WISSEN ▪ TECHNIK ▪ LEIDENSCHAFT

Backup-Konzept v2.4

Zentraler Informatikdienst

2014 - 2018

Datensicherung - Backup Konzept 3

Datensicherung - Backup Konzept

Unternehmensrelevante Daten, die zur Wiederherstellung von Geschäftsprozessen notwendig sind, sind zu sichern.

Der ZID verfügt über ein Backup-System für zentrale Server- und Speicher-Systeme mit dem z.B. auch Institutsserver im Rahmen eines Service-Angebotes gesichert werden können, wenn am Institut keine Möglichkeit einer Datensicherung existiert.

Welche Infrastruktur wird für das Backup verwendet?

Die Datensicherung der Dateiserver erfolgt als **Backup2Disk** auf Hardware der Fa. HP mit einer MSL8096 Bandstation mit zwei LTO6-Laufwerken, und das verwendete Softwareprodukt ist Simpana Version 11 vom Hersteller **CommVault**.

Die Sicherung der virtuellen Serverumgebung erfolgt als **Backup2Disk** auf Hardware der Fa. HP mit dem Softwareprodukt **Veeam** in der jeweils aktuellen Version.

Weiterhin kommen auf einigen Systemen die **Windows-Server Backup-Tools** zum Einsatz.

Die Datenbestände der Linux-Server werden durch **Simpana** gesichert oder speichern die System- und Anwenderdaten mit **Linux-Tools** als tar-flies oder legen mittels **rsnapshot** Generationen von Datensicherungen auf eingebundene Storage-Laufwerke ab. Die Betriebssysteme selbst sind gespiegelt ausgeführt.

Wie wird Datensicherheit erzeugt und wie erfolgt die Datensicherung?

Das mehrstufige Konzept berücksichtigt folgende Kriterien:

- Datenspeicherung durch Arbeitsstationen erfolgen auf Netzwerklaufwerken
- RAID5/6-Festplattensysteme in allen Servern eingesetzt
- Überwachung der Hardware durch Monitor-Systeme
- Storage-Server sind redundant ausgeführt
- inkrementelles Backup forever (täglich)
- Full-Backup am Server (wöchentlich)
- Full-Backup auf Tape (monatlich)
- IT-System-Administratoren können bei Bedarf weitere Sicherungen initiieren
- die Daten bleiben im Sicherungsbestand zumindest 30 Tage erhalten

In der virtuellen Serverumgebung wird aus dem täglichen inkrementellen Backup ein synthetisches Full-Backup auf dem Disksystem erzeugt und dies einen Monat vorrätig gehalten. Dies erfolgt für alle in der virtuellen Infrastruktur betriebenen Produktivsysteme.

Wie sind lokale Daten zu handhaben?

Es sind keine unternehmensrelevanten Daten lokal auf der **Arbeitsstation** (Kleinrechner, Workstation etc.) abzuspeichern, sondern immer ein gesichertes Netzlaufwerk (**Dateiserver**) zu verwenden. Datenverluste bei Schäden an lokalen Festplatten sind dadurch nahezu ausgeschlossen.

Welche Institutionen sind in die BU-Strategie eingebunden?

Die Daten auf den Microsoft-Dateiservern für die Einrichtungen der Verwaltung, das Rektorat und der vom ZID betriebenen Institutsserver werden zentral vom ZID gesichert.

Institute mit mehr als im **Datei- und Backupservice** festgelegten Speicherplatz, haben die Kosten für diesen Mehrbedarf zu tragen.

Für diese Infrastruktur wird auch ein weiteres Disksystem für „fast statische" Daten zur Verfügung gestellt (**Async-Speicher**). Unter dieser Datenart werden Daten verstanden, welche im direkten Zugriff verfügbar sein müssen, sich jedoch gar nicht oder selten ändern (z.B. abgeschlossene Projekte, Meß- & Bilddaten). Hier erfolgt jedoch keine Datensicherung mit dem Backup-System.

