

# Backup

Dr Peadar Grant

April 24, 2024

# Contents

<b>1</b>	<b>Backup</b>	<b>S.2</b>
<b>2</b>	<b>Backup granularity</b>	<b>S.9</b>
<b>3</b>	<b>Backup targets</b>	<b>S.16</b>
<b>4</b>	<b>Deduplication</b>	<b>S.26</b>

# 1 Backup

## 1.1 Backup terminology

**Production data** is actively used in the course of day-to-day operations.

**Backup:** additional copy of production data that is created and retained for the sole purpose of recovering lost or corrupted data.

**Archive:** store of data that is no longer actively used that is retained on low-cost secondary storage.



**Figure 1:** Backup operations

## 1.2 Backup purpose

**Disaster recovery** where production systems are damaged / destroyed.

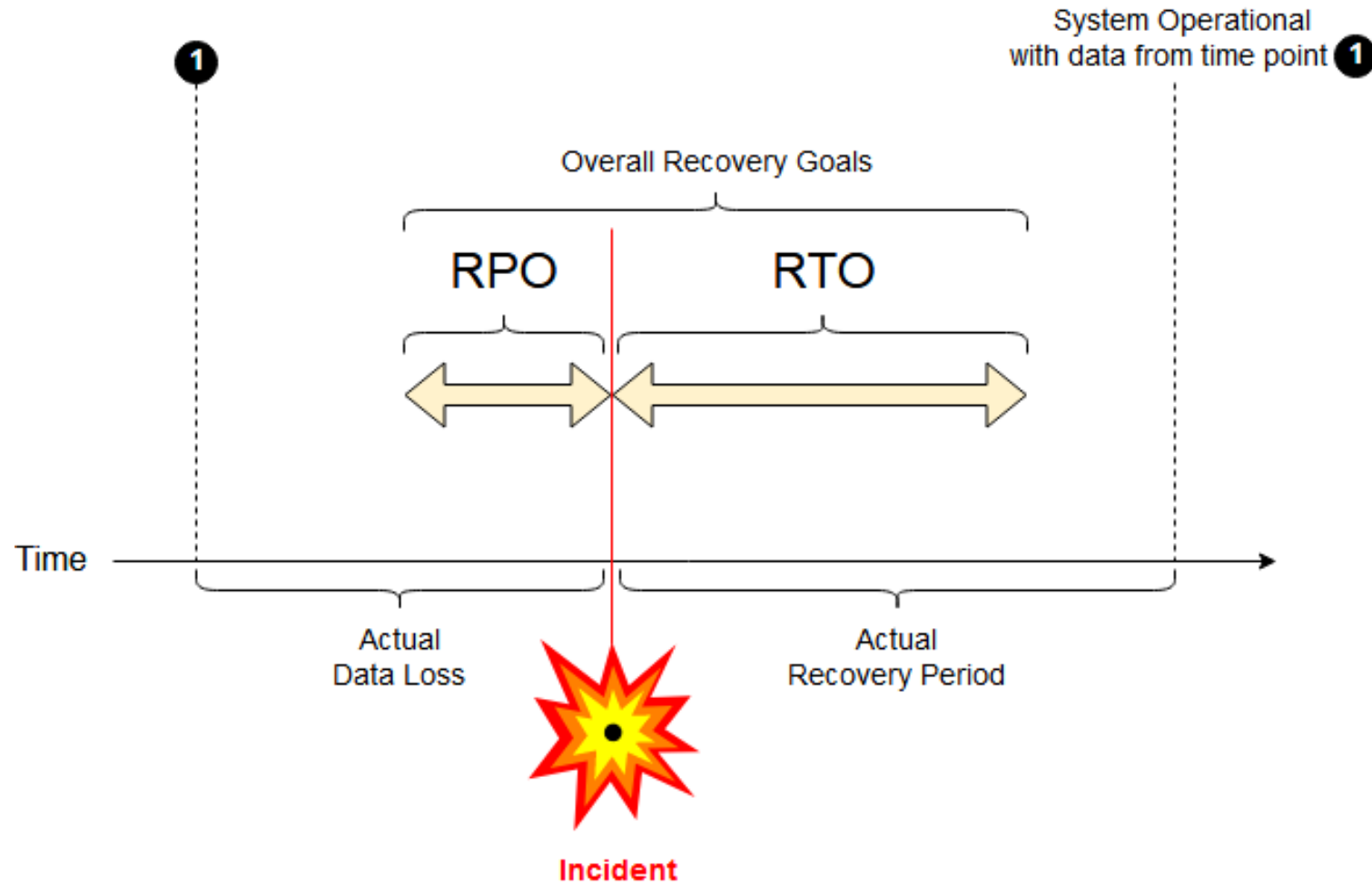
**Operational recovery** where data loss / corruption occurs due to application failure or user error (e.g. important email deleted).

