



ulm university universität  
uulm



## Bacula – Eine strategische Entscheidung

Thomas Nau

Kommunikations- und Informationszentrum der Universität Ulm, kiz

# Übersicht

- kiz @ Ulm
- Auslöser für den zeitnahen Wechsel
- Entscheidungsfindung
- Bacula Enterprise Edition (BEE) als Lösung
- Zusammenfassung
- Q & A

## kiz, Kommunikations- und Informationszentrum

- 2002 aus den ehemaligen Bereichen „Rechenzentrum“, „Bibliothek“, „Telekommunikation“ und der „Zentrale für Foto, Grafik und Reproduktion“ gegründet
  - in der Zwischenzeit ein gängiges Modell in ganz Deutschland
- Nutzung vorhandener Synergien
- 100+ Beschäftigte in 6 Abteilungen

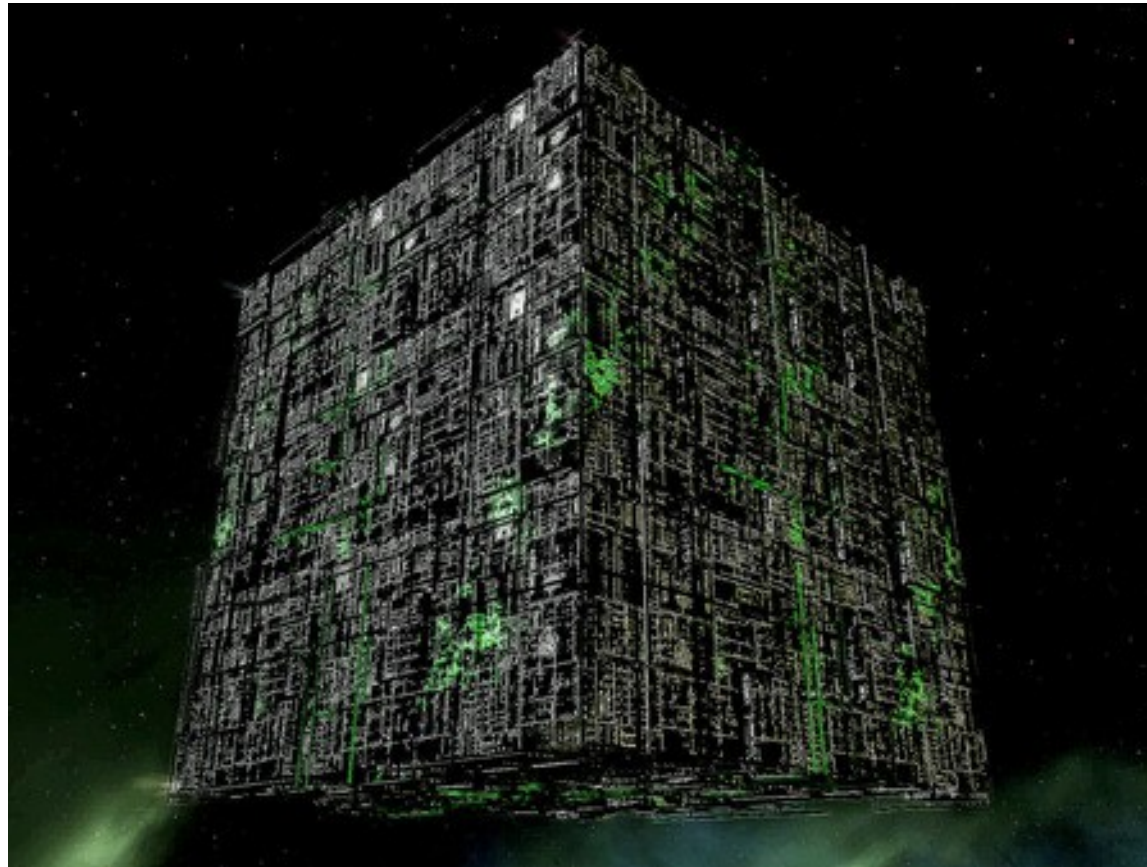
# Abteilung Infrastruktur

- tolles Team das Dienstleistungen für ~14.000 Personen und 4 Fakultäten erbringt
  - gesamte zentrale Server-, Verwaltungs-IT und Netzwerke
    - LAN, MAN und WLAN
    - Telefonie mit 14.000 Anschlüssen
    - Anbindung an BelWue mit 10Gbit demnächst 100Gbit
  - ca. 600 Desktop PCs und Notebooks mit Windows und Linux
  - Backup Service für mehrere Universitäten in Baden-Württemberg
  - Landes HPC Cluster mit Schwerpunkt „Theoretische Chemie“
  - Hype der Woche (IaaS, Cloud, Fog, Rain, ...)

## Die Vorteile

- ein Team, ein Ziel mit einer Verantwortung
  - Team umfasst neben Rechnerbetrieb auch IT- Security und Netzwerk; großes Plus durch „reality check“
  - macht es einfacher Entscheidungen zu treffen und gemeinsam zu tragen
- alle werden dazu ermutigt eigene Ideen zu entwickeln und sie im Gesamtkontext zu verfolgen

# Wir sind also eine Art „freundliches Kollektiv“



[http://scifi.wikia.com/wiki/File:Borg\\_cube.jpg](http://scifi.wikia.com/wiki/File:Borg_cube.jpg)

# Über mich

- von „Haus aus“ Physiker
- erste IT Berührung mit einer PDP-11
- Schwerpunkte
  - UNIX Systeme
  - File- und Storage Systeme
  - IT Betrieb