Alle Novell-Dateiservern wurden noch bis **Q3 2016** durch das zentrale Backup-System des ZID gesichert; danach wurden alle Server stillgelegt, entsorgt und der Softwarelizenzvertrag nicht verlängert. Andere Institute sind i.d.R. für die Sicherung ihrer Datenbestände auf selbst betriebenen Institutsservern auch selbst verantwortlich.

Institute mit eigener Linux- oder **Microsoft-Infrastruktur** können im Rahmen des **Fileservice**-Angebotes ab Q2 2017 mit dem Backup-System gesichert werden. Die Kosten für das Backup (Lizenzen, Hardware anteilig, Medien) sind von dem Institut zu tragen.

Sonstige Storage-Systeme wie z.B. NetApp können mangels Lizenzen nicht gesichert werden.

Der Zentrale Informatikdienst empfiehlt eine Sicherung der Systemkonfiguration bei jeder Änderung und eine regelmäßige Sicherung der Benutzerdaten über lokal am Institut installierte Backup-Einrichtungen.

Welche zusätzliche Möglichkeit eines Backup gibt es?

Am ZID steht eine Linux-Workstation mit einem LTO5-Laufwerk zur Nutzung zur Verfügung. Die Daten können entweder über das Netzwerk oder von einer externen Festplatte gesichert werden. Die dafür benötigten LTO5-Medien sind von den Nutzern (Institutionen) selbst zu finanzieren und im Vorwege zu bestellen, z.B. über das Büroservice.

Um auch nach längerer Zeit eine Datenwiederherstellung sicherzustellen, sollten die so erstellten Medien fachgerecht gelagert und jährlich umkopiert werden.

Wie werden die Daten von Notebooks gesichert?

Der einfachste Weg ist derzeit die Verwendung einer externen Festplatte; diese gibt es auch abgesichert mit Zahlen-Code oder Finger-Scan.

Im Netzwerk der TU Graz synchronisieren sie die Daten mit den **Fileservices**.

Das kostenlose Softwaretool **SyncToy** der Fa. **Microsoft** kann wertvolle Hilfe den Personen leisten, die mit Windows-Geräten viel außer Haus sind.

Apple-Notebooks sichern sie vorzugsweise mit der integrierten **Time Machine**.

Für **Linux**-Notebooks stehen ein Vielzahl an Programmen zur Verfügung, z.B. **Back-in-Time**.

Welche zentralen Server- und Speicher-Systeme werden vom o.g. System gesichert?

Von allen virtuell betriebenen Server (zB. Nameserver, Ldap, Project, SCCM, Software, vernetztes Lernen) wird das Betriebssystem mit Konfigurationsdateien etc. über die Software **Veeam** gesichert.

Die von diesen Servern verwalteten Datenbestände, die auf Storage-Systemen liegen, werden durch die Software **Simpana** gesichert.

Für nicht virtuell betriebene Server werden die Betriebssysteme (i.d.R. gespiegelt) und Konfigurationsdateien ebenfalls mittels der Software **Simpana** gesichert bzw. es werden Sicherheitskopien mittels Script oder Backup-Tools (z.B. **rsnapshot**) automatisch auf Backup-Laufwerke

übertragen. Datenbank-Server (zB. Galera) führen zusätzlich noch eine Backup-Historie auf den lokalen Festplatten durch.

Derartige physikalischen Server sind die Exchange-/ADM-/Sql-Server, der mySql-Galera-Cluster, die Software-Lizenzserver, die Server FTP/Matlab/ ownCloud/PLUTO/SVN etc., die Hochleistungsrechner bzw. -cluster u.v.m.

Werden E-Mails zentral gesichert?