**Archival** for regulatory compliance, historical record, later analysis etc.

## 1.3 Backup parameters

**Recovery Point Objective (RPO)** the most recent point in time to which we should be able to restore. Dictates backup frequency.

**Recovery Time Objective (RTO)** is the time taken to restore the system to the RPO point after an incident occurs.



**Figure 2: RPO-RTO graph**

## 1.4 Backup methods

**Hot / online:** production system is running during backup:

- Does not interrupt normal operation
- Issues with open files, particularly database backend stores.

**Cold / offline:** production system is shut-down for normal operations.



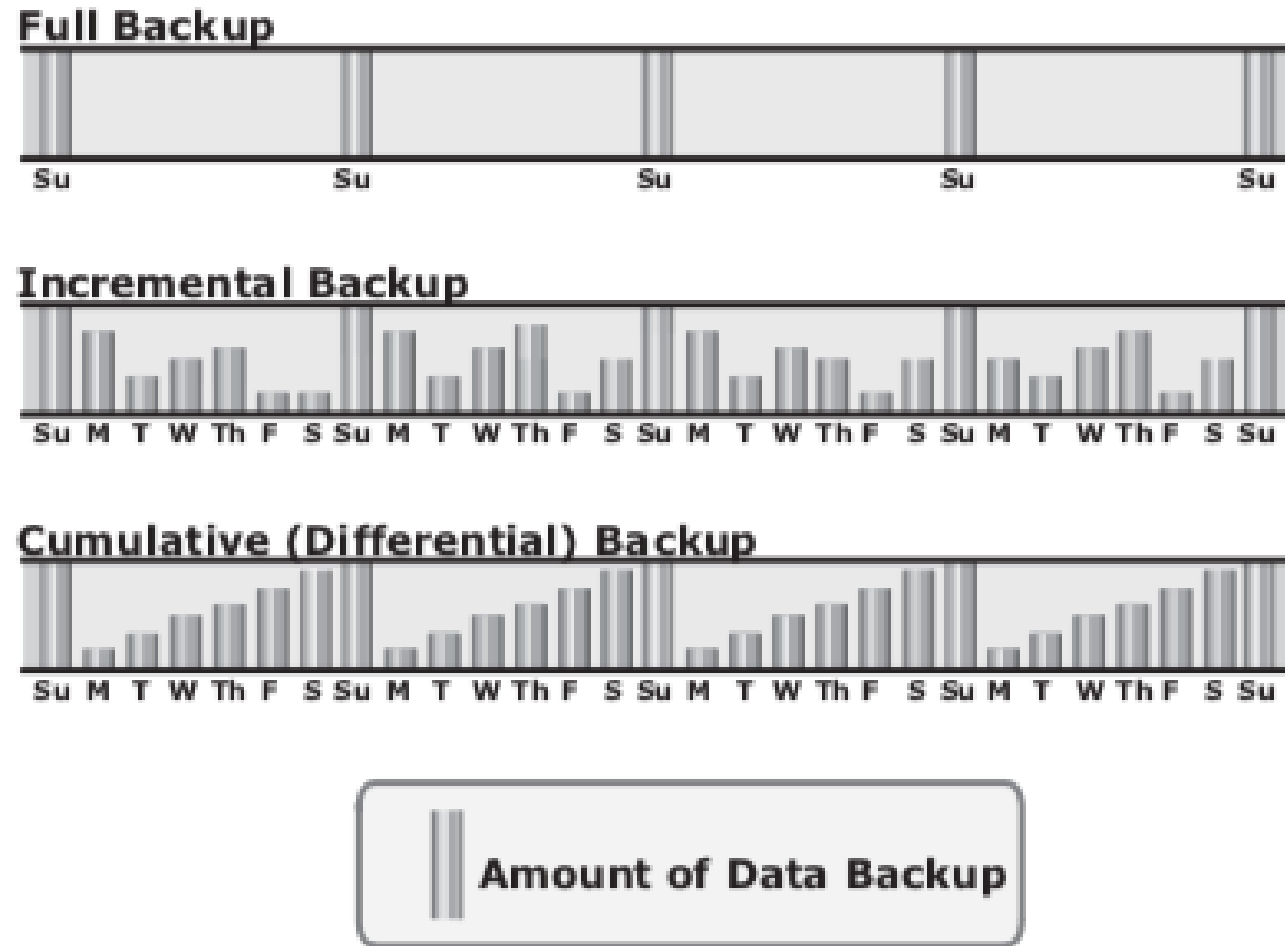
## 1.5 Backup scope

Scope refers to what is actually backed up:

- Files themselves
- Files and their layout (e.g. folders, symlinks)
- Metadata (permissions, last updated)
- Is a *Bare Metal Recovery* required?

## **2 Backup granularity**

There are three common granularities of backup, Figure 3. See Acronis article on different granularities for further info.



