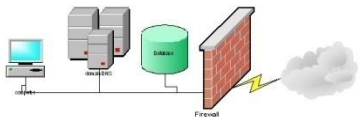
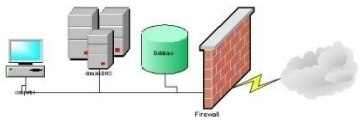
Management of Information Systems

Back-up and Restore

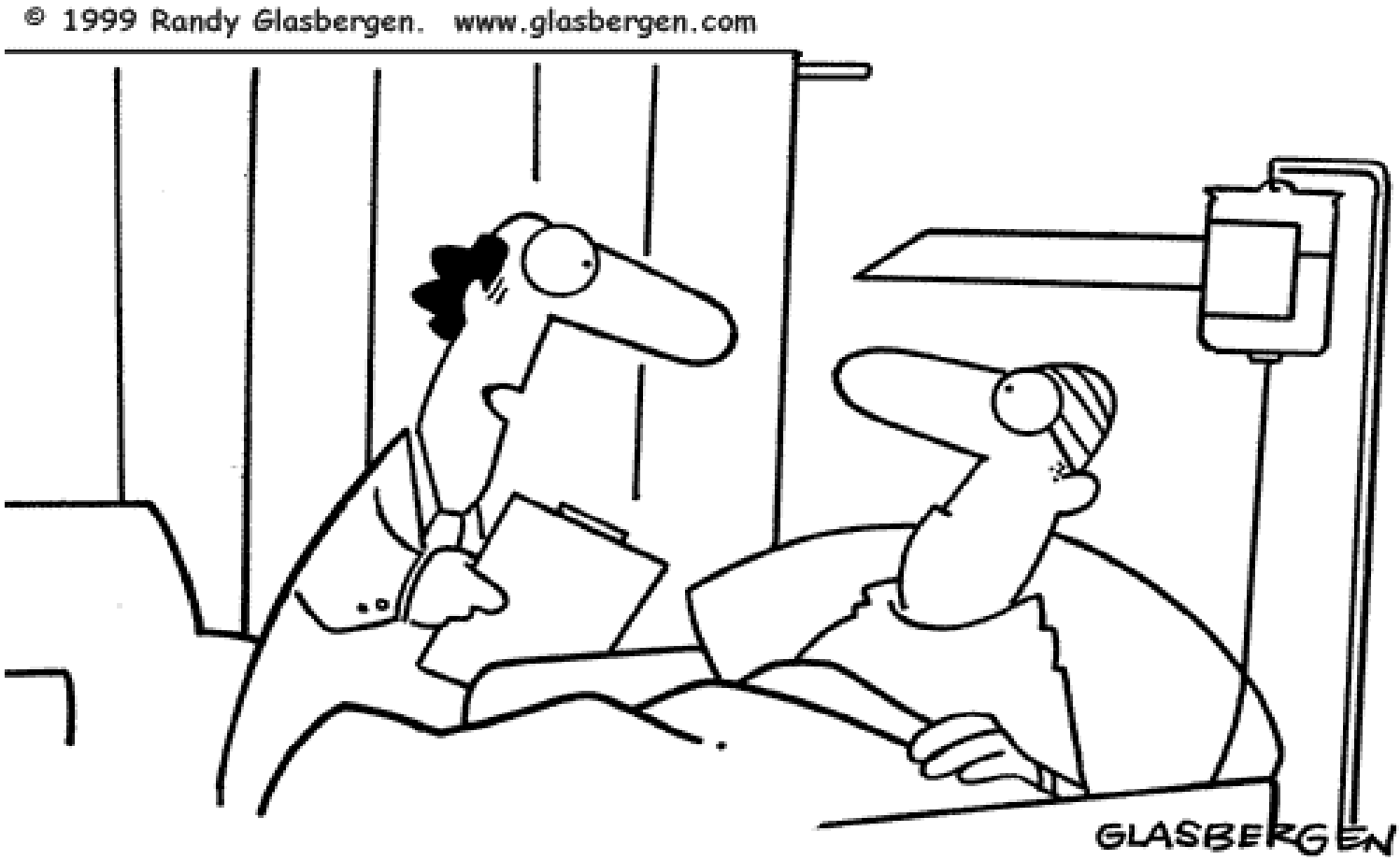


Back-up and Restore

- Back-up / archiving definition
- Back-up
  - Tape
  - Back-up systems
  - Back-up methods
    - Full, incremental, differential, progressive
- Archives
  - Archivation requirements
- Design of back-up
  - Example
- Restore



Back-up and Archive



"You caught a virus from your computer and we had to erase your brain. I hope you kept a back-up copy."

Back-up and Archive

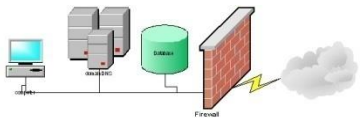
Goal of back-up / archiving: recovery safeguard, make copies for preventing data loss

Goal of back-up: guarantee business continuity

- Delete: the user accidentally / intentionally delete/overwrite
- Failure: storage device/system fails

Goal of archiving: to reproduce the entire data of a certain point in time
  - with a granularity of a quarter, half, or full year
- Business, legal (determined by law) etc. reasons: data serves as an evidence, basis for a comparision, reference
- Not used data have to be deleted: maintenance / evidence(!!) / legal(!!) reasons

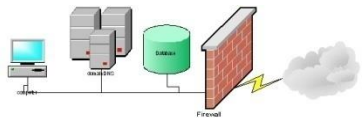
Tipically common base technology for back-up/archiving



Needs for Restoration

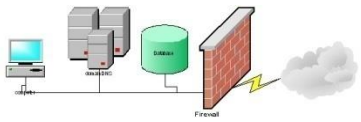
- Accidental file deletion
- Disk failure

- Archives



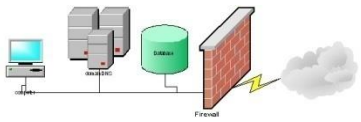
Accidental File Deletion

- User requirement: quick restoration
- Typical business environment:
  - Max. 1 day earlier state can be restored
  - Takes for max. 1 day to restore
  - New SW-s: users can restore on their own without the system administrator
    - But only if the tape is still in jukebox



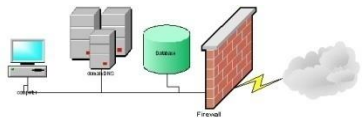
Disk Failure

- Disk failure
  - Disk hardware failure
  - Other HW / SW failure, that causes the complete loss of disk content
- Main consequences
  - data loss
  - service interrupt
- Slow, typically several GB/TB to restore
  - Restore in several steps
    - Full back-up
    - Incremental/Differential back-ups



Archives

- Goal is different
- On technical level
  - ALWAYS full back-up
  - Archive tapes must be treated separately from „normal” back-ups – duplicating, off-site storing
  - Store at different places
  - Long life (~10 years) so not only the tapes to keep, but
    - Devices making/reading archives
    - Tools (programs)



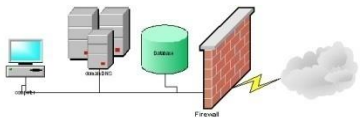
Why to use tapes?

