LECTURE 9
BACKUPS

Prepared By:
Amnir Hadachi and Artjom Lind
University of Tartu, Institute of Computer Science
amnir.hadachi@ut.ee / artjom.lind@ut.ee
LECTURE 9: BACKUPS

OUTLINE

1. General rules
2. Backup devices and media
3. Saving space and time with incremental backups
4. Setting up a backup regime with dump
5. Dumping and restoring for upgrades
6. Other archiving programs
7. Multiple files on a single tape
8. Commercial backup products
GENERAL RULES
• Intro:

QUOTE:

In this section we will present a set of rules and suggestion regarding back up. None of the presented suggestion is an absolute rule, however you figure out that more of them you will follow, the easier and smoother your task becomes regarding storing and backups task.
LECTURE 9: BACKUPS
GENERAL RULES

• We can categories the suggestion for backups as follows:

  - Perform all backups from a central location
  - Label your media
  - Pick a reasonable backup interval
  - Choose filesystems carefully
  - Make daily dumps fit on one piece media
  - Keep media off-site
  - Protect your backups
  - Limit activity during backups
  - Verify your media
  - Develop media life cycle
  - Design your data for backups
LECTURE 9: BACKUPS

GENERAL RULES

Perform all backups from a central location

• make all dumps from one machine
• `rdump` command allows you to perform remote dumps over the network.

Label your media

• for clear organisation, with unique id to identify their content
Lecture 9: Backups

General Rules

- Pick a reasonable backup interval
  - Reduce loss of data by making more often backups
  - Frequency of backups:
    - Heavy system every working day
    - Light system every week

- Choose filesystems carefully
  - Take care of backing up /etc/passwd everyday and store it in other partition.
  - No need for backing up /tmp directory.
GENERAL RULES

Make daily dumps fit on one piece media

• Use the same tape for all dumped filesystem of specific day.

• To minimise the impact with users: mount tape every day before you leave work and run dunks late at night from cron.

keep media off-site

• To protect the data from any physical damage keep backups copy off-site.

• secure and keep the storage facility in climate-controlled environment.
GENERAL RULES

LECTURE 9: BACKUPS

Protect your backups
- Locks and keys.

Limit activity during backups

Verify your media
- Monitoring backups procedure and functionality
- After each dump run a verification on the tape
- Run `restore t` to generate a table of contents for each filesystem and to store results on disk.
- Use `grep` to look up the filename and pick the newest instance.
LECTURE 9: BACKUPS

GENERAL RULES

- Develop media life cycle
- Design your data for backups
  - Rely on the following information when designing your storage architecture:
    - which type of data you will be dealing with
    - expected volatility of each type of data
    - backup frequency appropriate for your facility
- Be ready for the worst
2.
BACKUP DEVICES AND MEDIA
LEcTure 9: Backups
backup devices and media

• Choosing backup media and device
  • Several media and devices
    • each has advantages and disadvantages
  • E.g.
    • Optical media
    • Tape backups
    • External disk drives
    • Network backups
• Optical media
  • Media storing digitised data
  • Uses laser to write data, read data
  • E.g.: CDs, DVDs

• Requirement:
  • Recordable CD or DVD drive, software utility
  • CD-R (compact disc-recordable) / CD-RW (650 MB)
LECTURE 9: BACKUPS
BACKUP DEVICES AND MEDIA

• Optical media
  • CD backups
    • simple to restore from standard format.
    • Relatively low storage capacity
  • Recordable DVD
    • 4.7 GB on one single-layered side
    • Double-layered, two sides DVD
      • store up to 17 GB of data
    • Several different formats.
• Optical media

• Disadvantage:
  • Writing data takes longer than other media
  • Requires more human intervention
Tape Backups

The action of copying data to magnetic tape

simple

huge capacity of storage

Requirement:

Tape drive connected to network

management software

Backup media
Tape backups, Cases:

- Small network
  - stand-alone tape drives attached to each server
- Large Network
  - One large centralised tape backup device
- Extremely large
  - Robots retrieve, circulate tapes from vault (Tape storage library)
Lecture 9: Backups