**Figure 3:** Backup granularity

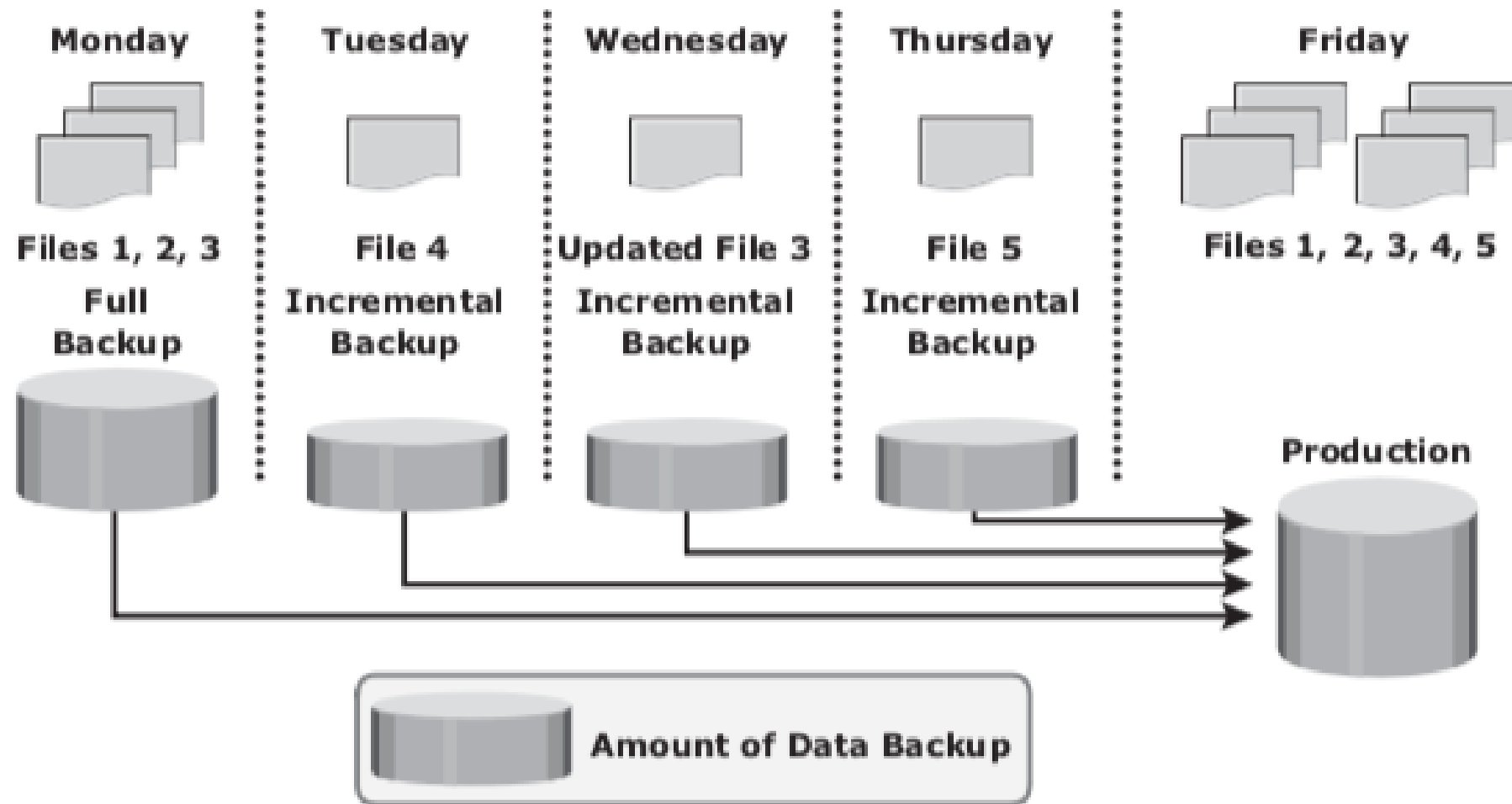
### **2.1 Full backup**

Full backup contains the entire dataset:

- Completeness.
- Frequent full backups may be prohibitive in terms of time and/or storage space.