Advantages:

- Cheap
- Tapes can be taken out – no continuous mechanical stress
- Long life – 30 years
- Stored data can easily be deleted

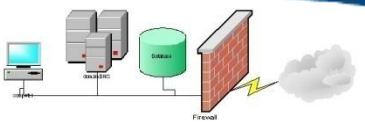
Disadvantages:

- Serial data access (slow)
- Tape insertation time
- More vulnerable



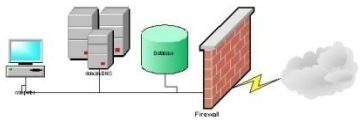
Linear Tape-Open (LTO) standard

- *Tape Format Standard*: IBM, HP and Certance (Seagate) cosortium
  - Open system standard technology
  - 7th generation
  - Backward compatibility
- Since coming out (2000) wide range of industrial acceptance, leading tape technology
- Actual technology: Ultrium 7
  - 2000: 100 GB
  - 2015: 6 TB (v7)
  - Since v3:WORM
  - New releases under standardisation

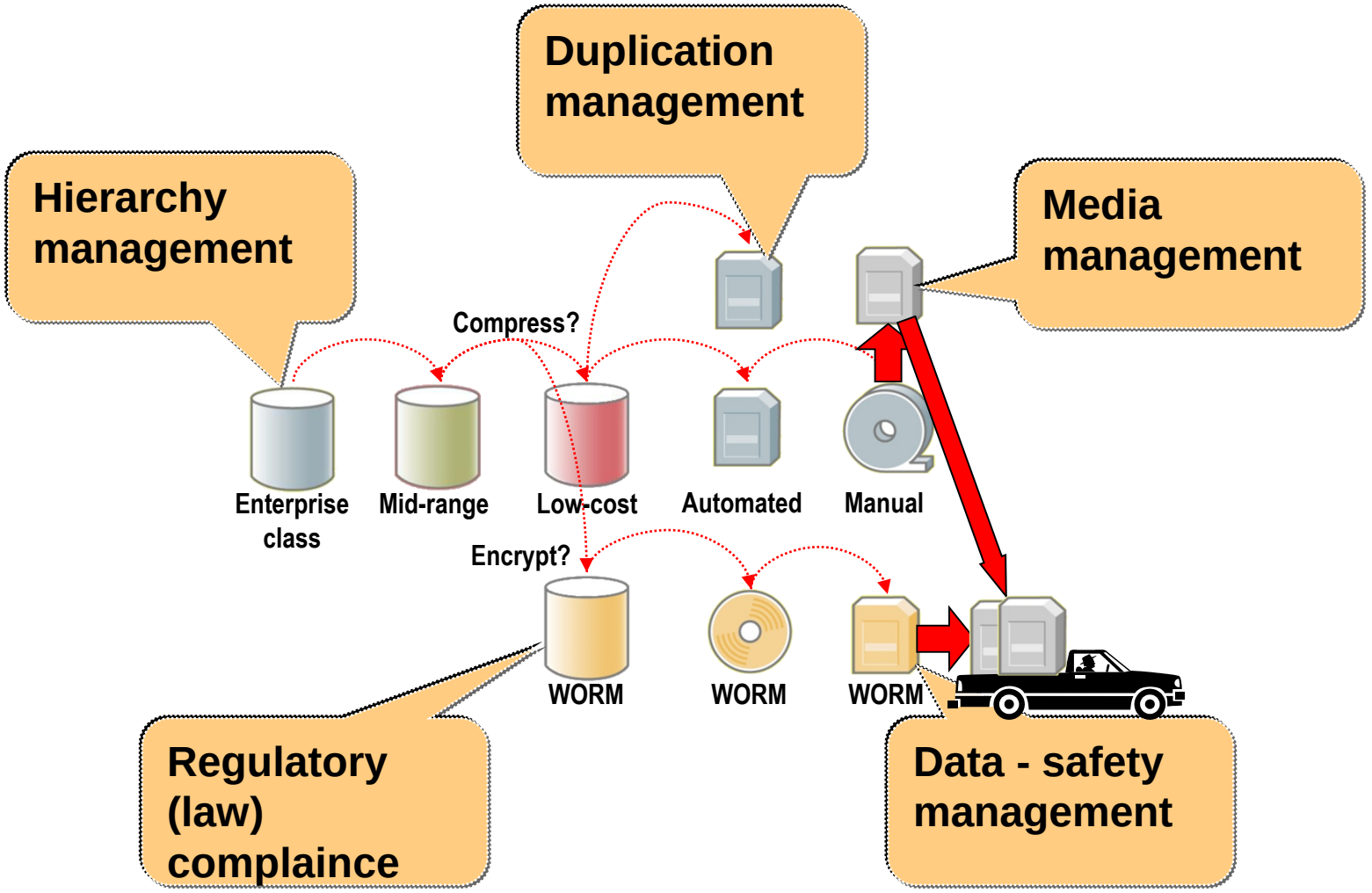


Linear Tape-Open (LTO) features

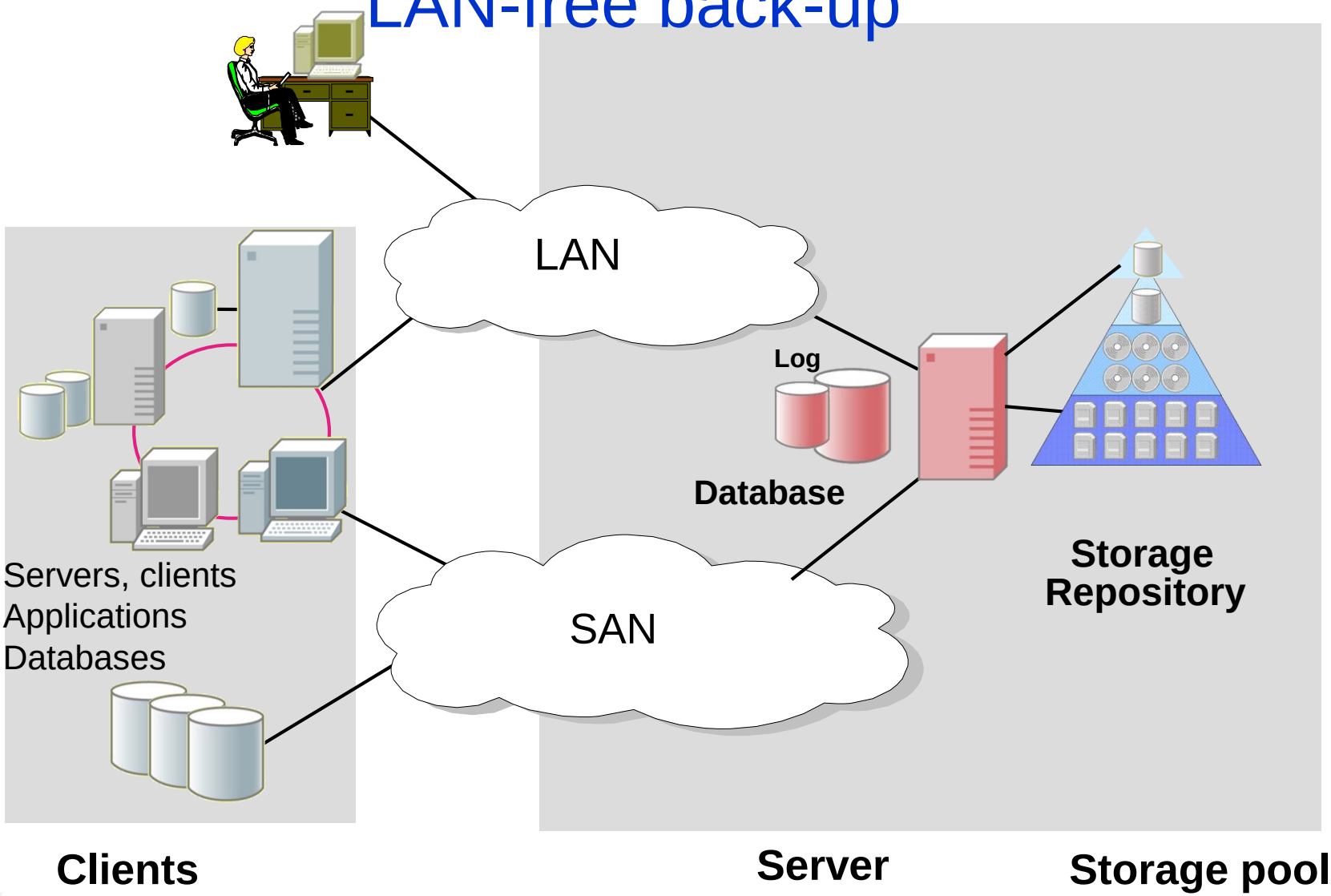
Type	Year	Capacity	Entire-tape reads/ writes	Approximate years of life assuming one tape filled...	
				per *month*	per *week*
LTO-1	2000	100 GB	200	17	4
LTO-2	2003	200 GB	250	21	5
LTO-3	2005	400 GB	364	30	7
LTO-4	2007	800 GB	200	17	4
LTO-5	2010	1.5 TB	200	17	4
LTO-6	2012	2.5 TB	—	—	—
