Maîtriser le stockage des données en entreprise — Jour 1

Jour 1 : rôle des données, types de stockage (fichier, BDD, bloc, objet), stockage On Premise (baies, RAID, SAN vs NAS) et TP initiateur iSCSI.

Stockage et Données | 65 min de lecture | Niveau Intermédiaire

Document généré le 25/06/2026 à 21h40 · nouvy.fr/wiki/maitriser-stockage-donnees-entreprise-jour1

Sommaire

41 section(s) · 65 min de lecture

Objectifs du jour 1

- Comprendre le **rôle des données** dans l'entreprise et les **différents types de stockage**.
- Distinguer **stockage de fichiers à plat** et **bases de données** et savoir les positionner.
- Découvrir l'**infrastructure On Premise** : baies de disques, **RAID**, différences **SAN / NAS**.
- Réaliser un **TP** : se connecter à une cible iSCSI (initiateur Linux) et monter un LUN.

Identifier les différents types de stockage en entreprise

En entreprise, les données sont stockées selon plusieurs modalités : **fichiers à plat**, **bases de données**, **stockage sur site (On Premise)** avec baies de disques, **SAN/NAS**, et **stockage en ligne (cloud)** en IaaS, PaaS ou SaaS. Ce module permet d'identifier et de positionner chaque type selon les usages et les contraintes (coût, performance, sécurité, conformité).

Objectifs d'apprentissage (module complet)

À l'issue des deux jours, vous serez capable de :

- Identifier le **rôle des données** et les **types de stockage** adaptés à chaque besoin.
- Distinguer **stockage de fichiers à plat** et **bases de données** et choisir le bon usage.
- Comprendre l'**infrastructure On Premise** : baies, **RAID**, **SAN** et **NAS** (zoning, LUN masking).
- Connaître le **stockage online** : acteurs, coûts, modes **IaaS**, **PaaS**, **SaaS**.
- Maîtriser les **protocoles d'accès** et l'**intégration avec un annuaire**.
- Appliquer les bonnes pratiques de **sécurité des données**.

Prérequis du module

- Notions de base en **réseau** (IP, LAN, protocoles).
- Connaissances de base en **systèmes d'exploitation** (fichiers, volumes, disques).
- Idéal : première expérience avec un **annuaire** (LDAP/Active Directory) et une **base de données** (requêtes, sauvegardes).

Le rôle des données dans l'entreprise

Les données structurent les processus métier, la relation client, la production et la décision. Leur **disponibilité**, **intégrité** et **confidentialité** conditionnent directement la continuité et la conformité de l'entreprise.

- **Stratégie** : données comme levier de différenciation et d'innovation.
- **Opérationnel** : données pour piloter la production, les stocks, les commandes.
- **Conformité** : RGPD, sectoriel (santé, finance), preuves et traçabilité.

Les données : l'ADN de l'entreprise

Les données constituent l'**ADN** de l'entreprise : elles décrivent les clients, les produits, les contrats, les flux. Une perte ou une fuite impacte la confiance, la réputation et peut engager la responsabilité juridique.

- **Valeur** : données structurées (BDD) vs non structurées (fichiers, logs, médias).
- **Cycle de vie** : création, utilisation, archivage, suppression (droit à l'oubli).
- **Gouvernance** : propriétaire des données, classification, politique de rétention.

Les différents types de stockage

Type	Caractéristique principale	Exemples
Fichiers à plat	Fichiers sur système de fichiers (NFS, SMB, objet)	Documents, images, sauvegardes, data lakes
Bases de données	Données structurées, requêtes, transactions	MySQL, PostgreSQL, SQL Server, Oracle
Block storage	Volumes disque (blocs) pour OS et applications	SAN, iSCSI, volumes cloud (EBS, disques Azure)
File storage	Partage de fichiers en réseau	NAS, NFS, SMB/CIFS
Object storage	Objets (clé + métadonnées + contenu)	S3, Azure Blob, Swift

Le choix dépend des **usages** (accès séquentiel vs aléatoire, partage, sauvegarde, analytics) et des **contraintes** (latence, débit, coût, conformité).

Le stockage de fichiers à plat : pour quels usages ?

Le **stockage de fichiers à plat** correspond à des fichiers stockés dans une arborescence (répertoires/fichiers), sans modèle relationnel.

Usages typiques :

- **Documents métier** : contrats, factures, PDF (partage via NAS ou partage cloud).
- **Médias** : images, vidéos, sons (CDN, archivage).
- **Sauvegardes** : dumps, archives (réplication, tiering froid).
- **Data lakes / analytics** : fichiers CSV, Parquet, JSON pour traitements batch (Spark, etc.).
- **Logs et traces** : fichiers texte ou binaires, rotation, rétention.

Avantages : simplicité, interopérabilité, coût souvent faible pour de gros volumes. **Inconvénients :** pas de requêtes complexes, pas de transactions, gestion des métadonnées et des doublons à la charge des applications.

Les bases de données : pour quels usages ?

Les **bases de données** stockent des données **structurées** avec un schéma (tables, relations) et permettent des **requêtes**, des **transactions** (ACID) et une **gestion fine des droits**.

Usages typiques :

- **Applications métier** : ERP, CRM, e-commerce (données transactionnelles).
- **Référentiels** : annuaires, catalogues, paramétrage.
- **Reporting et BI** : agrégations, tableaux de bord (souvent sur un entrepôt dédié).
- **Sessions et cache** : Redis, Memcached (clé-valeur, très faible latence).