### **2.2 Incremental backup**

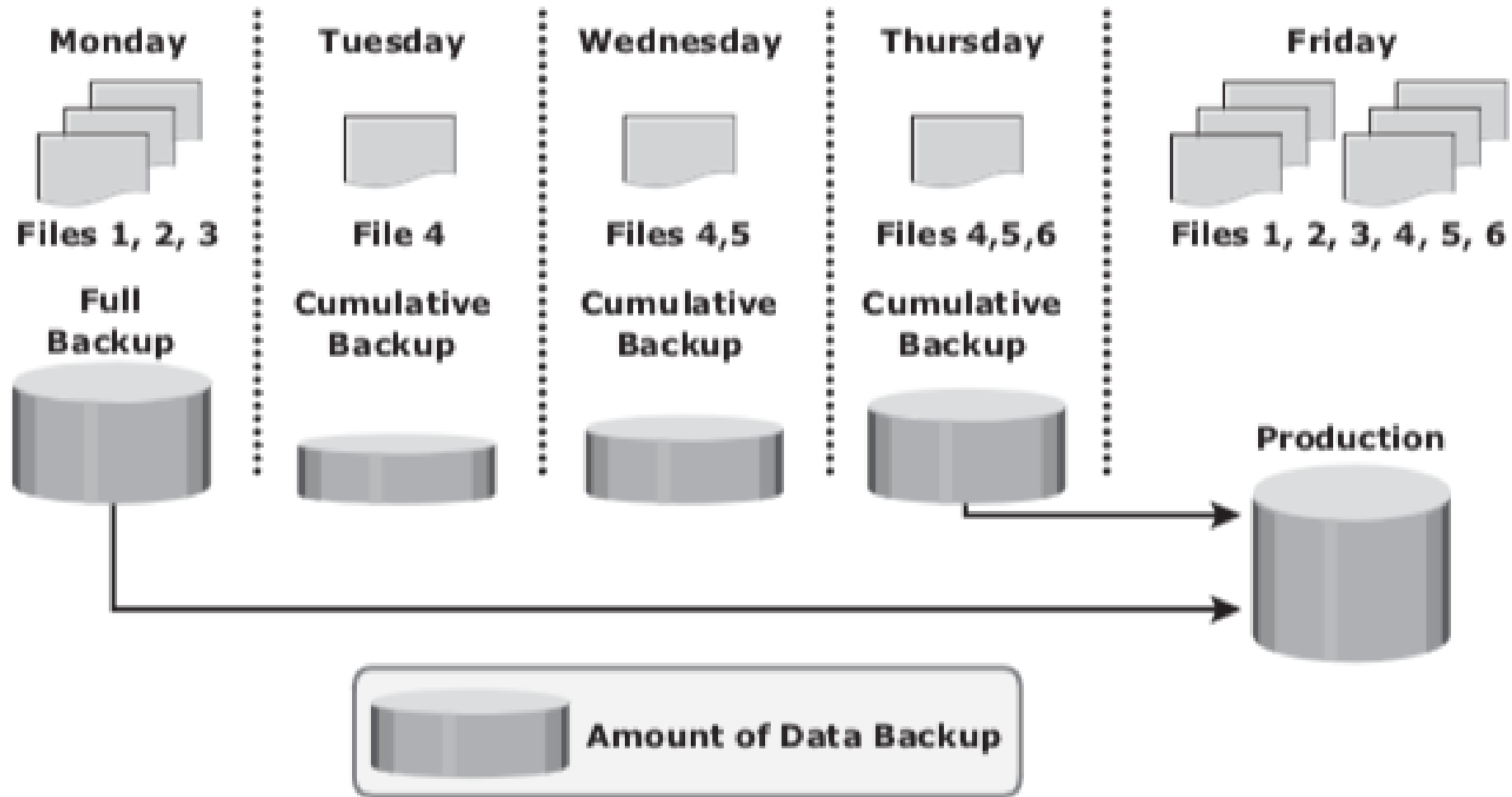
Incremental backups store only the files changed since the most recent full or incremental backup was taken. Restoration steps are shown in .