LTO-7	2015	6.0 TB	—	—	—



Modern Storage System Requirements

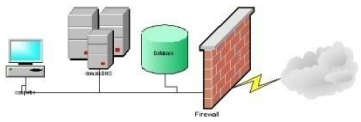


Storage System Architecture
LAN-free back-up



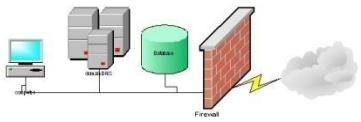
LAN-Free Back-up and Recovery

- LAN-Free client data transmission
  - Server manages the internal storage pool
  - The client moves the data from disk to tape, or to a SAN disk
  - Meta-data are moving on LAN network
  - LAN is not overloaded by heavy storage data traffic
  - Scalable



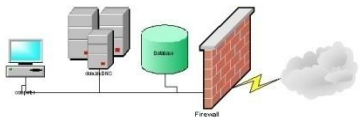
Back-up Methods

- Full back-up
- Incremental back-up
- Differential back-up
- Progressive Back-up Methodology



Full Back-up

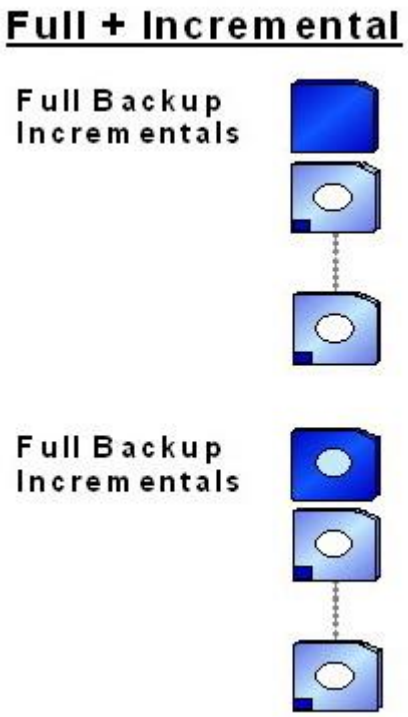
- Save the whole disk content every day
  - Huge amount of data
  - Slow
  - Low utilisation of tapes
    - Saved many times, even if not changed
- BUT:
  - Fast restore (one tape)



Incremental Back-up

- Full back-up only on the first day
- Then only the changes since the previous day
  - Small amount of data

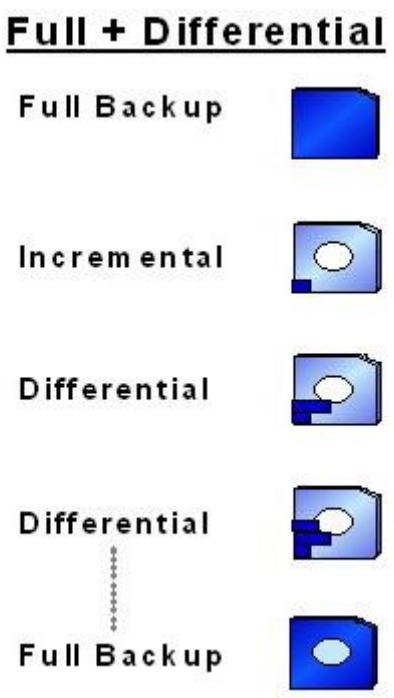
  BUT:
  - Restore is slow
  - Low utilisation of tapes
    - almost empty