Backup Devices and Media

- External disk drives
  - Removable disk drives
    - Attached temporarily to computer
    - Simple to use
    - Handling large amount needs:
      - Backups control features, higher storage capacity, faster read-write access.
• Network backup
  • process of saving data to another place or network
    • different server, another WAN location
    • w.g. NAS storage device
  • Online backup
    • saves data across internet
    • need for strict security measures
    • automated backup, restoration processes
    • provides also evaluation (e.g. test speed, security, recovery,...)
3. SAVING SPACE AND TIME WITH INCREMENTAL BACKUPS
All backup tools support at least two types of backups:

- Full backups
- Incremental backups
Lecture 9: Backups

Saving Space and Time

- All backup tools support at least two types of backups:
  - Full backups
  - Incremental backups

- Full backups:
  - All file systems
  - All the data and files in selected folders

- Incremental backups:
  - Only new or modified files and folders

- Differential backups:
  - All data since the last full

- Mirror backups:
  - Only new modified files and folders
# LECTURE 9: BACKUPS

**SAVING SPACE AND TIME**

<table>
<thead>
<tr>
<th>Backup type</th>
<th>Backup time</th>
<th>Restore time</th>
<th>Storage space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full backup</td>
<td>slowest</td>
<td>fast</td>
<td>high</td>
</tr>
<tr>
<td>Incremental backup</td>
<td>fast</td>
<td>moderate</td>
<td>lowest</td>
</tr>
<tr>
<td>Differential backup</td>
<td>moderate</td>
<td>fast</td>
<td>moderate</td>
</tr>
<tr>
<td>Mirror backup</td>
<td>fastest</td>
<td>fastest</td>
<td>highest</td>
</tr>
</tbody>
</table>
BOOST technique

Deduplication:

- process of discovering similar/identical files or data even if they are located in different directory. This way we ensure only one copy is kept on the disk.

- very helpful for limiting the size of backups.
SCHEDULING BACKUPS

In order to build the right schedule for your backups you have to take into account:

- Activity of your filesystems
- Capacity of your dump device
- Amount of redundancy you want
- Number of tapes you want to buy
LECTURE 9: BACKUPS
SAVING TIME AND SPACE

- Design a schedule
- Simple schedule
  - Do level 0 dumps of everyday
  - Reuse a group of tapes for very N days
  - Keep the tapes forever

Cost:
\[(365/N) \times \text{(tape price)}\]
LECTURE 9: BACKUPS
SAVING TIME AND SPACE

• Design schedule

• moderate schedule

<table>
<thead>
<tr>
<th>Week 1</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 2</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 3</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 4</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Days level 9 or 3 dump
Week level 5 or 2 dump
Month level 3 or 1 dump
4. SETTING UP A BACKUP REGIME WITH DUMP
To create a backup and restore points we need `dump` and `restore` commands.

Let’s imagine we choose an incremental regime thus:

- **Dump filesystem**: build up a list of files that have been modified or added since the last dump, then pack them into a single file to achieve to an external storage.

**Limitations:**
- **Individual dumping**
- Only local filesystems can be dumped but you can fix that by using `rdump` command to do it on remote location.

```
#rdump -fmachine1:/dev/rmt0 -c /usr
```

### `rdump` options
- `-b Blocks`
- `-B`
- `-c`
- `-d Density`
- `-L Length`
- `-s Size`
- `-u`
- `-w`
- `-W`
- `-level`
- `-f` Machine:Device
- `-n` Filesystem/DeviceName
• Dump arguments and functionality:

  #dump d s

• d specify the tape density in byte per inch, and s specify the size in kilobytes.