**Figure 4:** Restore from incremental

### 2.3 Cumulative / Differential backup

Differential backup includes files changed since the last **full** backup. Restoration steps are shown in

**Figure 5:** Restore from cumulative



## **3 Backup targets**

Backup data is written to targets. Like any other storage media, targets may be directly attached to a host or may be LAN or SAN attached.

### **3.1 Disk-based backup**

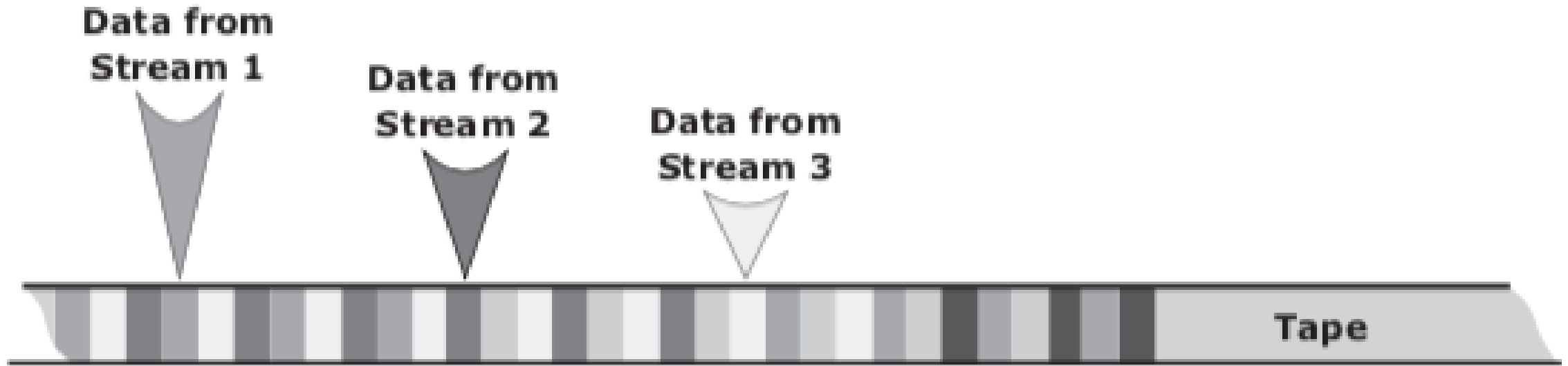
Disk-based backup has low cost, fast backup and recovery. Should be using monitored RAID to ensure backup availability.

### **3.2 Tape-based backup**

Backup traditionally has used tape media.

Tape drives are best operated such that they do not have to start/stop repeatedly during operation