- Die E-Mails sind auf RAID-Systemen gespeichert; die Exchange-Infrastruktur ist redundant aufgebaut, das **sbox-System** wird mit einem Ersatzsystem (Fallback-System) synchronisiert, d.h. aber auch, dass E-Mails, die von Anwendern gelöscht werden, am Ersatzsystem ebenfalls gelöscht werden!
- Moderne E-Mail-Clients lassen sich i.allg. so konfigurieren, dass beim Löschen E-Mails nicht sofort gelöscht, sondern z.B. zuerst nur in einen Papierkorb verschoben werden; das verhindert unabsichtliches Löschen durch den Benutzer und wird dringend empfohlen.
- Auf den Servern werden auch tatsächlich gelöschte E-Mails eine gewisse Zeit zwischengespeichert, d.h. auch auf irrtümliche Löschung durch Benutzer kann in einem **definierten Zeitfenster** von 90 (**Exchange**) bzw. 14 (**sbox**) Tagen reagiert werden (delayed expunge).
- Am Exchange-System kann der Benutzer die E-Mails selbst wieder herstellen (siehe **Project-X**); Nutzer der **sbox** wie Studierende und Alumni müssen sich an den Postmaster wenden.
- Ein dauerndes echtes Sichern (mit Historie bzw. Generationen) würde die Performance des Systems leider empfindlich stören und ist daher nicht umgesetzt.
- Es kommen täglich tausende neue E-Mails, die von den Anwendern oft sofort gelöscht werden (z.B. Spam) - wozu sollten diese also gesichert werden? Eine gesetzliche Notwendigkeit besteht im Falle der TU Graz nicht.
- E-Mails, die nicht (sofort) gelöscht werden, bleiben oft sehr lange am Server - wozu diese also jedesmal sichern? Eine inkrementelle Sicherung erbringt leider eine sehr zeitaufwendige Daten-restaurierung.
- Das o.g. zentrale Backup-System ist mit einer derart großen Zahl kleiner E-Mail-Dateien überfordert.

Wie sollten E-Mails gesichert werden?

Wir schlagen vor, dass die Anwender sich die E-Mails, die sie gerne gesichert hätten, auf ein lokales oder noch besser auf ein Netzwerk-Laufwerk kopieren, das (z.B. von dem Institutsadministrator) regelmäßig gesichert wird.

- Die meisten E-Mail-Programme bieten die Möglichkeit die E-Mails nach vom Anwender definierten Kriterien automatisch in verschiedene Folder zu verteilen, diese Folder können auch auf einer "lokalen" Platte liegen. Leider sind dann auch relativ aktuelle E-Mails von anderen Rechnern aus nicht mehr zugänglich, außer die „lokale" Platte ist in Wirklichkeit ein Netzlaufwerk, das der Anwender auch von anderen Rechnern aus erreichen kann.
- Eine andere Möglichkeit besteht darin, dass der Anwender in E-Mail-Programmen mit graphischen Oberflächen einzelne E-Mails oder ganze Ordner markiert und dann per Maus in einen anderen Ordner verschiebt, der z.B. auf einem Laufwerk liegt, das regelmäßig gesichert wird.
- Studierende/Alumni können aber auch ganze Ordner in eine Datei sichern, indem sie über **Webmail** einen Ordner aussuchen und diesen z.B. als ZIP-Datei herunterladen. Wenn sie später doch auf einzelne E-Mails dieser Datei zugreifen wollen, dann müssen die Anwender diese Datei entpacken, im E-Mail-Programm einen lokalen Ordner anlegen und die entpackte Datei über die dadurch entstandene leere Datei kopieren; die Daten können dann im E-Mail-Programm wieder bearbeitet werden.

Erfolgt eine Archivierung der Daten?

Der ZID führt keine generelle Datenarchivierung der Datei- und Mail-Server durch. Dies hat bei Bedarf an den Instituten selbst zu erfolgen.

Die Datenbestände des **TUGRAZonline** werden täglich gesichert und regelmäßig archiviert.