- balanciere zwischen Technik und Management



# Hintergründe zur „Strategie“



# The Good, the Bad and the Ugly

- nahezu alle Universitäten und Hochschulen in Baden-Württemberg setzen seit mehr als 15 Jahren TSM ein
  - Landesverträge mit Lizenzkontingent (Processor Value Unit) für je 5 Jahre
  - letzte Verlängerung 2012
- viele Synergieeffekte
  - Arbeitsgruppe trifft sich regelmäßig zum Erfahrungsaustausch
  - off-site Spiegel von Daten, ...
- diese langen Zeiträume machen auch träge und blind für Entwicklungen und Alternativen
  - „haben wir doch immer so gemacht“

# The Good, the Bad and the Ugly

- 2012 - 2014: Jahre der allgemeinen Lizenz Audits
  - viele Anbieter an vielen Hochschulen: Oracle, Microsoft, IBM, ...
- IBM beauftragt Wirtschaftsprüfer mit TSM Audit in Ulm
  - steht im Einklang mit AGBs und auch Rechenzentren überprüfen die Einhaltung Ihrer Nutzungsbedingungen
- enormer Personalaufwand auf Seiten des kiz
  - Erfassung und Aufbereitung dauert trotz Kenntnis aller Ansprechpartner in den Instituten einige Monate
    - Software bot keinerlei Hilfe
  - Abteilungen tauschen Systeme ohne Wissen des kiz aus
    - Änderung des PVU Wertes → unberechenbares Kostenrisiko

# The Good, the Bad and the Ugly

- unabhängig vom Anbieter zeigt das Ergebnis deutlich die Schwachstellen des Modells auf
  - verstehen die Anbieter das Hochschulumfeld?
    - Volumen und PVU Modelle sind in dezentralen Umgebungen nicht zukunftssicher planbar
  - Einstufung als „Server“ oft schwer verständlich und vermittelbar
    - Uni SAP Systeme kosten lediglich 30% eines Physik HPC
    - für einige Institute wäre dadurch Backup nicht mehr bezahlbar
- Kostenexplosion durch zusätzliche Lizenzen (mehr offline)
  - Mehrkosten werden für ein Jahr aus kiz Etat bezahlt
  - sportliche Vorgabe: alternative Lösung in diesem Zeitraum zu evaluieren und zu etablieren