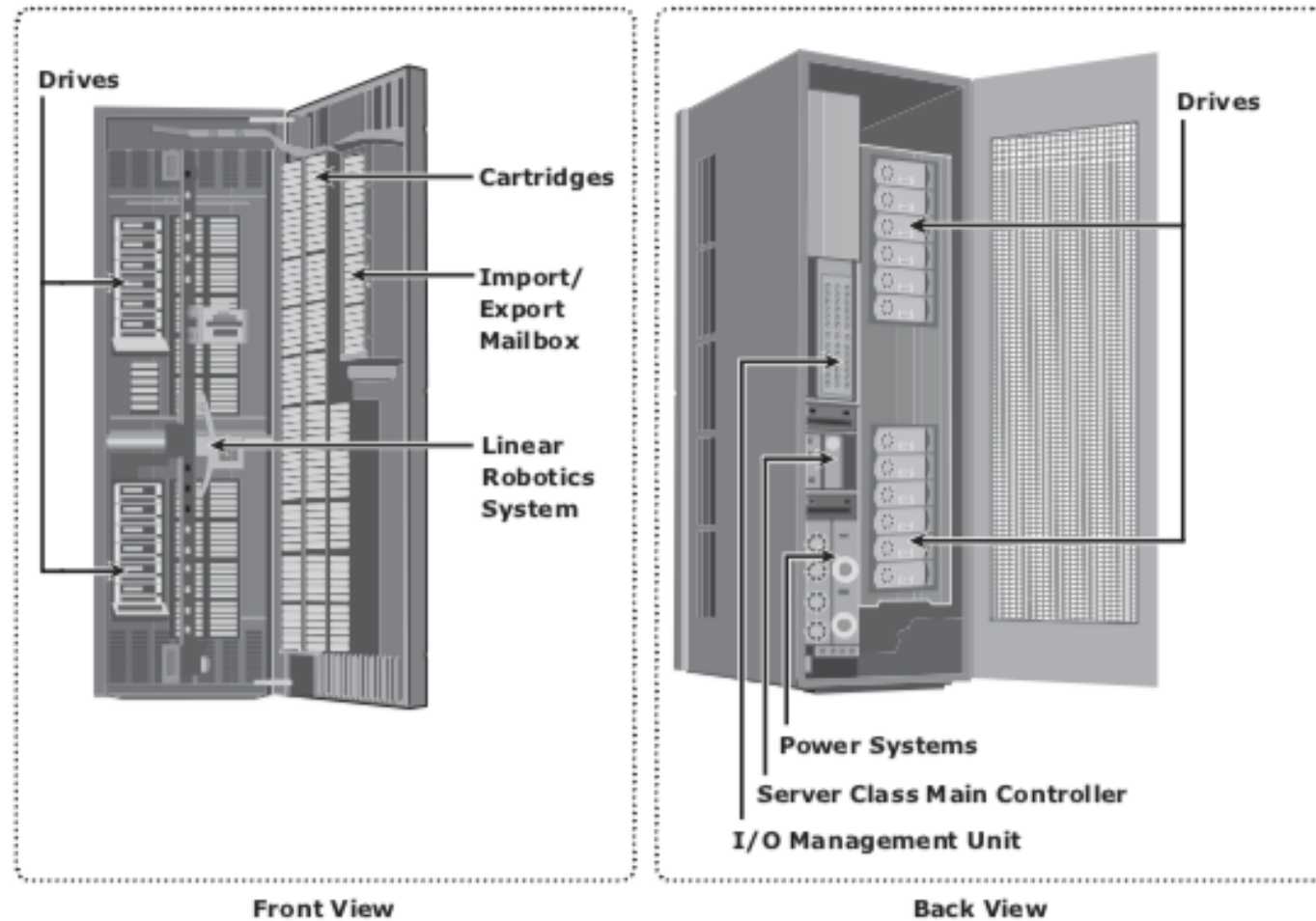
Interleaving can be used so that multiple streams of data can be written to a tape at speed,  
Figure 6.



