

## AUF EINEN BLICK

### AUFGABE

Einführung eines Datenmanagement-Systems mit Information Lifecycle Management Funktionalität

### SYSTEME & SOFTWARE

- 8 x IBM DS4800 Storage Subsysteme mit insgesamt 60 EXP810-Speichererweiterungseinheiten (insg. 413 TB)
- 2 x IBM SAN64-B2 und 2 x IBM SAN32-B2-FC Switches
- IBM SAN Volume Controller mit 8 Knoten
- 2 x IBM x3650 Server
- 4 x IBM POWER5 p520 Server
- 2 x IBM POWER5 p550 Server
- AIX und HACMP
- IBM General Parallel File System
- IBM TotalStorage Productivity Center
- IBM Tivoli Storage Manager

### VORTEILE

- Effizientes Speichermanagement durch Aufteilung des Speichers in mehrere Speicherklassen
- Virtualisierung auf Block-Ebene durch SAN Volume Controller
- Automatisiertes, regelbasiertes Abspeichern bzw. Migrieren der Daten in die Speicherklasse mit dem jeweils benötigten Service-Level durch GPFS
- Übergreifendes Management und Überwachung der Speicher-Infrastruktur durch IBM TPC
- Ausfallsicherheit durch Redundanz und Investitionsschutz durch Einhaltung von Standards wie SMI-S
- Einheitliche Ablage der Daten aus verschiedenen Projektgruppen und transparente Bereitstellung über bewährte Protokolle
- Datensicherung und Recovery über IBM Tivoli Storage Manager

## INFORMATION LIFECYCLE MANAGEMENT: OPTIMIERTER DATENZUGRIFF FÜR DAS GFZ

Speichervirtualisierung, -automatisierung und IBM TotalStorage Productivity Center sorgen für mehr Transparenz und Flexibilität im Datenmanagement des Helmholtz-Zentrum Potsdam (GFZ).

### DAS HELMHOLTZ-ZENTRUM POTSDAM (GFZ)

Das Helmholtz-Zentrum Potsdam / Deutsches GeoForschungsZentrum-GFZ ist das nationale Forschungszentrum für Geowissenschaften in Deutschland.

Mit mehr als 800 Beschäftigten umfasst es alle Disziplinen der Geowissenschaften von der Geodäsie bis zum Geoingenieurwesen und betreibt sie in einem engen interdisziplinären Verbund mit den benachbarten Naturwissenschaften. Die methodischen Kernkompetenzen des GFZ liegen in der Anwendung und Entwicklung von Satellitentechnologien und raumgestützten Messverfahren, im Betrieb globaler und regionaler geodätisch-geophysikalischer bodengestützter Messnetze, im Einsatz tomographischer Verfahren der geophysikalischen Tiefensondierung, in der Durchführung von Forschungsbohrungen, in der Labor- und Experimentiertechnik sowie in der Analyse und Modellierung von Geoprozessen.

### HINTERGRUND

Primäres Ziel des GFZ war die Einführung eines durchgängigen Information Lifecycle Management (ILM) Systems für die Ablage unstrukturierter Dateien.

Dabei war wichtig, dass die Anwender zukünftig ihre Daten auf Speicherressourcen mit verschiedenen Ausprägungen bezüglich definierter Parameter wie Verfügbarkeit, Zugriffszeit, Datensicherungsmodus etc. frei einstellen können sollten. Ziel war es, zukünftig die von den Anwendern beanspruchten Speicherkapazitäten entsprechend dem verwendeten Typ der Ressource differenziert auflisten zu können.

Es wurde eine Aufteilung in vier verschiedene Disk-Klassen (Tiers) geplant, die sich durch die definierten Zugriffszeiten, Transferraten, Dauern der Nutzungszeit usw. unterscheiden sollten.

**EIN DATENMANAGEMENTSYSTEM  
ÜBERNIMMT DIE AUFGABE  
DES INFORMATION LIFECYCLE  
MANAGEMENTS.**

**SVA LÖSUNG**

Die von SVA konzipierte und implementierte Lösung ist ein Datenmanagementsystem, das die Aufgabe des Information Lifecycle Managements übernimmt. Dabei wurde IBM Storage- und Server-Hardware weitgehend symmetrisch auf zwei Rechenzentren aufgeteilt, die ca. 600 m auseinander liegen. Mit IBM DS4800 Storage Subsystemen und IBM SAN Switches wurde ein Speichernetzwerk aufgebaut, ergänzt durch IBM SAN Volume Controller zur Virtualisierung des Speichers. IBM POWER5 p520-Server dienen als GPFS-Clusterknoten zur Bereitstellung des General Parallel File-Systems mit einem integrierten, policy-basierten und automatisierten Lifecycle-Management. Darüber hinaus sichern IBM POWER5 p550 Server die Systemumgebung mit IBM Tivoli Storage Manager.

