

- ab Mitte 2009: gezielte Suche nach Alternative
- primär: Storage-Appliance zur Vereinfachung von Maintenance, „alles aus einer Hand“
- zwei Hersteller: Sun mit Storage 7000, NetApp

Evaluation: NetApp

- Cluster aus $2 \times$ NetApp FAS3070, je $2 \times$ 1.8 GHz Dual-Core-Opteron, 8 GB RAM insgesamt $3 \times$ Diskshelves mit je $14 \times$ Platten, teilweise FC-AL, teilweise SATA, 16 GB-PAM-Modul
- Vorteil: lange Erfahrung im Filer-Markt
- Nachteile:
 - vorsichtige/langsame Weiterentwicklung (Beispiel: Kompression)
 - gewachsene Struktur, reichlich künstliche Limits (Aggregate)
 - inkonsistente Administration: z.T. über CLI, teilweise durch Editieren von Files im Root-Volume

Evaluation: NetApp (2)

- Nachteile:
 - Performance-Analyse nur für Spezialisten machbar (“black box”)
 - bei Auslastung des 10 GigE-Interfaces ist CPU am Anschlag, auch mit neuerem Modell keine entscheidende Verbesserung zu erwarten
 - viele Features einzeln lizenziert (Rollback)
 - schlechte Informationspolitik: kein Kundenzugang zu Dokumentation, Support-Website

Evaluation: Sun

- Cluster aus 2× Sun Storage 7410, je 4× 2.6 GHz 6-Core-Opteron, 128 GB RAM, insgesamt 3× J4400 mit 68× 1 TB SATA-Platten und 4× 18 GB SSDs, je 4× 1 GigE, 2× 10 GigE, “Try and Buy”
- Nachteil: neuer Player in diesem Markt, unsichere Zukunft des Produkts
- Vorteile:
 - supportete Version der neuesten ZFS-Features
 - vertrautes Administrationsmodell
 - zahlreiche Features nur hier verfügbar (Kompression)

Evaluation: Sun (2)

- Vorteile:
 - wesentlich bessere Performance: 30% CPU-Auslastung bei 10 GigE-Sättigung
 - einfache Performance-Analyse mit Analytics (“DTrace in a box”)
 - Lizenz umfaßt alles Features

Installation der Sun-Filer

- Betrieb als HA-Cluster: sehr hilfreich bei Wartungsarbeiten, Software-Updates
- länger kein NDMP-fähiges Backup-System, daher:
- initial nur Daten auf die Filer, die kein Backup brauchen
- Problem: CIFS nativ nur mit AD, kein NT4-Domain-Support
- daher: Resharing von NFS-Mounts mit Samba 3

Expansion: die nächste Ausbaustufe

- Ende 2010/Anfang 2011 stand Kapazitätserweiterung der S7410er und Replica-Filer an
- aber: J4400 waren ohne Vorwarnung zum Jahreswechsel abgekündigt worden, *Sackgasse*
- nach langem Hin und Her: S7410 als Replicas behalten, neuen S7420-Cluster mit Zielkapazität: je 2×1.86 GHz 6-Core-NehalemEX, 256 GB RAM, insgesamt $6 \times$ J4410 mit 136×2 TB SAS-2-Platten und 4×18 GB SSDs, je 4×1 GigE, 4×10 GigE

Migration: die weitere Reise

- Migration der bisherigen Daten auf die neuen Filer
- dazu: “shadow migration”
- Agenda:
 - Migration von alten auf neue Filer
 - Skripting von Projekt- und Volume-Erzeugung
 - mkacct-Anpassungen
 - Umzug (AGen-weise?)
 - Aufsetzen der Replicas