**Figure 6:** Interleave tape

### **3.3 Tape libraries**

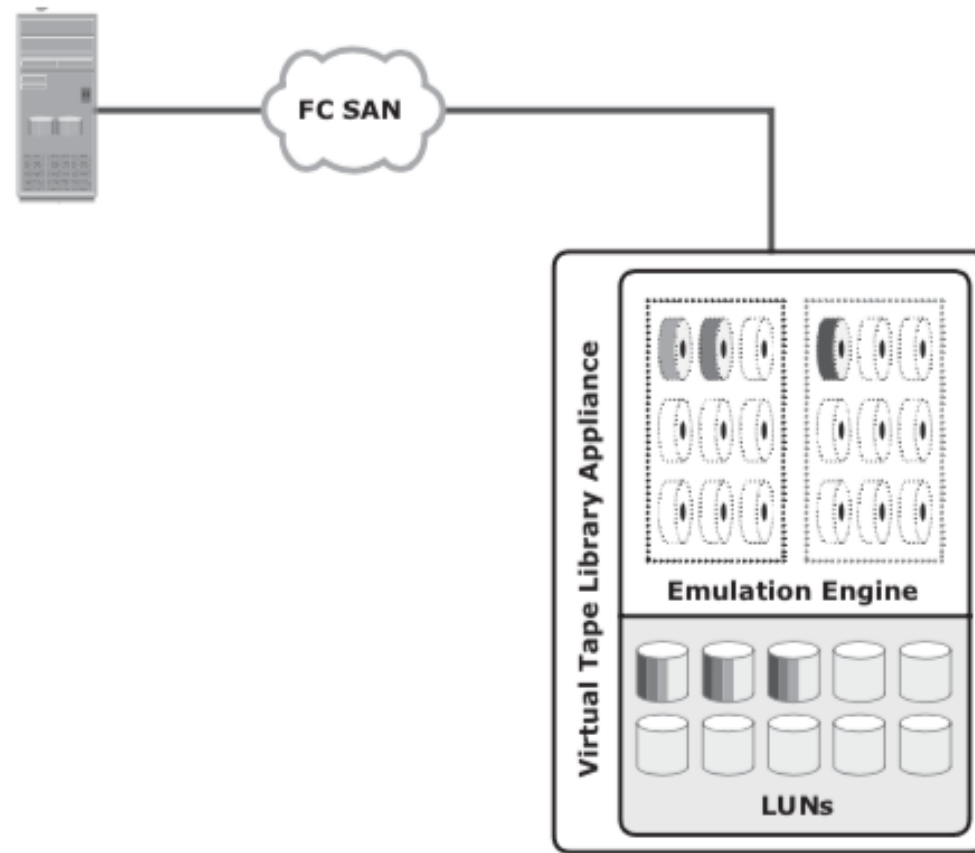
Tape backup can be provisioned on a tape library in medium-large installations, Figure 7.



**Figure 7:** Physical tape library

## **3.4 Virtual tape library**

Virtual tape library emulates a tape library but is actually provisioned usually from SAN LUNs, Figure 8.