# Entscheidungsfindung

# Alternative Lösungen

- „A“ Kriterien
  - langfristig planbares, einfaches und Risiko armes Lizenzmodell
    - keine versteckten Personalkosten für die Verwaltung von Lizenzen
  - tauglich für universitäre Größenordnungen und Hochschul übergreifenden Einsatz
    - Petabytes an Daten in Milliarden Files von tausenden (exotischen) Rechnern in allen möglichen internationalen Zeichensätzen
  - „gut genug“ als Entscheidungskriterium (mehr dazu später)
- „B“ Kriterien
  - kein Vendor lock-in aber Sicherstellung das Datenzugriff dauerhaft ohne Kosten möglich ist
  - offen gelegte und dokumentierte Formate, Schnittstellen, ...

## „A“ Kriterien

- planbares, einfaches und Risiko armes Lizenzmodell
  - Sondierungsgespräche mit potentiellen Anbietern zeigten das diese das Hochschulumfeld noch (immer) nicht verstehen
  - Beharren auf Volumen- oder PVU Modellen, teilweise mit zwingender jährlicher Steigerung
- tauglich für universitäre Größenordnungen
  - trennt Spreu vom Weizen
  - alle Anbieter von Enterprise Software erfüllen diesen Punkt wenngleich die Internationalisierung ein Kernproblem darstellt
- „gut genug“
  - Nerd Problem: Aufwand für Optimierung steht oft in schlechtem Verhältnis zum Aufwand

## „B“ Kriterien

- kein Vendor lock-in
  - wünschenswert
  - alternativ dynamisiert ein weiterer Anbieter im Land auch zukünftige Verhandlungen
    - kritische Masse muss hierbei erreicht werden
    - Ulm ist zu klein aber wir haben Mitstreiter in Konstanz und Tübingen
- offene und dokumentierte Formate, Schnittstellen, ...
  - Wunsch nach freier Wahl der Datenbank
    - vorhandenes know-how nutzen
    - dokumentierte Tabellen erlauben es eigene reports zu generieren
  - Sicherungen dauerhaft ohne Kosten lesbar

# (K)ein Dilemma, Open, Closed oder proprietär ?!

## Pro OpenSource

- kein Vendor lock-in
- offene und meist sehr gut dokumentierte Schnittstellen
- Kosten
  - es ist jedoch eine Vollkostenbetrachtung zwingend
- Community Support

## Cons OpenSource

- oft zu schnelle Release-Zyklen mit API Brüchen
  - OpenStack 6 Monats Zyklus
  - Filesystem des Tages
- Zukunftssicherheit bei kleineren Projekten
- Nerd „Selbstbeschäftigung“
- Community Support nicht „einforderbar“
  - auch bei kommerziellen Anbietern oft sehr schwierig



# Lösung: das Beste aus den Welten

- frei für viele Plattformen verfügbare OpenSource Basis
- ergänzt um kommerziellen Support
- ggf. ergänzt um Entwicklungen zur Erweiterung des Funktionsumfangs
  - closed source, oft temporär, durchaus akzeptabel solange Datenformate usw. davon unberührt sind
- idealer Weise **ein** Dienstleister der
  - groß genug ist um eine breite Kundenbasis bedienen zu können (Diversität bei Hard- und Software)
  - klein genug ist um engen Kundenkontakt pflegen zu können und dessen Umfeld zu kennen

# Lösung: das Beste aus den Welten

- das Erfolgsmodell „SOGo“ des Uni RZ Konstanz dient uns als Vorlage
- Bacula Enterprise Edition (BEE) von Bacula Systems
  - Empfehlung durch Kollegen der Universitäten Hannover und Gent
    - Stichwort „gut genug“
  - Verfügbarkeit für alle notwendigen Plattformen
    - Solaris 11 SPARC als Bacula Director und Storage Daemon
    - IBM Jaguar Bandlaufwerke und zugehörige Roboter
  - PostgreSQL als Datenbank
  - kompatibel zur Community Edition
    - erlaubt z.B. Zugriff auf alle Daten

