

DINI-Workshop:

Technische Aspekte des DINI-Zertifikats 2007

# Aspekte des Technischen Betriebs eines Repository-Servers

Christoph Ziegler

Berlin, 05.Juni 2007



TECHNISCHE UNIVERSITÄT CHEMNITZ  
UNIVERSITÄTSRECHENZENTRUM



# Zielstellung: Vortrag

- Erfahrungen und Anregungen vermitteln zu
  - Aufbau
  - Betrieb
  - Mitnutzungeines Repository Servers.

Erfahrungen resultieren aus

- Betrieb eines eigenen Dokumenten- und Publikationsservice seit 1995
- Eigener Entwicklung
- ca. 1500 Dokumenten
- Drei Server-Hardware-Generationen
- Zwei Betriebssystemen (Plattformen)

## ➤ Komponenten

- Software/Dienst -> Dokumenten- und Publikationsservice
  - Basis-Software -> WWW-Server, Datenbank, ...
  - Basisdienste -> Authentifizierung, Filesharing, ...
  - Betriebssystem/Plattform -> Linux/Windows/...
  - Hardware -> Server, Speicher, ...
- 
- ➔ Verantwortliche bestimmen/finden, die zum Projekt beitragen
  - ➔ Anforderungen festlegen (zukunftsorientiert)

•

•

- Anforderungen
  - 7x24h-Betrieb
  - Speicher für Dokumente und Metadaten
  - Lastverhalten beeinflusst durch
    - Parallele Zugriffe
    - Zu transferierende Datenmenge
    - Archivierung/Integration neuer Dokumente
    - Datensicherung

## Ziele

Hardware

Technologie

Zertifikat

Resümee

# Beispiel: Statistik 24-h-Betrieb



Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

## Stündliche Statistik

Jede Markierung (#, X) entspricht 918 Abfragen. # - lokal X - extern

0- 1 Uhr	3336 / 22794	##XXXXXXXXXXXXXXXX
1- 2 Uhr	4621 / 18366	###XXXXXXXXXXXX
2- 3 Uhr	3413 / 13195	##XXXXXXXX
3- 4 Uhr	2987 / 11663	##XXXXXXX
4- 5 Uhr	3734 / 11712	##XXXXXXX
5- 6 Uhr	4209 / 11596	##XXXXXXX
6- 7 Uhr	3823 / 14987	##XXXXXXXXXX
7- 8 Uhr	5994 / 18508	###XXXXXXXXXXXX
8- 9 Uhr	5611 / 31433	###XXXXXXXXXXXXXXXXXXXX
9-10 Uhr	7338 / 44250	###XXXXXXXXXXXXXXXXXXXXXXXXXXXX
10-11 Uhr	7510 / 57902	####XXXXXXXXXXXXXXXXXXXXXXXXXXXX
11-12 Uhr	9694 / 63158	#####XXXXXXXXXXXXXXXXXXXXXXXXXXXX
12-13 Uhr	7013 / 61373	###XXXXXXXXXXXXXXXXXXXXXXXXXXXX
13-14 Uhr	9035 / 58844	#####XXXXXXXXXXXXXXXXXXXXXXXXXXXX
14-15 Uhr	8905 / 65455	#####XXXXXXXXXXXXXXXXXXXXXXXXXXXX
15-16 Uhr	7505 / 66656	#####XXXXXXXXXXXXXXXXXXXXXXXXXXXX
16-17 Uhr	5457 / 59782	###XXXXXXXXXXXXXXXXXXXXXXXXXXXX
17-18 Uhr	4926 / 57540	###XXXXXXXXXXXXXXXXXXXXXXXXXXXX
18-19 Uhr	4675 / 44761	###XXXXXXXXXXXXXXXXXXXXXXXXXXXX
19-20 Uhr	5208 / 44291	###XXXXXXXXXXXXXXXXXXXXXXXXXXXX
20-21 Uhr	7347 / 41682	###XXXXXXXXXXXXXXXXXXXXXXXXXXXX
21-22 Uhr	3308 / 38908	##XXXXXXXXXXXXXXXXXXXXXXXXXXXX
22-23 Uhr	3305 / 37137	##XXXXXXXXXXXXXXXXXXXXXXXXXXXX
23-24 Uhr	6091 / 30944	###XXXXXXXXXXXXXXXXXXXX