Choix selon les besoins : SQL (relationnel) pour cohérence et requêtes complexes ; NoSQL (document, clé-valeur, colonnes) pour volume, flexibilité schéma ou distribution. Le stockage sous-jacent peut être **local** (disques serveur), **SAN** (volumes bloc) ou **cloud** (RDS, Cosmos DB, etc.).

Le stockage On Premise

Le stockage **On Premise** désigne des équipements et logiciels hébergés **dans les locaux** de l'entreprise (ou en datacenter dédié), sous sa responsabilité directe.

Composants principaux :

- **Baies de disques** : enceintes contenant disques durs et/ou SSD, contrôleurs, cache.
- **RAID** : redondance et performance au niveau des disques.
- **SAN (Storage Area Network)** : réseau dédié au stockage bloc (Fibre Channel, iSCSI).
- **NAS (Network Attached Storage)** : partage de fichiers (NFS, SMB) sur le réseau IP.

Les baies de disques : fonctionnement et positionnement des acteurs

Une **baie de disques** (storage array) regroupe des disques (HDD/SSD), des contrôleurs, du cache (RAM/NVMe) et des ports (FC, iSCSI, NFS/SMB). Elle fournit des **volumes logiques (LUN)** aux serveurs.

Fonctionnement :

- **Contrôleurs** : gestion du RAID, cache, réplication, snapshots.
- **Tiering** : répartition des données entre disques rapides (SSD) et lents (HDD) selon l'activité.
- **Réplication** : synchrone ou asynchrone vers une autre baie (DR, BC).

Acteurs (exemples) : Dell EMC (PowerStore, Unity), HPE (Nimble, Primera), NetApp, Pure Storage, IBM (FlashSystem), Huawei. Positionnement selon gamme (entrée de gamme / milieu / haut de gamme), protocoles et services (support, garanties).

Les RAID : les différents types de RAID et leurs avantages/inconvénients

Le **RAID** (Redundant Array of Independent Disks) combine plusieurs disques pour améliorer **performance** et/ou **tolérance aux pannes**.

Niveau	Description	Avantages	Inconvénients
RAID 0	Stripping (données réparties)	Très bon débit, capacité = somme des disques	Aucune redondance : 1 disque en panne = perte totale
RAID 1	Miroir (2 disques identiques)	Redondance simple, bonne lecture	Coût (2x capacité), écriture limitée
RAID 5	Parité répartie (N disques, 1 de parité)	Bon compromis capacité/redondance, lecture correcte	Écriture plus lente, risque en reconstruction (2e panne)
RAID 6	Double parité	Tolère 2 pannes de disques	Coût capacité, écriture plus lente que RAID 5
RAID 10	Miroir + stripping	Bonnes perfs et redondance	Coût (2x capacité minimum)

En entreprise, **RAID 5/6** pour capacité et coût, **RAID 10** quand la performance et la résilience sont prioritaires.

Les différences SAN / NAS dans une baie de stockage

Critère	SAN (Storage Area Network)	NAS (Network Attached Storage)
Unité exposée	**Bloc** (volume disque, LUN)	**Fichier** (partage de répertoires)
Protocoles	FC, iSCSI, FCoE	NFS, SMB/CIFS
Réseau	Réseau dédié (Fibre Channel) ou VLAN iSCSI	Réseau IP (LAN)
Usage typique	OS, bases de données, applications exigeantes	Partage de fichiers, sauvegardes fichier
Granularité	Volume entier géré par le serveur	Fichiers et dossiers partagés

Une **même baie** peut proposer les deux : **bloc** (SAN) pour les serveurs critiques et **fichier** (NAS) pour les partages. On parle alors de baie **unifiée**. Le **Jour 2** détaillera le réseau SAN, le zoning et le LUN masking.

TP 1 : Initiateur iSCSI et découverte des LUN (Linux)

Ce qu'on simule et pourquoi

En entreprise, un **SAN** repose sur une **baie de stockage** qui expose des **LUN** (volumes bloc) et des **serveurs** qui s'y connectent en **initiateurs** pour utiliser ces disques (OS, base de données, etc.). On n'a pas de baie physique en lab, donc on **simule** ce schéma avec deux machines :

- **VM 1 (Debian) — la « baie »** : on installe un logiciel **cible iSCSI** (target) qui expose un ou plusieurs « disques » (LUN) sur le réseau. Cela joue le rôle de la **baie de stockage**.
- **VM 2 (Debian ou autre Linux) — le « serveur »** : on installe un **initiateur iSCSI** qui se connecte à la cible et découvre les LUN. Pour l'OS, ça ressemble à un disque local. Cela joue le rôle du **serveur d'application** ou **serveur de BDD** qui consomme le stockage SAN.

Pourquoi faire ça ? Pour comprendre concrètement le **modèle client/serveur du stockage bloc** (qui parle à qui, découverte, IQN, LUN) avant d'aborder le zoning et le LUN masking au Jour 2. En production, la cible serait une vraie baie (NetApp, Dell, etc.) ou un NAS qui fait iSCSI ; en TP, une VM Debian avec **targetcli** suffit.

Prérequis : Deux VM Linux (Debian recommandé pour la cible) pouvant se joindre en réseau (IP). On commence par **créer la cible sur la première VM**, puis on **connecte l'initiateur depuis la seconde**.

Partie A : Créer la cible iSCSI sur une VM Debian (la « baie »)

À faire sur **VM 1** (Debian). Cette VM simule la baie de stockage : elle va exposer un volume bloc (LUN) en iSCSI.