Die archivierungsbedürftigen Daten (z.B. **SAP**) der Verwaltungs-einrichtungen erfolgen gemäß gesetzlicher Bestimmungen im Bundesrechenzentrum (**BRZ**) in Wien.

Werden die Hochleistungsrechner gesichert?

Auf allen **zentralen HLR-Systemen** (Linux-Cluster und andere HPC-Nodes) ist eine Datensicherung (inkl. Image der Nodes) eingerichtet, die für eine Systemwiederherstellung vorgesehen ist.

Aufgrund der sehr großen Datenmengen von wissenschaftlichen Berechnungen werden Anwender-Daten jedoch nicht extra durch ein Backup-System gesichert.

Diese Dateien liegen in einem redundanten RAID6-Festplattenstorage. Bei Bedarf ist eine Datensicherung auf andere Speichermedien von den Nutzern selbst durchzuführen!

Werden die Daten des Ceph-Cluster gesichert?

Der Ceph-Cluster ist ein objektorientiertes Storage-System, welcher die Daten z.B. für den **FTP-Server**, die **private Cloud der TU Graz**, den **Matlab-Server** und für **GPU-Nodes** vorhält und weiterhin Laufwerke zur Datensicherung via ceph-mount oder iSCSI zur Verfügung stellt.

Die Storage-Nodes und die Monitor-Server - welche das System überwachen und ggf. ausbalancieren - befinden sich an drei Standorten der TU Graz. Das System ist technisch so ausgestattet, dass es sich bei Ausfällen einzelner Disken oder ganzer Storage-Nodes selbst reparieren kann; die Datenbestände werden im System zweifach repliziert.

Aufgrund der großen Datenmenge ist ein zusätzliches Backup auf andere Medien nicht wirtschaftlich durchführbar und vom zeitlichen Aufwand her derzeit mit den bestehenden Backup-Ressourcen nicht realisierbar.

Die mit dem ownCloud-Client auf die Arbeitsstation(en) synchronisierten Dateien stellen bereits eine Sicherheitskopie dieser Daten dar. Zusätzlich ist hier auch die Versionierung aktiv, so dass jeder Nutzer prinzipiell auf ältere Versionen seiner Daten über das Webinterface zurückgreifen kann.

Für ein *disaster recovery* wird ein tägliches Backup der **ownCloud-Daten** inkl. Versionierung und der **FTP-Serverdaten** exkl. Versieonierung auf einen separaten Datenspeicher erstellt.

Welche Hardware-Redundanzen gibt es?

Da viele Linux-Server, die auf eine gemeinsame Storage-Infrastruktur zugreifen, vom gleichen Typ angeschafft wurden, wird eine Ersatzhardware bereit gehalten, die bei Hardware-Ausfällen im Bedarfsfall kurzfristig in Betrieb genommen werden kann.

Welche Planungen zum BU-System gibt es?

Das 2014 neu angeschaffte System wird kontinuierlich dem Bedarf angepasst, d.h. ggf. Lizenzen oder Hardware-Speicher nachgekauft und gewartet. In 2016 erfolgte ein Software-Upgrade sowie für 2017 ist die Anschaffung eines Media-Agent (Storage-Node) im Rahmen eines Lizenzmodellwechsels vorgesehen, um auch Instituten eine Datensicherung ihrer Infrastruktur zu ermöglichen. Weiterhin ist es dann lizenzrechtlich möglich, die Daten der Backups in einer separaten Hardware - vorzugsweise Ceph-Cluster - zusätzlich abzuspeichern. Für die Zeit nach 2018 ist zeitgemäß ggf. ein neues System anzuschaffen; entsprechender Budgetbedarf ist für das Jahr 2019 bekannt gegeben worden.

Anlagen

1. Aufstellung der gesicherten Linux/UNIX-Clients
2. Aufstellung der gesicherten Microsoft-Infrastruktur