Differential Back-up

- Full back-up only on the first day
- Then only the changes since full back-up
  - Greater, constantly groving amount of data

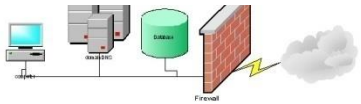
  BUT:
  - Shorter restore time (max. 2 tapes)
  - More tapes



Problem of Incremental / Differential Back-up

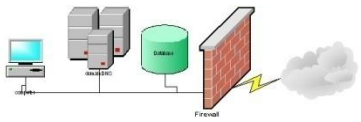
Day 1	Day 2	Day 3	Day 4	Day 5
File A	File A renamed to File F	File F	File F	File F deleted
File B	File B deleted			
File C	File C renamed to File G	File G	File G	File G
File D	File D moved to new location	File D (new location)	File D deleted	
File E	File E	File E	File E	File E

Files from Day 1 FULL backup		Files from Day 3 INCREMENTAL / DIFFERENTIAL backup		Hard Drive after a restore to Day 3
File A		File F		File A – **wrong** File F
File B	+		=	File B – **wrong**
File C		File G		File C – **wrong** File G
File D		File D (new location)		File D – **wrong** File D (new location)
File E				File E



Progressive Back-up Methodology

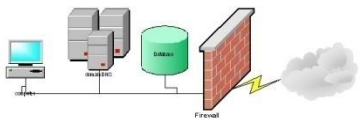
- Full back-up only once
- Then only incremental
- But back-up the file system, too
  - A bit more(!) more to save
- But at restore, we can find the actual state of a file (backward search)
  - Much faster when restore files that
    - modified several times
    - deleted



Advantage of Progressive Back-up Methodology

Day 1	Day 2	Day 3	Day 4	Day 5
File A	File A renamed to File F	File F	File F	File F deleted
File B	File B deleted			
File C	File C renamed to File G	File G	File G	File G
File D	File D moved to new location	File D (new location)	File D deleted	
File E	File E	File E	File E	File E

Required files from Day 1 FULL backup		Required files from Day 2 & Day 3 INCREMENTAL backups		Hard Drive after a restore to Day 3
	+	File F	=	File F
		File G		File G
		File D (new location)		File D (new location)
File E				File E

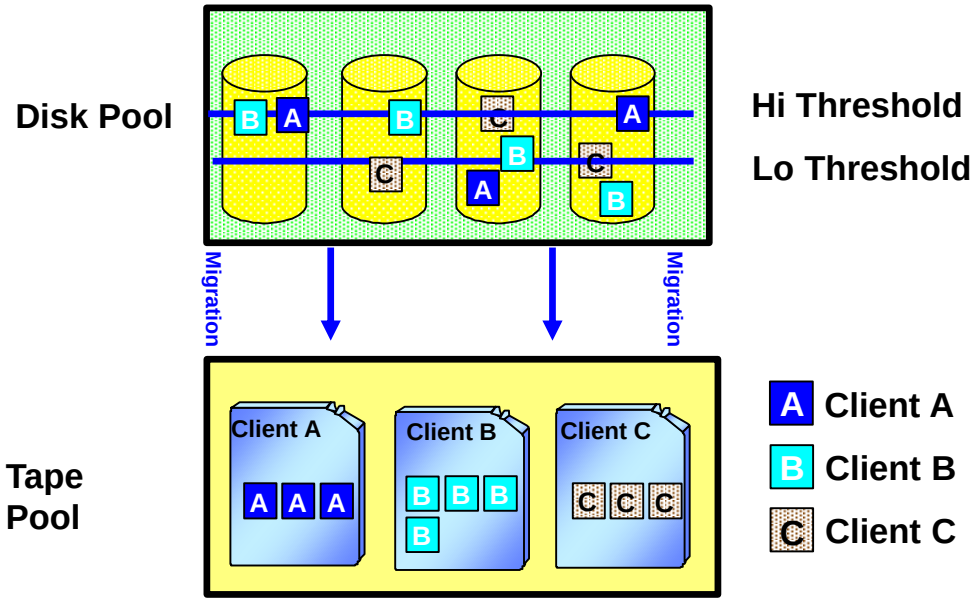


Colocation and Tape Reclamation

Colocation

Data of a client (group) to the same tape

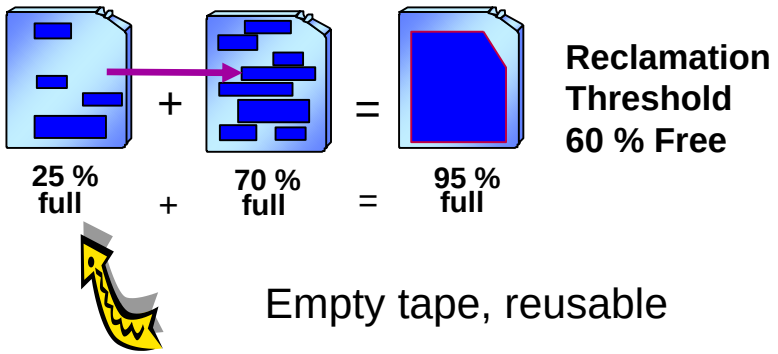
Shorter restoration, fewer tape exchange