1. **Installer le serveur cible iSCSI (LIO)**

```
sudo apt update
sudo apt install -y targetcli-fb
```

Copier

2. **Créer un « disque » à exposer (backstore)**
   Sans disque dédié, on utilise un **fichier** comme support (fileio). Créer par exemple un fichier de 1 Go :

```
sudo mkdir -p /var/lib/iscsi-disks
sudo dd if=/dev/zero of=/var/lib/iscsi-disks/lun0.img bs=1M count=1024
```

Copier

   (Pour utiliser un vrai disque ou partition, par ex. `/dev/sdb1`, on utiliserait un backstore **block** au lieu de **fileio**.)

3. **Lancer targetcli et configurer la cible**

```
sudo targetcli
```

Copier

   Dans le shell `targetcli`, exécuter les commandes suivantes (une par une) :

   Créer le backstore (fileio) :

```
/backstores/fileio create name=disk0 file_or_dev=/var/lib/iscsi-disks/lun0.img
size=1G
```

Copier

   Créer le target iSCSI (nom IQN au choix, ici exemple) :

```
/iscsi create iqn.2025-01.local.iscsi:tp-jour1
```

Copier

   Créer le LUN 0 (lier le backstore au target) :

```
/iscsi/iqn.2025-01.local.iscsi:tp-jour1/tpg1/luns create
/backstores/fileio/disk0
```

Copier

   Autoriser l'initiateur (on ajoutera l'IQN de la VM 2 à l'étape suivante). Pour

l'instant, en mode découverte sans ACL stricte, on peut créer un portal qui écoute sur toutes les interfaces :

```
/iscsi/iqn.2025-01.local.iscsi:tp-jour1/tpg1/portals create 0.0.0.0
```

Copier

(Si vous voulez restreindre par ACL : créer une ACL avec l'IQN de l'initiateur — voir plus bas.)

Sauvegarder et quitter :

```
saveconfig
exit
```

Copier

4. **Vérifier l'IP de la VM cible** (pour la suite sur l'initiateur)

```
ip -4 addr show
```

Copier

Noter l'adresse (ex. `192.168.100.10`). Les deux VM doivent pouvoir se ping.

5. **Optionnel — Restreindre l'accès par ACL (LUN masking simplifié)**
Sur la VM **initiateur** (VM 2), récupérer l'IQN :

```
cat /etc/iscsi/initiatorname.iscsi
```

Copier

Puis sur la VM **cible**, dans `targetcli` :

```
/iscsi/iqn.2025-01.local.iscsi:tp-jour1/tpg1/acls create
iqn.1993-08.org.debian:01:XXXXXXXX
```

Copier

(Remplacer par l'IQN affiché.) Seuls les initiateurs dont l'IQN est dans les ACL pourront se connecter (on approfondira au Jour 2).

Partie B : Se connecter à la cible depuis l'initiateur (le « serveur »)

À faire sur **VM 2** (Debian ou autre Linux). Cette VM simule le serveur qui utilise le stockage SAN.

1. **Installer l'initiateur iSCSI**

```
# Debian/Ubuntu
sudo apt update && sudo apt install -y open-iscsi
# RHEL/CentOS
sudo dnf install -y iscsi-initiator-utils
```

Copier

2. **Découvrir la cible** (remplacer `192.168.100.10` par l'IP de la VM cible)

```
sudo iscsiadm -m discovery -t st -p 192.168.100.10
```

Copier

Vous devez voir une ligne du type : `192.168.100.10:3260,1 iqn.2025-01.local.iscsi:tp-jour1`

3. **Se connecter (login)**
   Adapter l'IQN et l'IP si besoin (ceux affichés par la découverte) :

```
sudo iscsiadm -m node -T iqn.2025-01.local.iscsi:tp-jour1 -p 192.168.100.10 -l
```

Copier

4. **Vérifier le nouveau disque**

```
lsblk
# ou
sudo fdisk -l
```

Copier

Un nouveau disque (ex. `/dev/sdb`) doit apparaître : c'est le **LUN** exposé par la « baie » (VM 1).

5. **Créer une partition, un système de fichiers et monter** (adapter le device si différent)

```
sudo fdisk /dev/sdb   # n, p, 1, entrée, w
sudo mkfs.ext4 /dev/sdb1
sudo mkdir -p /mnt/iscsi_lun
sudo mount /dev/sdb1 /mnt/iscsi_lun
```

Copier

Tester : `echo "TP stockage Jour 1" | sudo tee /mnt/iscsi_lun/test.txt` puis `cat /mnt/iscsi_lun/test.txt`.

6. **Optionnel : connexion automatique au démarrage**

```
sudo iscsiadm -m node -T iqn.2025-01.local.iscsi:tp-jour1 -p 192.168.100.10 --
op update -n node.startup -v automatic
```

Copier

Critères de réussite : Vous avez **créé la cible sur une VM Debian** (Partie A) et **connecté l'initiateur** depuis une seconde VM (Partie B). Après login iSCSI, un disque bloc apparaît ; vous pouvez le partitionner, le formater et le monter. Vous comprenez ce qu'on a **simulé** (baie = cible, serveur = initiateur) et **pourquoi** (modèle SAN en lab sans matériel dédié).

Maîtriser le stockage des données en entreprise — Jour 2

Objectifs du jour 2

- Maîtriser en détail le **réseau SAN**, le **zoning** (Fibre Channel) et le **LUN masking**.
- Analyser des **études de cas** de stockage d'entreprise et proposer une stratégie complète de stockage des données.
- Comprendre le **stockage online** : acteurs, coûts, modes **IaaS**, **PaaS**, **SaaS**.
- Connaître les **protocoles d'accès**, l'**intégration avec un annuaire** et la **gestion de la sécurité des données**.