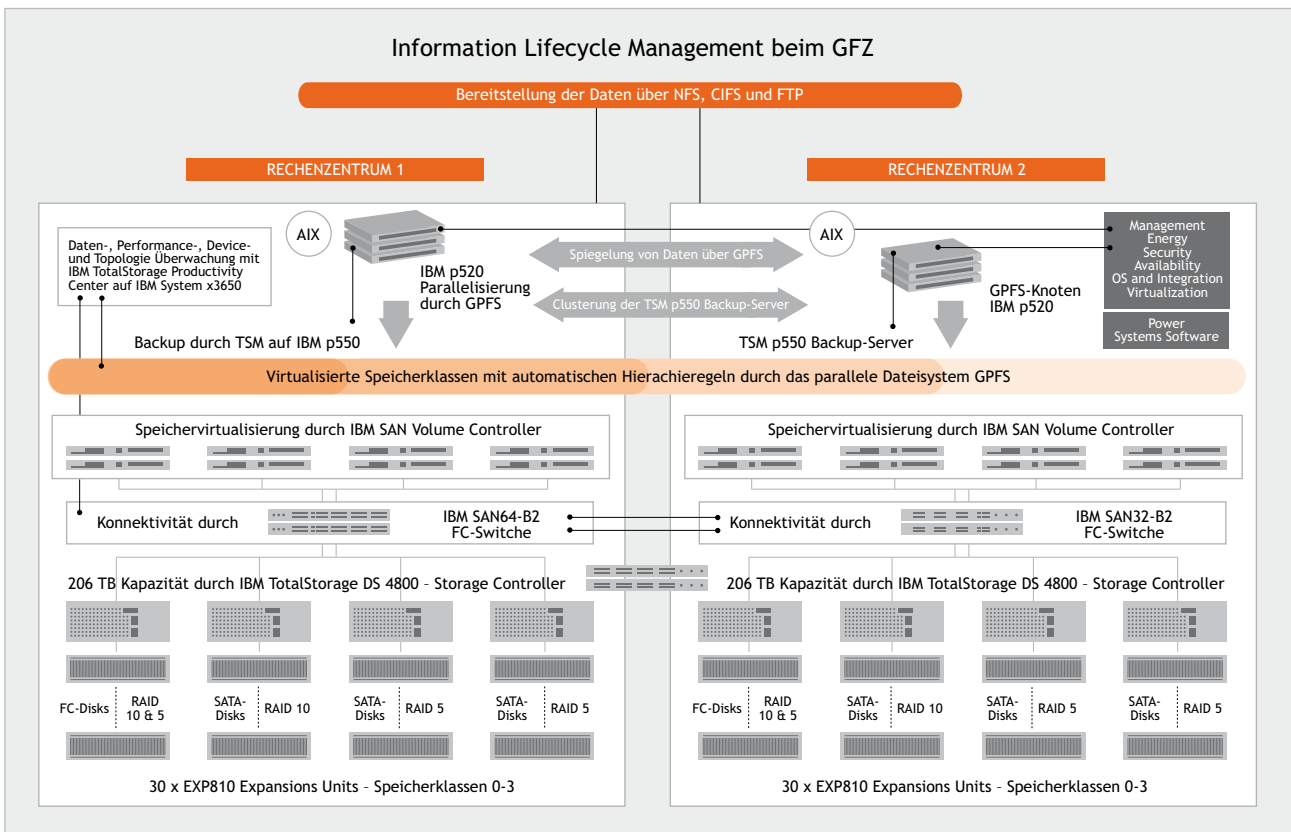


Abbildung: Schematische Darstellung der TPC Lösung für das GeoForschungsZentrum Potsdam

Auf einem IBM x3650 Server laufen Dienste des IBM TotalStorage Productivity Centers, unter anderem zur Überwachung der SAN-Infrastruktur und der Kapazitätsauslastung. Als Softwareprodukte kommen zusätzlich AIX als Betriebssystemplattform der POWER5 Server sowie HACMP zur Erhöhung der Ausfallsicherheit des GPFS zum Einsatz.

**DIE BLOCK LEVEL VIRTUALISIERUNG SCHAFFT DIE VORAUSSETZUNG FÜR BLOCKBASIERTE BEWEGUNG DER DATENBESTÄNDE ZWISCHEN EINZELNEN LEVELS.**

#### **DATENMANAGEMENT IM DETAIL**

Die Grundlage des Systems bilden Disk Storage Systeme, die in der Ebene „Block Devices“ die Tiers (Speicherklassen) 0 bis Tier 3 realisieren. Dabei kommen für Tier 0 und 1 Fibre-Channel Platten, für Tier 2 und Tier 3 SATA basierte Festplatten zum Einsatz. Oberhalb der physischen Storage Systeme liegt die Block Level Virtualisierung, die eine Entkopplung der Storage Systeme von den Servern und File-System-Strukturen realisiert. Damit wird es möglich, logische Volumes auf beliebigen Hardware-Systemen zu definieren und dabei auch physische Grenzen der Hardware nicht mehr als Limitierung zu berücksichtigen. Diese Fähigkeit der Block Level Virtualisierung schafft die Voraussetzung für die blockbasierte Bewegung der Datenbestände zwischen den einzelnen Levels (QoS).