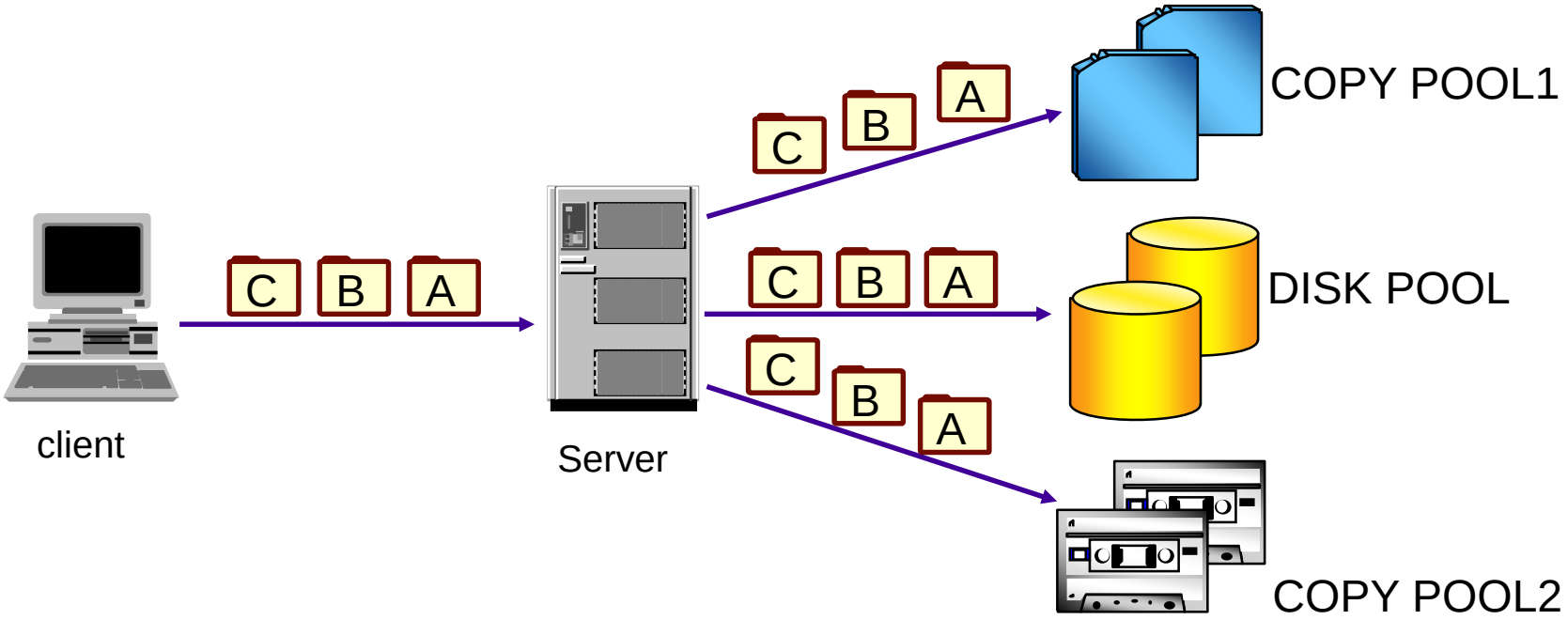
Tape Reclamation

Copy valid data to a new tape after a user-specified threshold

This copying can be timed



Parallel Back-up

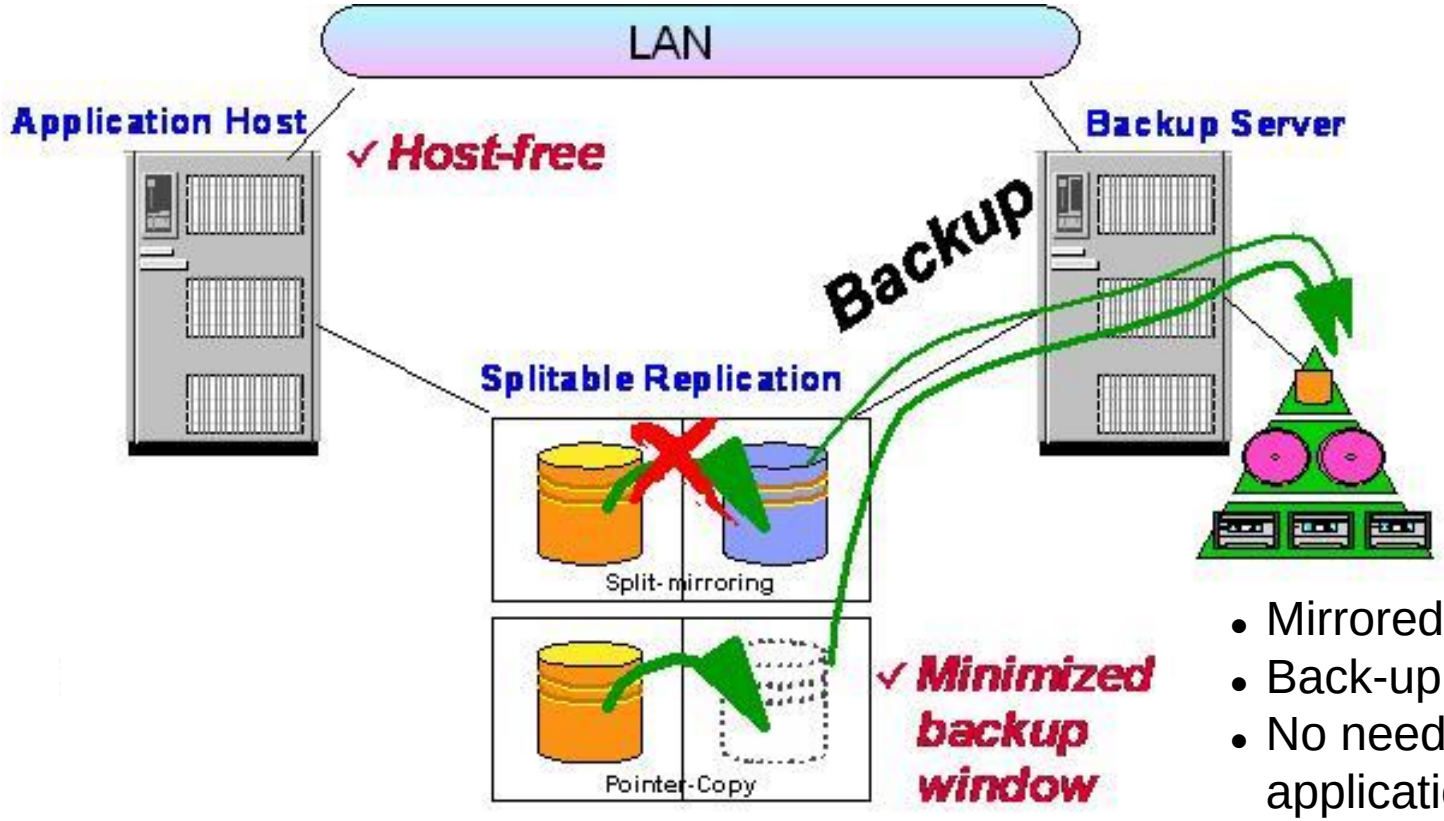


More back-up storage pools can be defined, simultaneous writing

The target storage pools can be of different types (tape, disk).