**Figure 8:** Virtual tape library

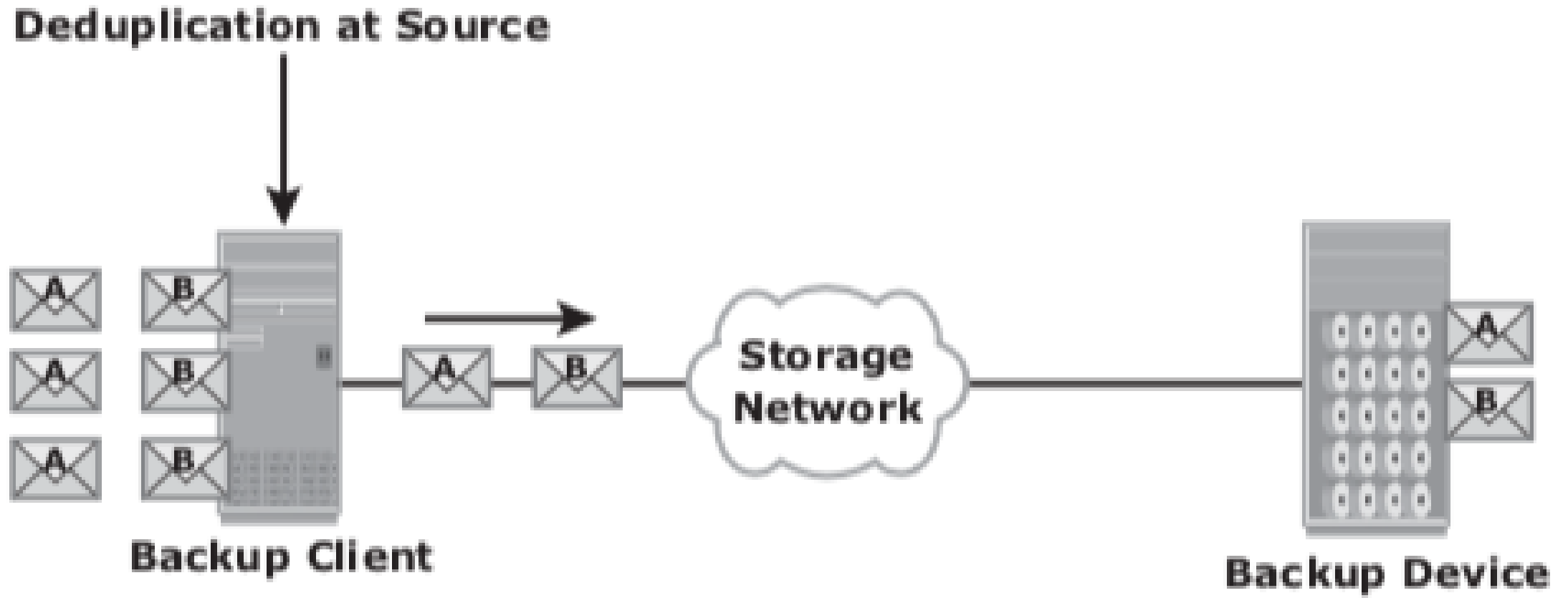


## 3.5 Comparison

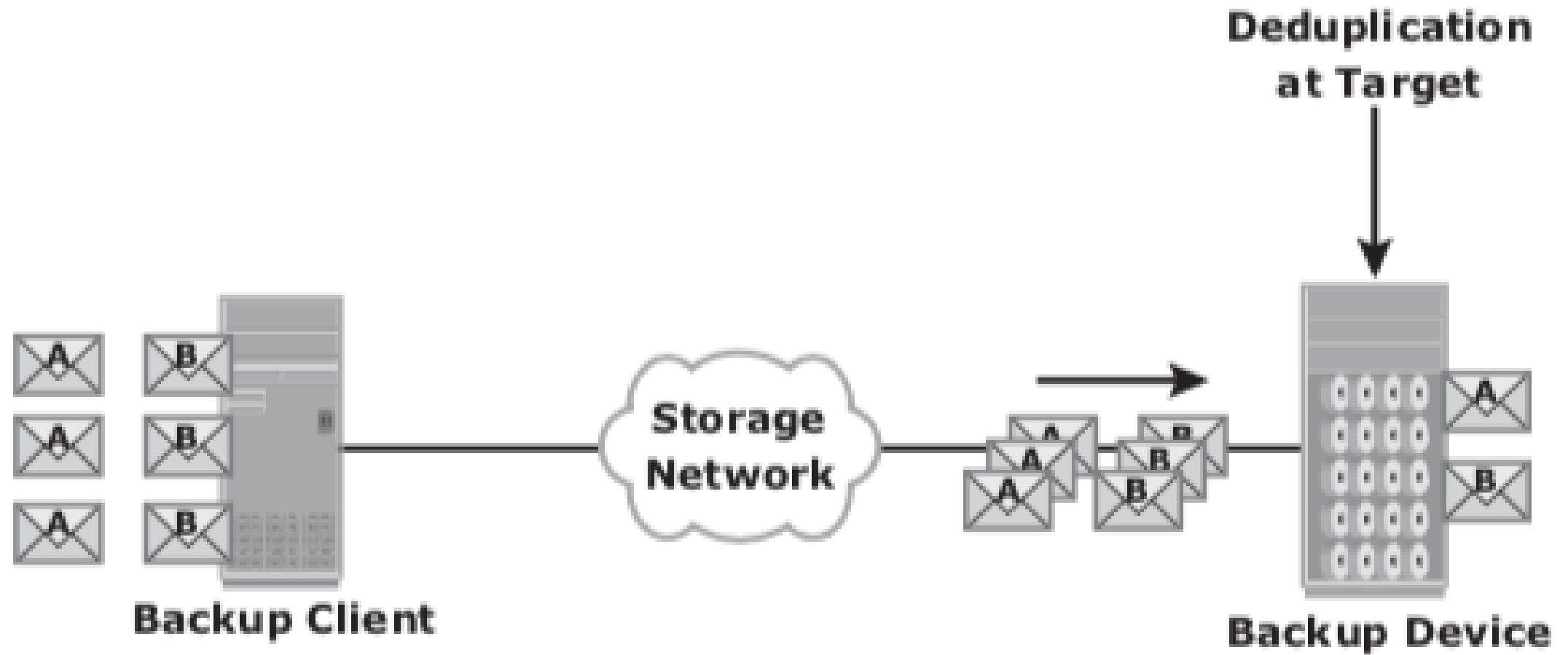
FEATURES	TAPE	DISK	VIRTUAL TAPE
Offsite Replication Capabilities	No	Yes	Yes
Reliability	No inherent protection methods	Yes	Yes
Performance	Subject to mechanical operations, loading time	Faster single stream	Faster single stream
Use	Backup only	Multiple (backup, production)	Backup only

**Figure 9:** Backup targets comparison

## 4 **Deduplication**



**Figure 10:** Deduplication at source



**Figure 11:** Deduplication at target