# BEE @ Ulm

## Zeitlicher Ablauf

- April – Juni '14 Proof of Concept (POC) zeigt Eignung für Verbund Ulm, Konstanz und Tübingen
- Juli – August '14 Förderung eines Pilotbetriebs durch MWK und Planung Produktivbetrieb
- September '14 Start Produktivbetrieb im kiz
- Dezember '14 alle kiz Server (SAP, DB, Mail, ...) von einer Person umgestellt
- ab Dezember '14 Migration aller anderen Server in Ulm sowie Start der Umstellung in Konstanz und Tübingen

## Aktueller Stand

- über 15.000 Backup Jobs ausgeführt
- > 400 TB mit > 750 Millionen Dateien
- Windows, Apple OSX, Solaris und diverse Linux Varianten werden regelmäßig gesichert
- „Außenstellen“ Konstanz und Tübingen angebunden
  - vor Ort Storage Daemons als „Cache Server“
  - Sicherung erfolgt auf Tape-Library in Ulm

## Vorhandene „exotische“ Hardware @ Ulm

- 2x Oracle SPARC-T4 Server
  - 128GB RAM pro System
  - 3 dual-port SAS-2 HBA plus 1 dual-port 16Mb Fiber-Channel HBA
- 2 JBODs je 60 x 4TB SAS-2 Enterprise Platten
  - 9x 6+2 RAIDZ2 plus 4 hot-spare pro Chassis
    - *filebench* multi-stream IO: ~3.5 GB/s
- 8x IBM 3592-E07 Fiber-Channel Bandlaufwerke
  - 250-600 MB/s abhängig von Kompression
- PostgreSQL 9.2 auf 6x 800GB SAS-2 Enterprise SSDs

# Backup Strategie

- alle Klienten sichern täglich inkrementell
- kleine Systeme (< ~ 500GB) sichern einmal pro Monat voll
- für große Systeme, derzeit bis 10TB, wird jeden zweiten Monat ein virtuelles Voll-Backup erzeugt
  - einige verwenden snapshots und ZFS replication
- alle Daten werden zuerst auf den Plattensystemen abgelegt
- alle inkrementellen und differenziellen Backups verbleiben für ihre Lebensdauer auf den Platten
- Voll-Backups werden bei Bedarf auf Band migriert

## Offene Punkte

- einfacherer Zugang für Instituts Admins
  - derzeit wird vieles zentral vom kiz erledigt, etwa restores
- Shibboleth Anbindung des Web-GUI
- **echte** Archiv Lösung
  - Bacula bietet ausschließlich Backup
- fine-tuning der Backup Strategie
  - für Systeme mit  $> \sim 15\text{TB}$  zwingend da Bacula keinen „incremental forever“ Ansatz verfolgt



# Zusammenfassung

- die langfristige Entscheidung für die „mostly open“ Lösung Bacula Enterprise Edition fiel bewusst
  - strategische Auswirkungen sind eine hilfreiche Zugabe
- Bacula und Bacula Systems haben beide unsere Erwartungen erfüllt
  - Zeitplan spricht für sich
  - hervorragenden Support auch **nach** Abschluss der Verträge
- **aber:** Bacula verfolgt andere Paradigmen wie TSM
  - das Team muss sich aktiv darauf einlassen und neue Wege gehen
  - organisatorische Anpassungen sind zwingend

# Zusammenfassung

- aus Sicht der Lizenz-Problematik haben wir den Zeitplan gehalten aber hatten ein etwas „mehr“ erwartet
  - notwendige Abstimmung mit Instituten bremst weit mehr als erwartet wenngleich das Feedback durchweg sehr positiv ist
- die Entscheidung von Konstanz, Tübingen und Ulm hat viel Interesse geweckt :-)