Disaster tolerant systems

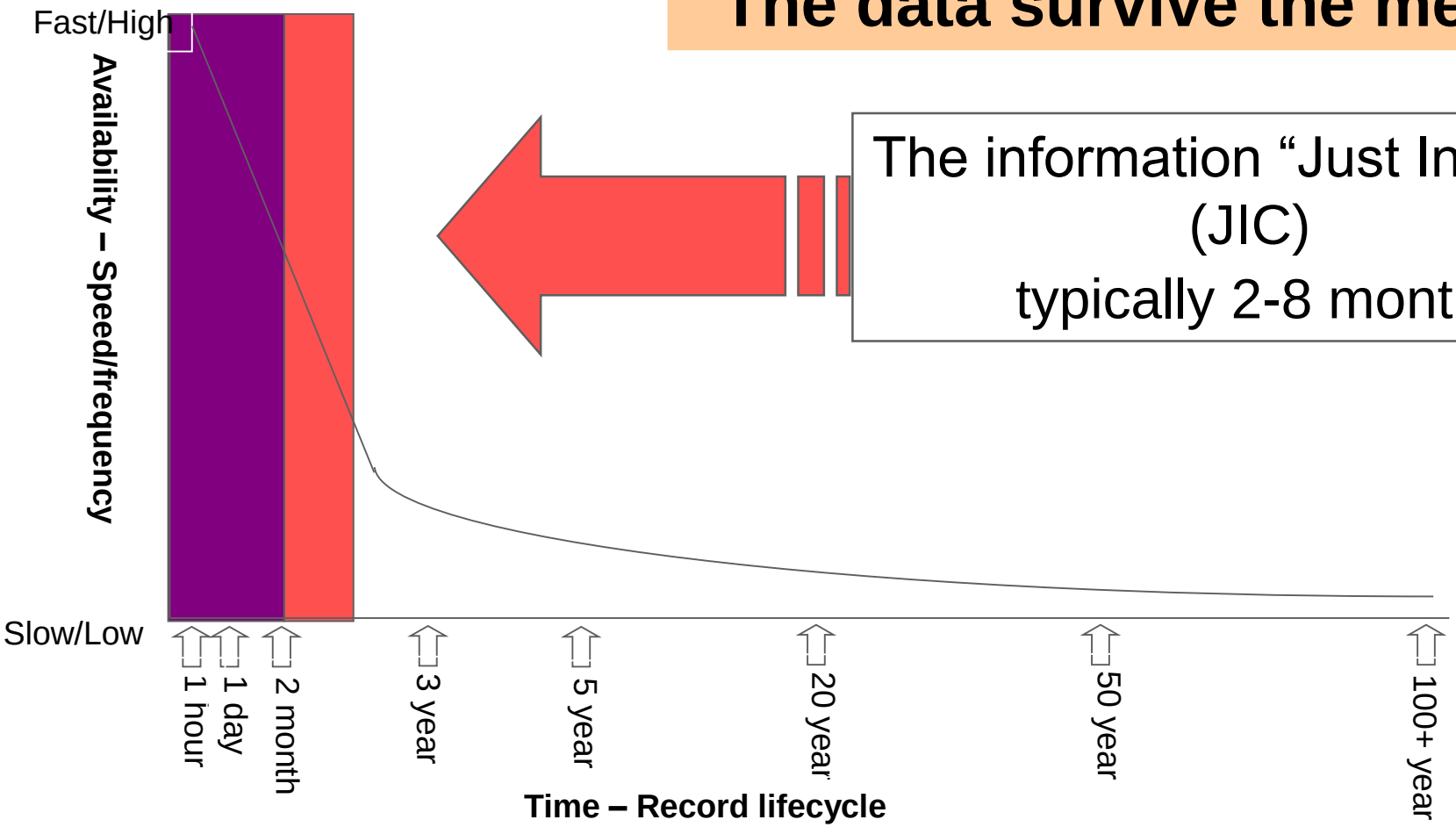
Zero down-time storage



- Mirrored volume
- Back-up from a mirror
- No need to stop the application for a long time

The JIC Information

The data survive the media!



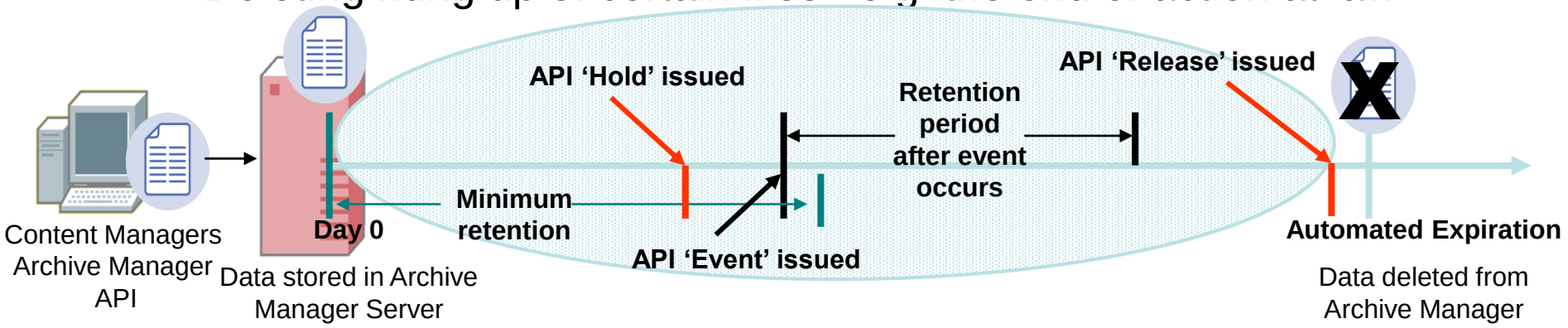
The information “Just In Case” (JIC)
typically 2-8 month

Special Archiving Requirements

Protect data from deletion till a predefined time / but deletes immediately after it

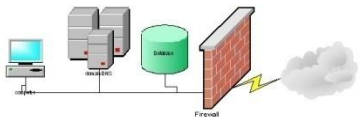
How to reach archiving functions

- – Chronological Retention time policy:
  - • Store the objects for a predefined time – e.g. 3 years –
- –Event-based retention protection:
  - • The storage time depends on an event – e.g. 70 years after client death
- – Deletion Suspend / Resume:
  - • Deleting hung up of certain files– e.g. the end of action at law



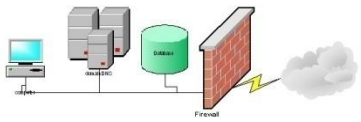
Back-ups must be planned

- Not enogh: „start at midnight”
  - Several back-up types!
  - *Back-up window* should not be the same
  - Back-up always reduces the performance of the system
  - Do at off-peak hours
    - But when are the peak hours?
  - Back-up outsourcing can lead to problems