Ce **Jour 2** est la **suite directe du Jour 1** : après les fondamentaux (types de stockage, RAID, baies, SAN vs NAS), on va se concentrer sur la **conception d'architectures complètes** et la **définition d'une stratégie globale de stockage des données d'entreprise**, à travers plusieurs **études de cas** présentées à l'oral sous forme de **slides**.

Le réseau SAN, le Zoning, le LUN Masking

Qu'est-ce qu'un réseau SAN ?

Un **SAN (Storage Area Network)** est un réseau **dédié** dont le seul but est de transporter des commandes **SCSI** (lecture/écriture en mode bloc) entre des **initiateurs** (serveurs, HBA) et des **cibles** (baies de stockage ou logiciels émulant une cible iSCSI). Contrairement au LAN qui transporte des paquets IP (fichiers, web), le SAN transporte des **blocs disque** : le serveur « voit » un disque distant comme un disque local.

Deux familles de protocoles :

Protocole	Réseau	Débit typique	Usage
Fibre Channel (FC)	Réseau dédié (fibre optique, switches FC)	8 / 16 / 32 Gbit/s	Environnements critiques, faible latence
iSCSI	Réseau IP (Ethernet)	1 / 10 / 25 Gbit/s	Coût moindre, réutilisation du réseau

Avec **FCoE** (Fibre Channel over Ethernet), les trames FC sont encapsulées dans Ethernet ; le SAN reste logiquement FC (zoning, WWN) mais physiquement sur le réseau IP.

Composants d'un SAN

- **HBA (Host Bus Adapter)** ou **CNA** : carte dans le serveur qui se connecte au SAN (port FC ou NIC iSCSI). Chaque port possède un identifiant unique : **WWN** (World Wide Name) en FC, ou **IQN** (iSCSI Qualified Name) en iSCSI.
- **Switch SAN** : en FC, c'est le **fabric switch** (Brocade, Cisco MDS) qui interconnecte tous les ports ; le **zoning** se configure sur ce switch. En iSCSI, on utilise des switches Ethernet (souvent des **VLAN dédiés**).
- **Baie de stockage** : elle expose des **LUN** (Logical Unit Number). Chaque LUN est un volume bloc (souvent issu d'un pool RAID). Les ports de la baie ont aussi des WWN (FC) ou IQN (iSCSI).

Flux typique : Serveur (initiateur) → Switch → Baie (cible). Le serveur envoie des commandes SCSI (I/O) vers une cible (adresse + LUN). Sans zoning ni LUN masking, tous les initiateurs pourraient en théorie voir tous les LUN de la baie : risque d'accès accidentel, de corruption ou de faille de sécurité.

Zoning (Fibre Channel)

Le **zoning** se configure sur le **switch FC** (fabric). Il définit des **zones** : un ensemble de ports (ou de WWN) qui ont le droit de communiquer entre eux. Un port peut appartenir à plusieurs zones ; la règle est : **seuls les membres d'une même zone peuvent s'échanger des trames**.

Objectifs :

- **Isolation** : le serveur A ne voit que les ports de la baie qui lui sont affectés ; le serveur B ne voit pas les LUN de A.
- **Éviter les réponses au mauvais endroit** : sans zoning, un initiateur pourrait recevoir des réponses destinées à un autre, provoquant erreurs ou corruptions.
- **Sécurité** : un serveur compromis ou mal configuré ne peut pas scanner toute la fabric.

Types de zoning :

Type	Définition	Avantage	Inconvénient
Zoning par port	La zone contient des adresses de port (Domain, Area) sur le switch	Simple, performant	Si on change de port physique (câble, carte), il faut mettre à jour la zone
Zoning par WWN	La zone contient les **WWN** (World Wide Name) des ports initiateur et cible	Si on déplace un câble, la zone reste valide (le WWN suit la carte)	Dépendance à la stabilité des WWN des équipements

En pratique, le **zoning par WWN** est souvent préféré pour la flexibilité. Une configuration courante est le **zoning « single initiator »** : une zone par serveur (ou par cluster), contenant les ports de ce serveur et les ports des baies qu'il a le droit d'utiliser.

Exemple de zones (schéma logique) :

```
Zone_Server_BDD_1 :
  - WWN du HBA du serveur BDD-1 (initiateur)
  - WWN du port 0 de la baie (cible)

Zone_Server_BDD_2 :
  - WWN du HBA du serveur BDD-2 (initiateur)
  - WWN du port 0 de la baie (cible)

Zone_Backup :
  - WWN du HBA du serveur Backup
  - WWN des deux ports de la baie (redondance)
```

Copier

Une fois les zones activées dans la **zone set** (configuration active du switch), chaque initiateur ne découvre que les cibles présentes dans ses zones.

LUN Masking

Le **zoning** limite « qui peut parler à qui » au niveau du **réseau** (switch). Le **LUN masking** limite **quels LUN** sont visibles par **quel initiateur**. Il se configure soit côté **baie** (le plus courant), soit côté **initiateur** (logiciel qui filtre les LUN découverts).

Pourquoi les deux ?

- **Zoning** : évite que le trafic SCSI aille vers le mauvais équipement et isole les flux (sécurité et stabilité du fabric).
- **LUN masking** : même si un serveur « voit » un port de la baie (grâce au zoning), la baie peut ne lui exposer que **certains LUN**. Ainsi, le serveur de prod ne voit pas les LUN de la preprod ; le serveur Windows ne voit pas les LUN destinés au cluster Linux, etc.

Où se configure le LUN masking ?