Ziele  
Hardware  
Technologie  
Zertifikat  
Resümee



# Beispiel: Statistik 7-Tage-Betrieb



Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe		
Jede Markierung (#, X) entspricht 821 Abfragen. # - lokal X - extern		
01. Dec. 2006	1731 / 37047	##XX
02. Dec. 2006	629 / 26810	XX
03. Dec. 2006	252 / 28173	XX
04. Dec. 2006	1195 / 39895	#XX
05. Dec. 2006	1323 / 35407	#XX
06. Dec. 2006	1430 / 33798	#XX
07. Dec. 2006	1155 / 35580	#XX
08. Dec. 2006	1006 / 32343	#XX
09. Dec. 2006	272 / 20436	XX
10. Dec. 2006	215 / 23867	XX
11. Dec. 2006	805 / 37087	XX
12. Dec. 2006	1647 / 37665	##XX
13. Dec. 2006	1061 / 37760	#XX
14. Dec. 2006	3677 / 32131	###XX
15. Dec. 2006	1620 / 28782	#XX
16. Dec. 2006	200 / 17741	XX
17. Dec. 2006	182 / 20013	XX
18. Dec. 2006	1729 / 32792	##XX
19. Dec. 2006	3552 / 32902	###XX
20. Dec. 2006	1453 / 28981	#XX
21. Dec. 2006	1430 / 27111	#XX
22. Dec. 2006	1489 / 22064	#XX
23. Dec. 2006	13 / 14817	XX
24. Dec. 2006	18 / 13550	XX
25. Dec. 2006	92 / 13478	XX
26. Dec. 2006	1 / 17197	XX
27. Dec. 2006	65 / 21730	XX
28. Dec. 2006	121 / 20014	XX
29. Dec. 2006	41 / 29728	XX
30. Dec. 2006	10 / 17079	XX
31. Dec. 2006	65 / 12983	XX

Ziele  
Hardware  
Technologie  
Zertifikat  
Resümee



# Hardware: Auswahlaspekte

- Einsatz redundanter Komponenten, um einem Serverausfall vorzubeugen:
  - Netzteile (Stromversorgung)
  - Lüfter
  - Netzkarten
  - Hot-swap-Platten (Spiegelung des Systems)
- erhöhte Qualität und Robustheit (*Lüftung vs. Wärmeentwicklung, Lüfter, Netzteile, ...*)
- bessere Wartungseigenschaften bzw. Reparierbarkeit (*hot-swap-Platten, öffnen Seiten-/Rückwand*)
- mehr interner Platz für zusätzliche Komponenten
- höhere CPU-Leistung und/oder größerer RAM
- Wartungsvertrag?

- ◆ Nicht überdimensionieren
- ◆ Erweiterungsmöglichkeiten werden selten genutzt
- ◆ Testinstallation/Testbetrieb auf vorhandener Hardware sollte Anregungen geben, ist jedoch niemals vergleichbar mit Anforderungen aus Produktionsbetrieb



# Hardware: Speicherbedarf

- Achtung: Datenredundanz berücksichtigen!
- Beispiel TU Chemnitz (MONARCH):
  - 1436 Dokumente belegen 7,320 Gbyte
  - 5,2 Mbyte je Dokument