EXAMPLE:

  #dump 5usdf 60000 6250 /dev/rst0 /work

  #dump [flag] [size] [density] [f:send output to this tape device] [filesystem]

• dump has many argument and flags but it does not change that much from a platform to another. (you should check the manual)
LECTURE 9: BACKUPS
SETTING UP A BACKUP REGIME

• Features:

  • Incremental regime:

    • no restriction reading the length of the filenames. (e.g hierarchies can be arbitrarily deep and long name and they are handled correctly)

  • dump arguments:

<table>
<thead>
<tr>
<th>Flag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U flag</td>
<td>make dump to automatically update /etc/dumpdates</td>
</tr>
<tr>
<td>F flag</td>
<td>make dump send its output elsewhere</td>
</tr>
<tr>
<td>D/s flag</td>
<td>to specify the tape density in bytes per inch and tape length in cubits</td>
</tr>
</tbody>
</table>
LECTURE 9: BACKUPS
SETTING UP A BACKUP REGIME

• Restoring from dump
  • Extracting data from dump tapes
  • Two types of restoring:
    • Restoring a set of files
    • Restoring entire filesystem
• **Restoring individual files:**
  • you have to determine which tape contain the version of files user wanted.
  • Create a temporary directory
  • To restore uses arguments `i`, `ls`, `cd` and `pwd`
  • to mark the files you want to restore use `add` command
  • After selecting the files, use `extract` to pull files from tape.

  *Example: looking for restoring a file: `/users/janet/jamlost` from a remote tape.*

```
$ sudo mkdir /var/restore
$ cd /var/restore
$ sudo ssh tapehost mt -f /dev/nst0 fsf 3
$ sudo restore -i -f tapehost:/dev/nst0
restore> ls
.:
janet/ garth/ lost+found/ lynda/
restore> cd janet
restore> ls
afile bf ile cfile iamlost
restore> add iamlost
restore> ls10
afile bf ile cfile iamlost*
restore> extract
start with the last volume and work towards the first.
Once the restore has completed, give the file to janet:
$ cd /var/restore
$ ls janet
iamlost
$ ls ~janet
afile bf ile cfile
$ sudo cp -p janet/iamlost ~janet/iamlost.restored
$ sudo chown janet ~janet/iamlost.restored
$ sudo chgrp staff ~janet/iamlost.restored
$ cd /; sudo rm -rf /var/restore
$ mail janet
```
Restoring entire filesystem:

First... Check what caused the filesystem to fail
Second... Create and mount the treated filesystem
Third... Go to the mount point of the new filesystem using `cd`
Fourth... Select the first most recent level 0 dump tape and put it in the tape drive
Fifth... Execute `restore r` command
Sixth... Mount and restore the incremental dumps chronologically
5. DUMPING AND RESTORING FOR UPGRADES
Before doing any OS upgrade, you should do a back up of all file system with level 0 dump and restore them.

Next you have to back up and restore any system specific files that are in "/" or "/usr". (e.g. /etc/passwd, /etc/shadow, ...).

Then you have to do a complete setoff level 0 dumps immediately after an upgrade.
6.

OTHER ARCHIVING PROGRAMS
LECTURE 9: BACKUPS

ARCHIVING

• Other archiving program:
  • Program Tar (tape archive) is useful for archiving and transmitting files or data.
  • Program dd is a file copying and conversion, it literally just copy files its input file to its output file.
• Tar:

• Reads multiple files or directories and packages them into one file, often a tape device.

EXAMPLE:

```
sudo tar -cf - fromdir | ( cd todir ; sudo tar -xpf - )
```

Create a copy of the directory tree fromdir in todir
**LECTURE 9: BACKUPS**

**ARCHIVING**

- **Tar**

<table>
<thead>
<tr>
<th>Options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>-c</code></td>
<td>create a new archive</td>
</tr>
<tr>
<td><code>-d</code></td>
<td>compare the files store interfile with others</td>
</tr>
<tr>
<td><code>-r</code></td>
<td>append other files to the end of the existing archive</td>
</tr>
<tr>
<td><code>-t</code></td>
<td>list the names of files stored in archive</td>
</tr>
<tr>
<td><code>-u</code></td>
<td>update</td>
</tr>
<tr>
<td><code>-x</code></td>
<td>extract or get</td>
</tr>
</tbody>
</table>
LECTURE 9: BACKUPS
ARCHIVING

- dd:
  - Copies its input files to as output files

INFO:
If a user brings you a tape that was written on a non-UNIX system, dd may be the only way to read it.