- **Côté baie** : on crée des **host groups** (ou « host » selon le constructeur) : on associe un ou plusieurs initiateurs (WWN ou IQN) à un groupe. Ensuite, on assigne à ce groupe une liste de **LUN**. Seuls les LUN assignés à ce groupe sont présentés aux initiateurs de ce groupe. C'est la méthode la plus fiable et centralisée.
- **Côté initiateur** : certains logiciels (multipath, outils de découverte) permettent de masquer des LUN par numéro. Moins centralisé, utile en complément ou en lab.

Exemple (logique) :

- Baie expose LUN 0 (OS), LUN 1 (data BDD), LUN 2 (logs), LUN 3 (autre serveur).
- **Host group Prod-DB** : initiateur = WWN du serveur BDD-1 → LUN 0, 1, 2.
- **Host group Autre-Serveur** : initiateur = WWN de l'autre serveur → LUN 3 uniquement.

Résultat : le serveur BDD-1 ne voit que LUN 0, 1, 2 ; l'autre serveur ne voit que LUN 3. Pas de risque de formater le mauvais disque.

Synthèse : ordre logique et bonnes pratiques

1. **Créer les LUN** sur la baie (depuis des pools RAID).
2. **Définir les zones** sur le switch FC (ou VLAN/ACL pour iSCSI) pour que chaque serveur ne voie que les ports cibles qui le concernent.
3. **Configurer le LUN masking** sur la baie : host groups + assignation des LUN par groupe.
4. **Découverte côté serveur** : rescan SCSI/iSCSI pour que l'OS voie les nouveaux disques.
5. **Multipathing** : si plusieurs chemins (plusieurs ports) vers la baie, configurer le multipath (DM-MPIO, PowerPath, etc.) pour basculement et répartition de charge.

Pour **iSCSI**, il n'y a pas de « zoning » au sens FC ; l'isolation se fait par **VLAN dédiés**, **CHAP** (authentification initiateur-cible), et surtout par **LUN masking** côté cible (logiciel iSCSI ou baie). Ces notions peuvent être simulées en laboratoire avec des cibles iSCSI logicielles, sans matériel FC dédié.

Le stockage Online

Le **stockage online** (cloud storage) désigne des capacités de stockage fournies via Internet par un tiers (cloud public ou privé managé). Les données sont hébergées dans des datacenters du fournisseur.

Intérêts : élasticité, pas d'investissement matériel initial, gestion opérationnelle déléguée, multi-sites et résilience. **Points de vigilance :** coût à long terme, latence, conformité et souveraineté des données, dépendance au fournisseur.

Les acteurs du marché : volumétries proposées et coûts

Acteur	Offres principales	Volumétries / coûts (ordres de grandeur)
AWS	S3, EBS, EFS, Glacier	S3 : ~0,023 $/Go/mois (Standard) ; EBS : ~0,10 $/Go/mois
Azure	Blob, Disques, Files	Blob : ~0,018 €/Go/mois ; Disques managés selon type
Google Cloud	Cloud Storage, Persistent Disk	Storage : ~0,02 $/Go/mois ; PD selon type
OVH	Object Storage, Block Storage, NAS	Object : ~0,01 €/Go/mois ; volumes bloc et NAS selon gamme
Scaleway	Object Storage, Block, NFS	Object : ~0,01 €/Go/mois ; Block et NFS variables

Les **volumétries** sont quasi illimitées côté objet ; les **coûts** varient selon classe

(hot/cool/archive), réplication et région. Il faut comparer stockage seul + trafic sortant + requêtes (objet).

Le stockage en mode IaaS

En **IaaS** (Infrastructure as a Service), le fournisseur livre des **ressources brutes** : VM, réseaux, **volumes de stockage bloc** (disques virtuels) et/ou **stockage objet**.

- **Volumes bloc** : attachés à une VM (ex. EBS, disques Azure, volumes Block OVH). Usage : OS, BDD, applications.
- **Stockage objet** : buckets (S3, Blob, etc.) pour fichiers, sauvegardes, data lakes.
- **Responsabilité** : vous gérez OS, BDD, applications et politiques de sauvegarde ; le fournisseur assure la disponibilité et la durabilité du stockage sous-jacent.

Le stockage en mode PaaS

En **PaaS** (Platform as a Service), vous consommez des **services managés** (BDD, file d'attente, stockage) sans gérer les VM ni les disques.

- **Bases de données managées** : RDS, Azure SQL, Cloud SQL — le stockage est inclus et géré (backups, patches).
- **Stockage objet** : mêmes services que en IaaS (S3, Blob) mais souvent intégrés à des runtimes (Lambda, Functions) et à des services analytics (Athena, BigQuery).
- **Avantage** : moins d'ops, scaling et haute dispo gérés par le fournisseur.

Le stockage en mode SaaS

En **SaaS** (Software as a Service), vous utilisez une **application** qui stocke les données chez l'éditeur (ex. Google Drive, Dropbox, SharePoint, Salesforce). Le stockage est **inclus** dans le service ; vous ne gérez ni disques ni buckets.

- **Usages** : documents collaboratifs, CRM, messagerie, partage de fichiers.
- **Points d'attention** : conditions d'hébergement, conformité (RGPD, clauses contractuelles), export et réversibilité des données.

Les protocoles d'accès, l'intégration avec un annuaire

Protocoles d'accès

- **Bloc** : **iSCSI** (IP), **Fibre Channel**, **FCoE** — pour montage de volumes par les OS.
- **Fichier** : **NFS** (v3/v4), **SMB/CIFS** — pour partages réseau.
- **Objet** : **S3** (REST), **Swift** (OpenStack) — accès par API HTTP et SDK.

Le choix dépend du type de stockage (bloc / fichier / objet) et de l'écosystème (Linux/Windows, cloud, on premise).