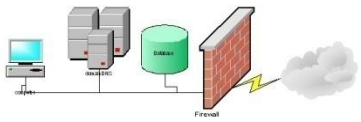
Plan of Back-up

- Corporate Guidelines
- Service Level Agreement (SLA)
- Back-up and Restore Policy
- Back-up Schedule



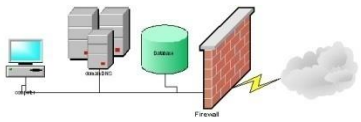
Corporate Guidelines

- Valid for the whole company
- Defines terminology and dictates requirements for data-recovery systems
  - based on legal requirements
  - types of data to back-up
- Do not deal with every specific implementation detail



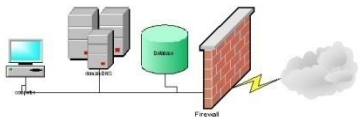
Determine the SLA

- An SLA is a written document that specifies what kind of services and performance are to be provided
- Created with involving the customers
- Defines:
  - types of back-up
  - requested restoration times
  - how often to back-up
  - how long to keep back-ups
  - back-up windows



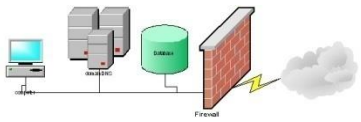
SLA Example

- Customers shall be able to get back any file
  - with a granularity of 1 business day for the past 6 months
  - with a granularity of 1 month for the last 3 years.
- Disk failures shall be restored in 4 hours, with no more than 2 days of lost data
- Archives shall be full back-ups on separate tapes generated quarterly and kept forever
- Critical data will be stored on a system that retains user-accessible snapshots made every hour from 7 AM until 7 PM, with midnight snapshots held for 1 week
- Databases and financial systems shall have higher requirements that shall be determined separately



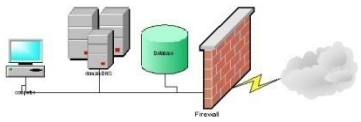
Back-up Policy

- When SLA approved, determine the policy how to achive the requirements
- Typically obvious:
  - From SLA Example:
    - Daily back-ups
    - Tapes will be retained as specified in SLA
    - The policy determines how often full versus incremental back-ups will be performed



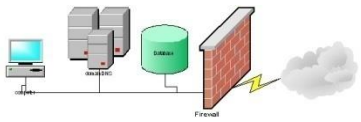
Back-up Schedule

- The back-up schedule lists details down to which partitions of which hosts are to be backed-up and when
- SLA changes rarely, but back-up schedule often
- Typically not written – stored in the configuration of the back-up system



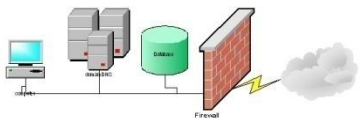
SLA Example - Again

- Customers shall be able to get back any file
  - with a granularity of 1 business day for the past 6 months
  - with a granularity of 1 month for the last 3 years.
- Disk failures shall be restored in 4 hours, with no more than 2 days of lost data
- Archives shall be full back-ups on separate tapes generated quarterly and kept forever
- Critical data will be stored on a system that retains user-accessible snapshots made every hour from 7 AM until 7 PM, with midnight snapshots held for 1 week
- Databases and financial systems shall have higher requirements that shall be determined separately



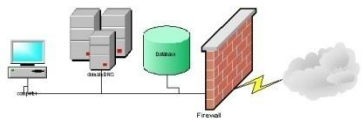
Preparation of Back-up Schedule

- Snapshots made by the system automatically
- Granularity of back-up is 1 business day
- Disk restore requirement 2 days – back-up needed not only on business days
- Since full back-ups take significantly longer than differentials, schedule them for Friday night and let them run all the weekend.
- Sunday through Thursday nights differential back-ups are performed



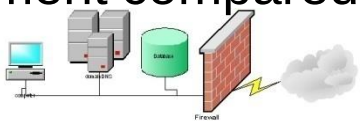
Preparation of Back-up Schedule

- Today's back-up systems are automated
  - It is common to simply list all partitions that need to be backed-up
  - Software generates a schedule based on the requirements
  - The back-ups are performed automatically, and email notification is generated when tapes must be changed
- But when to perform full back-up and when differential?



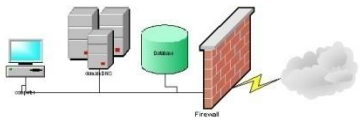
Back-up Schedule Example

- Size of a partition: 4GB
- Full back-up to make: every 4 weeks (28 days)
- Suppose, size of differential back-up grows by 5% every day
  - Day1: Full back-up, 4 GB
  - Day2: 200 MB
  - Day3: 400 MB, etc.
  - Day10: 2 GB
  - Day 11: 2.2GB
  - Just these two days require more than a full back-up
  - Worth perform full back-ups at every 10 day!
  - But full back-up shall be on week-ends: 7, 14, 21, 28 day cycles possible
  - In this example, 7 day (30%), 14 day (40%), other (~60%) tape requirement compared to everyday full back-up



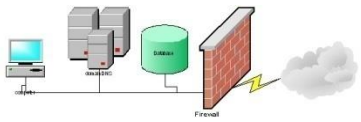
Back-up Schedule Example 2.

- Previous example is not typical
- „80/20 rule”: 80% of accesses reach only 20% of data
- Suppose: 20% of data is accessed and half of it (10%) is modified daily
- First differential back-up: 10% of the partition
- Grows only by 1% daily (10% of 10%)
- Optimal 14 day cycle (~37%), 21 day (39%), 28 day (41%), 7 day (42%)



Back-up Policy – Tell to Users

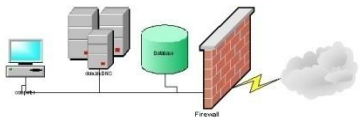
- Back-ups are performed only on data stored on servers (your PC's Z: drive, or UNIX /home directory) every night between midnight and 8 AM.
- *We never do back-ups of your PC's local C: drive.*
- If you need a file recovered, go to [*insert URL*] for more information, or send email to "help" with the name of the server, the file's complete path, and which date you need the restore from.
- Access problems, simple restores are done in 24 hours.



Example

- We have **2TB** data on a server
- Use incremental back-up
- Change 10% / day
  - a. In case of a weekly cycle how large amount of data is to be backed up in 4 weeks?