## Entwicklung, Speicherbedarf je Dokument:

- 1995: 1,8 MByte (größtes Dokument 17 Mbyte)
- 1996: 9,2 MByte (größtes Dokument 45 Mbyte)
- 1997: 7,8 MByte (größtes Dokument 55 Mbyte)
- 1998: 3,9 MByte (größtes Dokument 98 Mbyte)
- 1999: 10,6 MByte (größtes Dokument 76 Mbyte)
- 2000: 5,0 MByte (größtes Dokument 119 Mbyte)
- 2001: 4,4 MByte (größtes Dokument 35 Mbyte)
- 2002: 6,1 MByte (größtes Dokument 104 Mbyte)
- 2003: 4,7 MByte (größtes Dokument 84 Mbyte)
- 2004: 5,4 MByte (größtes Dokument 114 Mbyte)
- 2005: 4,5 Mbyte (größtes Dokument 43 Mbyte)
- 2006: 4,5 Mbyte (größtes Dokument 39 Mbyte)
- 2007: 5,3 Mbyte (größtes Dokument 49 Mbyte)

Ziele

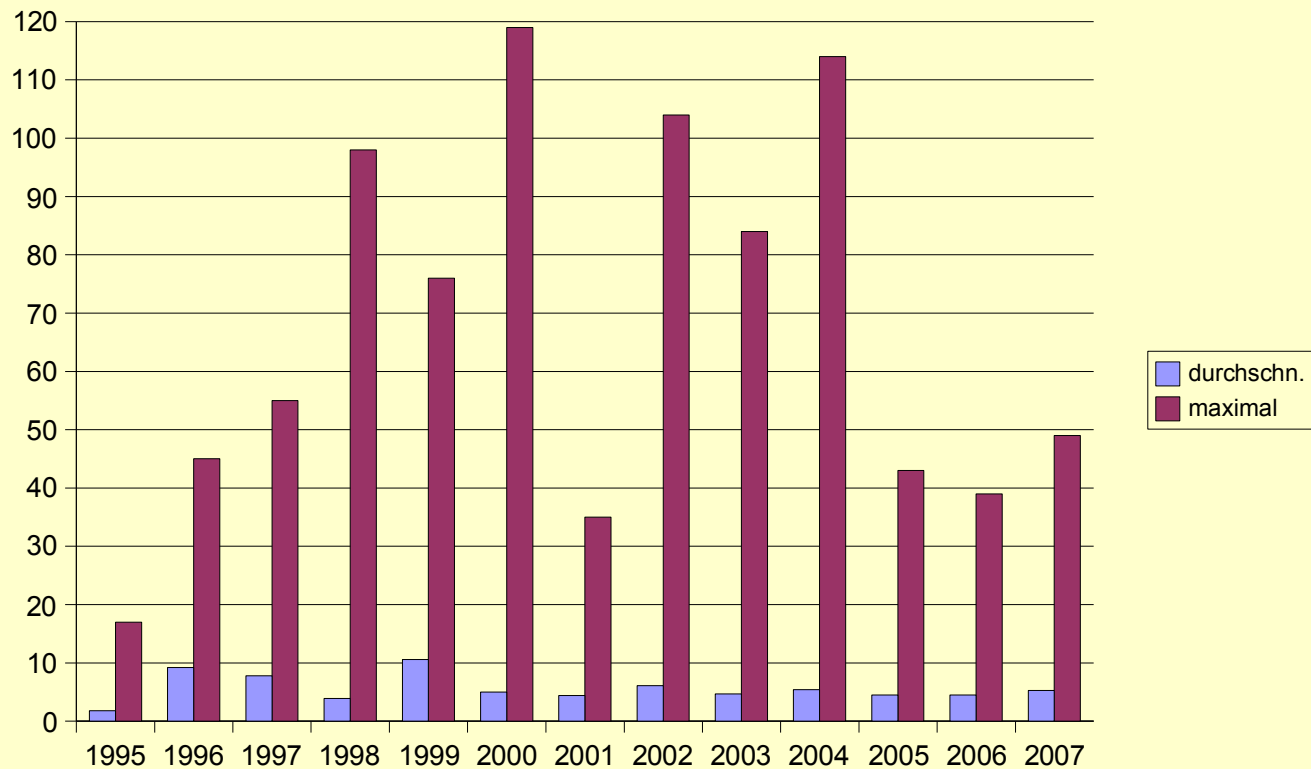
Hardware

Technologie

Zertifikat

Resümee

## ➤ Speicherbedarf für Dokumente (Diagramm)



# Hardware: Speichertechnologien

- DAS (direct attached storage)
  - Lokale Platten, lokales Raid
    - keine Skalierbarkeit; keine Redundanz; kein Weiterbetrieb bei Ausfall dieses Repository Servers
    - Raid – keine Datensicherung; Redundanz nur bei Ausfall weniger Platten
- SAN/NAS (storage area network/network attached)
  - Incl. Speicherdienste
  - Von Repository Server unabhängig -> Redundanz
  - Skalierbarkeit
  - Datenträgermigration
- Offline Datenträger
  - Optische Datenträger, Tapes
  - zur separaten offline-Datensicherung (read only)
  - Nur sinnvoll, wenn physisch separat (im Safe) gelagert

# Hardware: Aufstellung

- Server gehören in ein Data Center, nicht an den Arbeitsplatz
- optimale Umgebung (für Hardware!)
  - USV, Stromversorgung
  - Klima (Temperatur, Luftfeuchte)
  - Datennetzinfrastruktur
  - Aufstellplatz – Fokus liegt auf Technik
- Brandschutz
- Zugang
- policy notwendig (was, wer, wie - Beteiligte)

# Technologie: Installation/Konfiguration



- Ziel: identische Server bzw. identische Ersatz-/Nachfolgeserver
  - Installation
  - Netzwerk
  - Integration in das Identitätsmanagement der Einrichtung (Authentication, Authorisation, Accounting)
  - Server- und Dienstzugang
  - Monitoring-Technologie, logging
  - Regelmäßige zeitabhängige Aktivitäten (cron)