You can also use dd to make a copy of a magnetic tape. With two tape drives: /dev/st0 and /dev/st1

```
$ dd if=/dev/st0 of=/dev/st1 cbs=16b
```

With one drive (/dev/st0):

```
$ dd if=/dev/st0 of=tfile cbs=16b
/* Change tapes. */
```

```
$ dd if=tfile of=/dev/st0 cbs=16b
$ rm tfile
```
7. MULTIPLE FILES ON A SINGLE TAPE
LECTURE 9: BACKUPS
MULTIPLE FILES ON A SINGLE TAPE

• Magnetic tape:

INFO:

magnetic tape contains one long string of data, but it has some structure.

When dump or some other command writes a stream of bytes out to a tape device and then closes the device file, the driver writes an end-of-file marker on the tape.

consequence

the stream are separated with marker EOF it is used for automatic reading
LECTURE 9: BACKUPS
MULTIPLE FILES ON A SINGLE TAPE

- Magnetic tape:
  - to position the tape on a specific stream or file set you use `mt` command.

  ```
  #mt [-f tapename] command [count]
  ```

  - `-f`
    - `forest`
  - `-rew`
    - rewinds the tape to the beginning
  - `-offl`
    - put the tape offline. Some dump tape scripts use this to eject the tape.
  - `-status`
    - prints information about the current status of tape drive
  - `-fsf`
    - fast-forwards the tape. [count] is default to 1.
  - `-bsf`
    - backspace count files. If you are too far forward, your best bet is to rew it and start again from beginning
8.
COMMERCIAL BACKUP PRODUCTS
LECTURE 9: BACKUPS
COMMERCIAL BACKUP PRODUCTS

- ADSM / TSM (Tivoli Storage Manager):
  - Developed by IBM and purchased by Tivoli.
  - Pros:
    - Very low failure rate
    - reasonable price and excellent documentation
  - Cons
    - none friendly interface
    - design is incremental forever
LECTURE 9: BACKUPS
COMMERCIAL BACKUP PRODUCTS

• EMC NetWorker:
  • Owned by Storage Behemoth after they acquired Legato Networker.
  • It is a program for backup management and it centralises, automates, and accelerates data backups and recovery.

• Pros
  • Competitive price
  • Support diverse client platforms

• Cons
  • Significant overlap among EMC products
LECTURE 9: BACKUPS
COMMERCIAL BACKUP PRODUCTS

- Backups as a service (Cloud):

<table>
<thead>
<tr>
<th>Product</th>
<th>Storage Space</th>
<th>Backup multiple machines</th>
<th>File versioning</th>
<th>Web Access to Backups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mozy home</td>
<td>2 GB</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>SugarSync</td>
<td>2 GB</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Dropbox</td>
<td>2 GB</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>iDrive</td>
<td>2 GB</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Live Skydrive</td>
<td>25 GB</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>ZumoDrive</td>
<td>2 GB</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
# LECTURE 9: BACKUPS

## COMMERCIAL BACKUP PRODUCTS

- **Cloud backup options:**

<table>
<thead>
<tr>
<th>Options</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managed service provider (MSP)</td>
<td>• Simplicity</td>
<td>• Complete dependency on the MSP for all the aspects of the backups</td>
</tr>
<tr>
<td></td>
<td>• Cost-effective</td>
<td>• Control is handed to MSP</td>
</tr>
<tr>
<td>Cloud enabled backup applications</td>
<td>• Extends and supplements existing backup infrastructure</td>
<td>• Requires a cloud-enabled backup application</td>
</tr>
<tr>
<td></td>
<td>• Except for backup data location, control remains with the customer</td>
<td>• Little impact on IT resource requirements</td>
</tr>
<tr>
<td>Cloud gateways</td>
<td>• Works with any backup application that supports backup to disks</td>
<td>• Introduction of an additional IT infrastructure competent</td>
</tr>
<tr>
<td></td>
<td>• Extends and supplements existing backup infrastructure</td>
<td>• Higher cost</td>
</tr>
<tr>
<td></td>
<td>• Except for backup data location, control remains with the customer</td>
<td></td>
</tr>
</tbody>
</table>