System Administration

Week 09, Segment 1
Backups: Core Concepts

Department of Computer Science
Stevens Institute of Technology

Jan Schaumann
jschauma@stevens.edu
https://stevens.netmeister.org/615/
Backups vs. Restores

Backups are boring.
Backups are tedious.
Nobody likes doing backups.

But…
People really like being able to restore data!
Backups vs. Restores

Backups are just a means to accomplish a specific goal:

To have the ability to restore data.
Basic Terminology, Concepts, and Considerations

- "full backup"
- "incremental backup"
- “differential backup”
- file level vs. block level
- meta data (e.g., file- and filesystem), file data, live data / open files
- journalling vs. snapshots
- Recovery Point Objective (RPO) / Recovery Time Objective (RTO)
- Business Continuity Plan (BCP)
Full Backups

Sun | Mon | Tue | Wed | Thu | Fri | Sat
--- | --- | --- | --- | --- | --- | ---
7 TB | 7 TB | 7 TB | 7 TB | 7 TB | 7 TB | 7 TB

Total: 49 TB
Full Backups

Restore

Back up all data every time:
- slow backup process
- requires lots of storage / bandwidth
- fast restore
Differential Backups

Sun: 7 TB
Mon: 2 TB
Tue: 3 TB
Wed: 4 TB
Thu: 4 TB
Fri: 5 TB
Sat: 6 TB

Total: 31 TB
Back up only data that has changed since the last full backup:
• improved backup performance
• better storage / bandwidth utilization
• slower restore than full backup
• at most two data sets required for recovery

Differential Backups
Incremental Backups

<table>
<thead>
<tr>
<th>Day</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun</td>
<td>7 TB</td>
</tr>
<tr>
<td>Mon</td>
<td>2 TB</td>
</tr>
<tr>
<td>Tue</td>
<td>1 TB</td>
</tr>
<tr>
<td>Wed</td>
<td>1 TB</td>
</tr>
<tr>
<td>Thu</td>
<td>1 TB</td>
</tr>
<tr>
<td>Fri</td>
<td>1 TB</td>
</tr>
<tr>
<td>Sat</td>
<td>1 TB</td>
</tr>
</tbody>
</table>

Total: 13 TB
Incremental Backups

Restore
Incremental Backups

Back up only data that has changed since the last incremental backup:

- highest backup performance
- best storage / bandwidth utilization
- slowest restore
- higher risk of failed full recovery since backup across sets is chained
Data Storage Media and Properties

- magnetic tape
- traditional hard disk
- solid-state drive
- the cloud, why not

- I/O performance (read/write, sequential/random, …)
- reusability and degradation
- longevity
- data integrity assurance (e.g., WORM - write once, read many)
- data compression, encryption
- deduplication
When do we need backups?

- long-term storage / archival
- recover from data loss
When do we need backups?

• long-term storage / archival
  • complete backup
• recover from data loss
When do we need backups?

• long-term storage / archival
  • complete backup
  • separate set from regular backups
• recover from data loss
When do we need backups?

• long-term storage / archival
  • complete backup
  • separate set from regular backups
  • usually stored off-site
• recover from data loss
When do we need backups?

• long-term storage / archival
  • complete backup
  • separate set from regular backups
  • usually stored off-site
  • recovery / retrieval takes time
• recover from data loss
When do we need backups?

• long-term storage / archival
  • complete backup
  • separate set from regular backups
  • usually stored off-site
  • recovery / retrieval takes time
  • limited granularity
• recover from data loss
When do we need backups?

• long-term storage / archival
  • complete backup
  • separate set from regular backups
  • usually stored off-site
  • recovery / retrieval takes time
  • limited granularity
  • storage media considerations
• recover from data loss
When do we need backups?

• long-term storage / archival
  • complete backup
  • separate set from regular backups
  • usually stored off-site
  • recovery / retrieval takes time
  • limited granularity
  • storage media considerations
  • backup encryption and recovery key management
• recover from data loss
When do we need backups?

• long-term storage / archival
• recover from data loss due to e.g.:
  • user failure
  • software bugs
  • equipment failure
  • security breach
  • natural disaster

Type of restore:
• individual file(s)
• individual system recovery
• disaster recovery

Think of your backups as insurance: you invest and pay for it, hoping you will never need it.
System and Disaster Recovery

• loss of e.g. entire file system
• leads to downtime (of individual systems)
• RAID may help
• takes long time to restore
• may require retrieval of archival backups from long-term storage
• often involves some data loss
• 3-2-1 Rule:
  • keep at least 3 copies of your data
  • keep at least 2 copies on different storage media
  • keep at least 1 copy offsite
Beware: disasters scale up much faster than your backup strategy!
Trusting your backups

• Backing up data requires superuser privileges!
• A backup is a copy of the data. If the data is corrupt, your backup may become corrupt.
• To restore data from a trusted backup, you can only use trusted tools.
• Verify the authenticity and integrity of your backups!

Schrodinger’s Backup

“The condition of any backup is unknown until a restore is attempted.”
Practical examples:
• `dump(8) / restore(8)`
• `tar(1)`
• `rsync(1)`

Recommended exercise:
https://stevens.netmeister.org/615/backup-exercise.html
Links

• https://en.wikipedia.org/wiki/Disaster_recovery
• https://en.wikipedia.org/wiki/Write_once_read_many
• https://www.oreilly.com/catalog/unixbr/