  - Spezifische Files/Verzeichnisse incl. Zugriffsrechte
  - Softwareinstallation
  - Anbindung an zentralen Speicherdienst

- Systematische Administration!
- Spezialisierung, Knowhow, Themenverantwortung je
  - Dienst (Dienstverantwortlicher)
  - Sicherheit
  - Basisdienste
  - Netztechnologie
  - Betriebssystem
- Automatisierung notwendiger bzw. häufiger zu wiederholender Vorgänge (Routinetätigkeit)
- Ziel: Wiederherstellbarkeit
- Beschreibung/Dokumentation (Wiederholung, Vertretung)
- policies je Verantwortungsbereich

- Ziel: single point of failure (SPOF) vermeiden
  - Betreiben mehrerer (aktiver) Server
  - Verwendung einer Alias-Technologie (DNS-Alias) (*an den Dienstnamen gebundene Adresse (nicht an den Servernamen)*)
  - zeitversetzte Wartungsarbeiten
  - gleichzeitige Nutzung mit spezieller Netztechnologie (layer-4-switching-Technologie)
  - mehrere Dienste auf jedem Server möglich, bei Notwendigkeit dedizierbar

# Technologie: Wartung

- Wartungszeiten/Wartungsfenster
- Ausführung aller geplanten Wartungsarbeiten
- generelles Wartungsfenster oder Ankündigung (*Ankündigung wegen anonymer Nutzung problematisch*)
- **Nur** geplante und getestete Aktivitäten durchführen
- Zeitplan einhalten: definierte Deadlines beachten
- Testung vs. Wartung vs. Produktionsbetrieb



- systematische Erfassung, Beobachtung oder Überwachung
- "Historisches Monitoring" ( *Performance Monitoring, zur automatischen Erzeugung von Langzeit-Statistiken*)
  - Verfügbarkeit
  - Ausnutzung
  - Durchsatz

# Technologie: Monitoring (2)

## ➤ "Real-Time Monitoring"

- laufende Überwachung der Funktionalität von Diensten
- Registrierung von Diensteausfällen
- Sofortige aut. Benachrichtigung des/der Verantwortlichen  
Ziel: Verantwortlicher hat Kenntnis, bevor der (erste) Nutzer den Ausfall bemerkt

## ➤ Überwachung von

- Speicherplatz, RAM/CPU -> Lastverhalten, Kapazitätsprobleme, Wachstumsraten
- Netz (Traffic)
- Systemfehler, Ereignisse
- Lüfter, USV, Prozessor, Board (Temperatur)
- Dienste, Protokolle (WWW – httpd, Filesharing – Speicherverfügbarkeit), Erreichbarkeit
- (Ausfall) redundanter Komponenten