Intégration avec un annuaire

L'**annuaire** (LDAP, Active Directory) centralise les **identités** et **groupes**. Pour le stockage :

- **Authentification** : accès NAS/SMB ou services cloud (S3, partages) via identité annuaire (SSO, SAML, OAuth).
- **Autorisations** : mapping utilisateur/groupes annuaire → droits sur partages, buckets ou volumes (ACL, politiques IAM côté cloud).
- **Audit** : corrélation logs d'accès stockage ↔ identités annuaire.

Cela permet une **gestion centralisée** des droits et une **traçabilité** des accès.

Gestion de la sécurité des données

La sécurité du stockage couvre **confidentialité**, **intégrité** et **disponibilité**.

Domaine	Mesures principales
Accès	Authentification forte, annuaire, principe du moindre privilège, révision des droits
Chiffrement	Au repos (FDE, chiffrement baie/volume) ; en transit (TLS, FC crypté si applicable)
Sauvegardes	Règles 3-2-1, sauvegardes hors site, tests de restauration, rétention conforme
Conformité	RGPD, sectoriel, classification des données, droit à l'oubli et purge
Réseau	Zoning SAN, LUN masking, VLAN dédiés, pare-feu et filtrage pour NFS/SMB et APIs
Monitoring	Logs d'accès, alertes (anomalies, quota, panne), corrélation avec l'annuaire

En **cloud**, s'appuyer sur IAM, politiques bucket/container, chiffrement managé et options de conformité du fournisseur (certifications, clauses contractuelles).

Études de cas : stratégie complète de stockage des données (travail en groupes)

Les **études de cas** se font en **5 groupes**, chacun travaillant sur une entreprise différente avec des contraintes propres.
Pour chaque cas, le livrable est une **présentation orale** (8–10 minutes) appuyée par **quelques slides**.
Vous devez **analyser le contexte**, **cartographier les données**, puis **proposer VOTRE stratégie de stockage** (pas de solution imposée dans l'énoncé).

Étude de cas 1 – PME industrielle « MECAPRO » (On Premise historique)

Contexte (à réutiliser dans vos slides) :

- 150 personnes, 1 site principal (usine + bureaux) + 1 petit site distant (atelier / dépôt).
- Infra actuelle : 3 serveurs physiques vieillissants (AD, fichiers, ERP) dans une petite salle serveur sans vraie climatisation.
- Stockage : disques internes en RAID 5, aucun stockage mutualisé ; quelques disques USB externes pour les sauvegardes hebdomadaires.
- Données critiques : fichiers de production (plans, CAO, programmes machines), base de données ERP, partages bureautiques (compta, RH, direction).
- Aucun site de secours ; les sauvegardes ne quittent presque jamais le bâtiment.

Inventaire des serveurs et du stockage (situation de départ) :

- **Serveur 1 – Infra / Annuaire**
  - Rôle : Active Directory, DNS, éventuellement DHCP et impression.
  - Matériel : serveur physique tour ou rack d'ancienne génération.
  - Stockage : 4 disques internes en RAID 5 (~2 To utiles) contenant à la fois OS et données AD.
- **Serveur 2 – Fichiers bureautiques**
  - Rôle : partage de fichiers pour bureautique, RH, compta, direction, etc.
  - Matériel : serveur physique dédié.
  - Stockage : 6 disques en RAID 5 (~6 To utiles), taux de remplissage ~80 %, pas de séparation claire données actives / archives.
- **Serveur 3 – ERP**
  - Rôle : application ERP de production + base de données (SQL) associée.
  - Matériel : serveur physique avec CPU et RAM dimensionnés à l'origine pour une charge plus faible.
  - Stockage : 4 disques en RAID 5 (~4 To utiles) avec forte activité disque en journée ; sauvegardes BDD locales sur le même serveur.
- **Sauvegardes & stockage externe**
  - 3–4 disques USB externes tournants pour les copies hebdomadaires.
  - Aucune baie SAN/NAS centralisée, pas de stockage dans un second site.

Contraintes / éléments à prendre en compte :

- Budget contraint mais perte de plusieurs jours de production = coût très élevé.
- Fenêtres de maintenance réduites (l'usine tourne en 2x8).
- Sensibilité à la **perte de données** historiques (traces de production, anciens plans).
- Croissance estimée des données : ~10 To aujourd'hui, +15 %/an.

Questions à traiter dans la présentation :

- Quelles **catégories de données** distinguez-vous ? (production, ERP, bureautique, logs, sauvegardes…)
- Quelles **priorités** en termes de disponibilité et de reprise (RPO/RTO) pour chaque

catégorie ?

- Quelle **architecture de stockage globale** imaginez-vous pour les 5 prochaines années ? (sans détailler tous les produits, mais en expliquant les grands blocs : local / distant, mutualisé / non mutualisé, etc.)
- Comment organisez-vous la **sauvegarde**, l'**archivage** et la **reprise après sinistre** ?
- Quelles **règles de sécurité et de gouvernance** (droits, séparation des environnements, accès distant) vous semblent indispensables ?

Étude de cas 2 – Start-up SaaS « BookMe » (100 % cloud)

Contexte (à réutiliser dans vos slides) :

- Start-up qui vend une application SaaS de réservation (restaurants, salles de sport, co-working…).
- Infra hébergée chez un **fournisseur de cloud public** (un seul pour l'instant).
- Données manipulées :
    - Comptes utilisateurs (clients finaux, restaurateurs…)
    - Réservations, historique des actions
    - Fichiers associés (logos, photos, pièces jointes simples)
    - Logs applicatifs et techniques
    - Sauvegardes / snapshots de bases de données
- L'activité est très saisonnière (pics le week-end et sur certaines périodes).