Bei der eingesetzten Virtualisierung handelt es sich um eine „In-Band-Virtualisierung“. Es werden durch die Virtualisierungs-Engines zusätzliche Performance-Ressourcen, wie z. B. globaler Cache, zur Verfügung gestellt. Bei erforderlichem Wachstum des Gesamtsystems können diese Virtualisierungs-Engines linear skalierend mitwachsen. Die für QoS-Level 0 erforderlichen SAN Volumes werden den entsprechenden Servern durch die Block Level Virtualisierung direkt über Fibre Channel zur Verfügung gestellt und erfüllen damit alle Anforderungen an Verfügbarkeit und Performance.

**BESSERE SKALIERBARKEIT UND FEHLER-TOLERANZ DURCH GPFS.**

#### **FILE MANAGEMENT MIT IBM GPFS**

In der nächsten Ebene erfolgt eine File Level Virtualisierung auf der Basis des hochperformanten, shared-Disk basierten General Parallel File System (GPFS). Das GPFS bietet gegenüber einem Standard File-System wesentlich bessere Skalierbarkeit und Fehlertoleranz, da der Zugriff auf die physische Ebene parallelisiert über mehrere Nodes erfolgt. Dieses File-System ist aufgrund seiner Funktionen (wie z. B. Policy Driven Automation, Automatic Journaling, Concurrent File Access, Data Protection) in der Lage, alle vom GFZ geforderten Eigenschaften abzubilden.



**DAS SYSTEM ERFÜLLT GESAMT-  
HEITLICH DIE ANFORDERUNGEN  
AN EIN STORAGE SERVICE LEVEL  
MANAGEMENT.**

Das File-System wird physisch auf einem GPFS-Cluster realisiert, der auf die SAN Volumes der Block Level Virtualisierung zugreift. Für die Anwender erfolgt die Bereitstellung der Files, wie gefordert, über NFS, CIFS und FTP Protokoll.

#### **MONITORING MIT TPC**

Für die Überwachung der ILM-Lösung im Backend wurde mit dem TotalStorage Productivity Center eine Lösung realisiert, die auf der Basis der vorhandenen Standards wie SMI-S und SOA ein End-to-End Monitoring und Management der gesamten ILM Infrastruktur erlaubt. Dabei wurden speziell die Anforderungen einer service-orientierten Architektur berücksichtigt, wie z. B.:

- Einsatz von Standards
- Loose Coupling
- Accessibility and Reuse
- SOA Governance.

Ebenso sind Definitionen der IT Information Library (ITIL) und Data Retention Compliance in diese Lösung integriert und das System erfüllt somit gesamt-heitlich die Anforderungen an ein Storage Service Level Management. Auf der Hardware-Ebene werden alle Komponenten unterstützt, die den SMI-S Versionen 1.0.2 und 1.1 entsprechen. Neben den in die Lösung integrierten Komponenten können damit auch alle Produkte, die dieser Spezifikation entsprechen, mit in das Monitoring und Management integriert werden.

#### **VORTEIL: OPTIMIERTE DATENVERTEILUNG**

Die implementierte Lösung bietet nun eine performante Virtualisierung auf Block-Ebene, die automatisierte Einhaltung/Anwendung, vielfältige Anwendungsmöglichkeiten der GPFS-Policies sowie umfangreiche Möglichkeiten zur Überwachung und zum Management mit TPC. Die Ablage der unstrukturierten Nutzerdaten kann nun durch Einsatz des Datenmanagementsystems zentral erfolgen und das Management der Daten ist weitgehend automatisiert und für den Endnutzer transparent realisiert worden. Die Endanwender aus den verschiedenen Projektgruppen können ihre Daten einheitlich und unter Nutzung der bereits vor Systemeinführung verwendeten Protokolle ablegen.

Im Backend sorgt das implementierte System für eine den Anforderungen der jeweiligen Daten entsprechende Speicherung entweder auf Speicherklassen mit hoher Performance oder in kostengünstigen, kapazitätsorientierten Speicher- klassen. Die deutliche Zeitersparnis und Flexibilität für das Management des Datensystems führt zu höherer Effizienz und gleichzeitiger Kostensenkung.

#### **KONTAKT**

SVA System Vertrieb Alexander GmbH  
Borsigstraße 14  
65205 Wiesbaden  
Tel 06122-536-0  
Fax 06122-536-399  
mail@sva.de  
www.sva.de

© SVA GmbH  
Alle Marken- und Produktnamen sind  
Warenzeichen und werden als solche  
anerkannt.