# Beispiel: Monitoring Ressourcen



Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

big brother



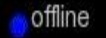
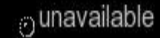
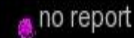
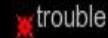
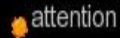
la  
Fr Jun 1 15:38:21

Suche nach Host:

*WEB+MONARCH*

[connfp](#) [cpu](#) [disk](#) [http](#) [md](#) [memory](#) [msgs](#) [procs](#)

	connfp	cpu	disk	http	md	memory	msgs	procs
<a href="#">ham</a>								
<a href="#">sem</a>								



- Ziele
- Hardware
- Technologie
- Zertifikat
- Resümee



# Beispiel: Monitoring connections



Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe

big brother



Fri Jun 1 15:...

Thu May 31 15:43:57 2007

Fr

16 17 18 19 20 21 22 23 0 1 2 3 4 5 6 7 8 9 10 11 12

## ham - connfp

Last 24 Hours

100%	0%	0%	0%	0%	0%
[Total may not equal 100 %]					

### Last 50 log entries (Full HTML log)

Date	Status	Duration
Thu Mar 8 11:18:51 2007		85 days 03:25:06
Thu Mar 8 11:09:12 2007		0:09:39
Sun Nov 12 19:40:39 2006		115 days 15:28:33
Sun Nov 12 17:09:47 2006		2:30:52
Wed Oct 25 06:39:41 2006		18 days 11:30:06
Wed Oct 25 06:34:50 2006		0:04:51
Wed Jul 26 06:48:24 2006		90 days 23:46:26
Wed Jul 26 06:43:18 2006		0:05:06
Thu Mar 23 06:49:32 2006		124 days 22:53:46
Thu Mar 23 06:39:42 2006		0:09:50
Tue Mar 14 10:02:32 2006		8 days 20:37:10
Tue Mar 14 09:57:35 2006		0:04:57
Thu Mar 9 07:15:36 2006		5 days 02:41:59
Thu Mar 9 07:12:39 2006		0:02:57

- Ziele
- Hardware
- Technologie
- Zertifikat
- Resümee

- Automatisch
- in Speichertechnologie integrierte Sicherung zu empfehlen
- wichtig: Technologie der Wiederherstellung
- Wann wird Notwendigkeit der Wiederherstellung erkannt?
- Gründe
  - defekte Hardware (Datenträger)
  - RAM, CPU, Bussysteme, Kabel, ... - kann auch zu Datenveränderung führen (wie feststellbar?)
  - Verlust der Datenträger (Diebstahl, Vandalismus, Brand, Wasser, ...)
  - Verlust der Daten (Manipulation, Zerstörung)
  - Bedien-/Programmierfehler (versehentliches Löschen, Verändern)
  - Archivarische Gründe (disaster recovery, juristische Auseinandersetzungen, o.ä.)

- Betriebskonzept gewährleistet eine angemessene Verfügbarkeit des Systems.
  - Versionsangaben, technische Parameter
  - Zugangsregelung
  - Regelung zur Wartung
- Dokumentation des technischen Systems
- Technologie zur Sicherung und Wiederherstellung
- Sichere Installation von System und Software
- Regelmäßige Systemwartung
- Kontrollierte und nachweisbare Aufnahme von Dokumenten

- Kundenzufriedenheit ist wichtig
- Den eigenen Anforderungen gerecht werden!
- Den Zertifikatsanforderungen gerecht werden!

# Aspekte des Technischen Betriebs eines Repository-Servers

Danke für Ihr Interesse!

[christoph.ziegler@hrz.tu-chemnitz.de](mailto:christoph.ziegler@hrz.tu-chemnitz.de)

<http://archiv.tu-chemnitz.de/pub/2007/0075>



TECHNISCHE UNIVERSITÄT CHEMNITZ  
UNIVERSITÄTSRECHENZENTRUM