Inventaire des serveurs et du stockage (situation de départ) :

- **Frontends web**
    - Plusieurs instances (VM ou containers) derrière un équilibreur de charge managé.
    - Code statique et assets parfois servis directement depuis un bucket de stockage objet ou un CDN.
- **API / backend**
    - Pool de serveurs applicatifs (VM/containers) connectés à la base de données principale.
    - Pas de stockage persistant local significatif, hors logs temporaires.
- **Base de données principale**
    - Service de base relationnelle managée du cloud (cluster mono-région).
    - Volumes bloc répliqués au sein de la même région ; snapshots automatiques journaliers.
- **Stockage de fichiers**
    - 1 ou plusieurs buckets de stockage objet pour logos, photos et pièces jointes.
    - Versioning parfois activé, mais sans politique de cycle de vie claire.
- **Logs & métriques**
    - Service de logs managé, rétention courte (quelques jours/semaines).
    - Export ponctuel vers stockage objet pour certains audits.

Contraintes / éléments à prendre en compte :

- Forte croissance prévue (x2 utilisateurs / an).
- Objectifs de disponibilité élevés (SLA marketing 99,9 %).
- Respect du **RGPD** : données d'utilisateurs européens, droit à l'oubli, portabilité.
- Maîtrise des **coûts cloud** (stockage mais aussi trafic sortant, requêtes…).

Questions à traiter dans la présentation :

- Quelles **familles de données** identifiez-vous et quelles sont leurs caractéristiques (taille, fréquence d'accès, durée de conservation) ?
- Comment répartiriez-vous ces données entre **stockage "chaud"**, "tiède" et "froid" ?
- Quelle stratégie globale de **sauvegarde** et de **rétention** mettriez-vous en place (sans forcément choisir des services précis) ?
- Comment gérez-vous les aspects **sécurité / confidentialité / RGPD** pour ces différentes familles de données ?
- Quels sont, selon vous, les **risques principaux** (techniques, réglementaires, économiques) sur le stockage, et comment votre stratégie les adresse-t-elle ?

Étude de cas 3 – Hôpital « Saint-Martin » (santé, haute criticité)

Contexte (à réutiliser dans vos slides) :

- Hôpital général (urgences, blocs opératoires, imagerie, hospitalisation).
- Systèmes principaux :
  - Dossier Patient Informatisé (DPI)
  - Système d'imagerie médicale (PACS)
  - Applications administratives (admission, facturation, RH)
- Stockage actuel dispersé : plusieurs baies anciennes, NAS non homogènes, sauvegardes historiques sur bandes.
- Des interruptions de service ont déjà provoqué des retards de prise en charge.

Inventaire des serveurs et du stockage (situation de départ) :

- **Plateforme DPI**
  - 2 serveurs applicatifs virtualisés (cluster de virtualisation interne).
  - 1 serveur base de données (physique ou VM dédiée) très sollicité.
  - Stockage : LUN présentés par une baie SAN ancienne génération, RAID mixte (5/10).
- **Plateforme PACS (imagerie)**
  - 1 serveur d'indexation / orchestration.
  - Plusieurs serveurs de stockage d'images connectés à un ou plusieurs NAS.
- **Serveurs administratifs**
  - Plusieurs VMs (admission, facturation, RH, finances) hébergées sur la même infrastructure de virtualisation.
  - Données réparties entre NAS et disques locaux selon l'historique des projets.
- **Stockage central**
  - 1 baie SAN principale approchant la saturation en capacité et en IOPS.
  - 2 à 3 NAS de marques/générations différentes (PACS récent, ancien PACS,

fichiers administratifs).

- **Sauvegardes**
    - Bibliothèque de bandes LTO sur le site principal.
    - Jobs de sauvegarde nocturnes pour DPI, PACS, applicatifs ; quelques bandes stockées à l'extérieur mais sans procédure rigoureuse.

Contraintes / éléments à prendre en compte :

- Données de santé à **forte sensibilité** : secret médical, réglementation spécifique.
- Exigence très forte de **disponibilité** pour certains services (urgences, bloc).
- Durées de conservation longues, parfois plusieurs décennies, voire à vie.
- Besoin d'un **PRA multi-site** réaliste (rôle de l'hébergeur de données de santé, éventuel cloud, liens entre sites).

Questions à traiter dans la présentation :

- Comment **classifiez-vous** les données de l'hôpital (critères, catégories, priorités) ?
- Quelle **architecture de stockage** imaginez-vous entre le site principal, un éventuel site de secours et/ou un hébergeur externe ?
- Comment conciliez-vous **performance**, **capacité**, **coût** et **conformité réglementaire** ?
- Quels mécanismes de **traçabilité** des accès aux données de santé jugez-vous indispensables ?
- Comment organisez-vous la **continuité d'activité** en cas de panne grave de la plateforme de stockage ?

Étude de cas 4 – Groupe de retail « FashStore » (multi-sites)

Contexte (à réutiliser dans vos slides) :

- 50 magasins physiques répartis sur le territoire + un siège avec datacenter.
- Dans chaque magasin : un ou plusieurs points de vente (caisse), un petit serveur local ou une box "tout-en-un".
- Données :
    - Tickets de caisse, transactions du jour
    - Stocks locaux
    - Paramètres de prix / promotions
    - Données clients (fidélité) synchronisées avec le siège
- Au siège : consolidation globale des ventes, reporting BI, pilotage des stocks, campagnes marketing.