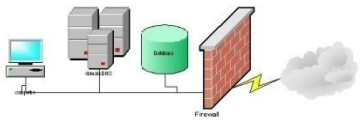
- Full back-up: 2 TB
- Incrementals: 2TB * 10% = 0.2 TB (each day)
- One week: 2TB + 6*0.2 TB = 3.2 TB
- Four weeks: 4 * 3.2 TB = 12.8 TB



Example cont.

b. How large will be the back-up window if the writing capacity of the back-up device is 100 GB/h?

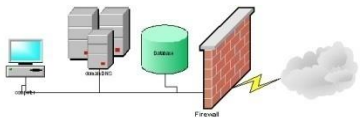
- Sunday (full back-up)
  - 2 TB / 100 GB/h = 20 (!!) hours
- Other days:
  - 0.2 TB / 100 GB/h = 2 hours



Example cont.

c. How many devices are needed if the maximal allowed back-up window is 8 hours?

- Worst: Sunday: 20 hours
- 3 devices needed



Example cont.

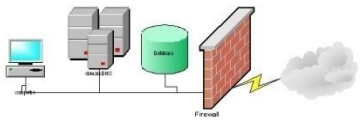
d. How many tapes are needed if we use new tape(s) every day and the capacity of a tape is 500 GB?

Sunday: 2 TB / 500 GB = 4 tapes

Other days: 0.2 TB (= 200 GB) = 1 tape

Total: 4+ 6*1 = 10 tapes / week

40 tapes / 4 weeks



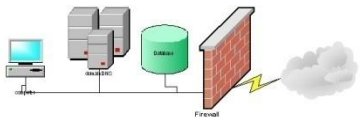
Example cont.

e. Maximum how many tapes needed to restore the content of a given day?

Worst: Saturday

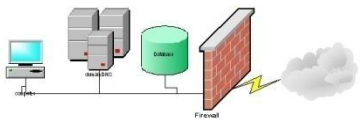
Restore: 1 full + 6 incrementals

4 + 6*1 = 10 tapes needed



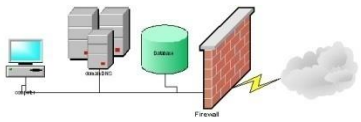
Time and Capacity Planning

- Restores and back-ups are constrained by time
- Restores need to be finished within the time permitted by the SLA
- Service may be disabled until the restore is complete
- The speed of a back-up is limited by the slowest of the following factors:
  - read performance of the disk,
  - write performance of the back-up medium,
  - bandwidth and latency of the network between the disk and the back-up medium



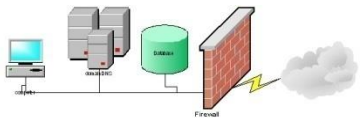
SLA Example - Again

- Customers shall be able to get back any file
  - with a granularity of 1 business day for the past 6 months
  - with a granularity of 1 month for the last 3 years.
- Disk failures shall be restored in 4 hours, with no more than 2 days of lost data
- Archives shall be full back-ups on separate tapes generated quarterly and kept forever
- Critical data will be stored on a system that retains user-accessible snapshots made every hour from 7 AM until 7 PM, with midnight snapshots held for 1 week
- Databases and financial systems shall have higher requirements that shall be determined separately



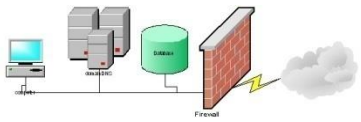
Tape requirement

- Back-up Policy affects the number of tapes needed
- *Example:*
  - tapes containing the incrementals can be reused after 6 months
  - tapes containing the full back-ups (except for the archives) can be reused after 3 years
    - Full back-ups: every 14 days
    - Archives: every 3 months



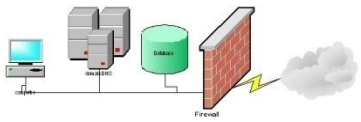
Restoration

- Slow...
- Reading and writing speed of a tape often very different + access time!!
  - Often longer than to restore a partition!
- Speed of restoration is typically determined by the writing speed of the file descriptors!!
- Tricks to perform the back-up faster (e.g. incremental back-up) makes the restore slower
- Hardware limits
  - Tape fast, if the data arrives with exactly the same speed as the writing speed...
- Fastening: typically independent, dedicated network for back-up and restore



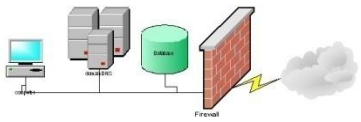
Restoration: safety issues

- Who has the right to claim the restoration (and usage) of a given file? – validation!
- File access rights change after restoration?
- File to be restored at the original place with the original access rights or at a different place with probably different access rights?
- Overwrites existing data?



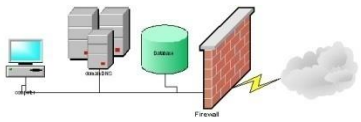
Centralization

- Centralization typically reduces costs of:
  - Equipments (expensive, because require high speed and high precision mechanics and high reliability).
  - Tape replacement (expensive, because needs human's work)
- Disadvantages of distributed back-up
  - Back-up device to every server – for high reliability: 2!
  - Tape replacement takes long time
- Network back-up systems
- Jukeboxs



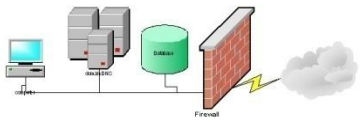
Tape Inventory

- A set of back-up tapes with no index or inventory is only slightly more useful than no back-ups at all…
- Automatic inventory generation
  - No – read every tape backward in time…
  - Partition level
  - File level – fast, but large
  - (Automatic) restoration of inventory
- How to restore if the restore system fails?
  - At least minimal info on the tape itself



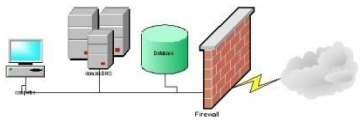
Fire drills

- The only time to learn the quality the back-up system is when doing a restore
  - Restoring a randomly chosen file
  - Restoring a whole disk
    - rarely needed, maybe forget how to do
    - large amount of data – capacity/bandwidth is really enough?



Changes in Technology

- Technology canges are different
  - Disk: almost linear (capacity doubles in 1-1.5 years)
  - Tape: capacity remains the same for years and the big change
    - The tape units (jukeboxes) are expensive – not to replace ***them*** often
- When new tape type comes out – retain 1 (2) pieces from the old platform!



Back-up and Restore - Summary

- Back-up / archives
- Types of back-ups
  - Full, incremental, differential, progressive
- Design of back-up
  - Corporate Guidelines, SLA, Back-up policy, Back-up schedule, Time and Capacity Planning, Tape need calculation
- Restoration