Inventaire des serveurs et du stockage (situation de départ) :

- **Dans les magasins :**
    - 1 à 2 postes de caisse (terminaux POS) par magasin, avec stockage local minimal (quelques dizaines de Go) pour bufferiser les tickets de la journée.
    - 1 petit serveur ou box locale jouant le rôle de cache et de relais (base articles/prix, synchronisation avec le siège) avec 1 ou 2 disques, parfois sans RAID.
- **Au siège :**

    - 1 cluster de virtualisation hébergeant les serveurs d'applications back-office et les outils de reporting.
    - 1 ou plusieurs serveurs de bases de données centrales (ventes, stocks, fidélité) connectés à une baie de stockage.
    - 1 serveur de fichiers ou NAS pour exports, rapports BI et échanges interservices.
- **Sauvegardes :**
    - Sauvegardes centralisées au siège (BDD + fichiers) sur disque/NAS puis sur bande ou vers un autre site.
    - Sauvegardes locales dans les magasins très variables (de la simple copie manuelle à aucune sauvegarde formalisée).

Contraintes / éléments à prendre en compte :

- Les magasins doivent pouvoir **continuer à encaisser** même en cas de coupure WAN.
- Les données doivent remonter régulièrement au siège pour l'analyse et les décisions.
- Les magasins n'ont pas de personnel IT dédié (simplicité et robustesse importantes).

Questions à traiter dans la présentation :

- Quelles données doivent absolument rester **disponibles localement** en magasin ? Lesquelles peuvent être **uniquement centrales** ?
- Comment imaginez-vous le **flux des données** entre magasins et siège (fréquence, sens, mode) ?
- Quel compromis faites-vous entre **simplicité dans les magasins** et **contrôle centralisé** au siège ?
- Comment limitez-vous l'impact d'un incident de sécurité dans un magasin (ransomware, vol de matériel…) sur le reste du système ?
- Quelles priorités de **restauration** donneriez-vous en cas de sinistre au siège ?

Étude de cas 5 – Agence de communication « Pixel&Co » (gros fichiers, collaboration)

Contexte (à réutiliser dans vos slides) :

- 80 personnes (directeurs artistiques, graphistes, monteurs vidéo, chefs de projet).
- Typologie des fichiers : photos haute résolution, vidéos 4K, maquettes InDesign, livrables clients.
- Le travail se fait souvent en équipe sur les mêmes projets, avec beaucoup de versions successives.
- Télétravail fréquent, recours régulier à des freelances externes.

Inventaire des serveurs et du stockage (situation de départ) :

- **NAS principal au siège**
    - Rôle : stockage des projets en cours et d'une partie des archives récentes.
    - Capacité : plusieurs dizaines de To, proche de la saturation.

    - Accès : partages SMB pour tous les postes internes du studio.
- **NAS secondaire / disques externes**
    - Utilisés pour décharger le NAS principal (archives plus anciennes, projets terminés).
    - Organisation hétérogène (dossiers par années, clients, parfois doublons).
- **Postes de travail fixes (graphistes / vidéo)**
    - Stations puissantes avec SSD local pour le travail en cours sur certains médias.
    - Synchronisation manuelle ou semi-automatisée entre SSD local et NAS.
- **Accès distant / freelances**
    - Accès via VPN au NAS du siège ou via un service de partage de fichiers cloud mis en place progressivement.
    - Certains projets partagés en parallèle sur le cloud et le NAS, sans gouvernance claire.

Contraintes / éléments à prendre en compte :

- Performance sur les **gros fichiers** (éviter les téléchargements multiples énormes).
- Besoin de **collaboration** fluide (plusieurs personnes sur un même projet, parfois en même temps).
- Nécessité de garder une **trace des versions** et de pouvoir revenir en arrière.
- Besoin d'accès externe sécurisé (freelances, clients pour validation).

Questions à traiter dans la présentation :

- Comment organisez-vous la **vie d'un projet** du point de vue des données (création, travail actif, validation, archivage) ?
- Où placez-vous la **frontière** entre stockage interne et stockage dans un service externe (cloud, SaaS...) ?
- Comment assurez-vous la **cohérence** et la **non-duplication** des projets (éviter 10 copies d'un même rush vidéo) ?
- Quelle est votre approche pour la **sauvegarde** et l'**archivage** des projets terminés (où, combien de temps, sous quelle forme) ?
- Comment conciliez-vous **télétravail**, **performance** et **sécurité** des données des clients ?

Glossaire des principaux acronymes

Acronyme	Signification	Rappel rapide
SAN	Storage Area Network	Réseau dédié au stockage bloc, exposant des LUN aux serveurs.
NAS	Network Attached Storage	Serveur/baie qui expose du stockage sous forme de partages de fichiers (NFS/SMB).
RAID	Redundant Array of Independent Disks	Ensemble de disques combinés pour la performance et/ou la tolérance aux pannes.
LUN	Logical Unit Number	Volume logique bloc présenté par une baie ou une cible iSCSI à un serveur.
RPO	Recovery Point Objective	Perte de données maximale acceptable en cas d''incident (en temps).
RTO	Recovery Time Objective	Durée maximale acceptable pour remettre un service en route après incident.
IOPS	Input/Output Operations Per Second	Nombre d''opérations d''E/S par seconde que peut traiter un stockage.
DPI	Dossier Patient Informatisé	Dossier médical électronique dans le contexte hospitalier.
PACS	Picture Archiving and Communication System	Système d''archivage et de diffusion d''imagerie médicale.
SLA	Service Level Agreement	Engagement de niveau de service (disponibilité, performance) entre un fournisseur et un client.
VPN	Virtual Private Network	Tunnel chiffré permettant de relier un utilisateur ou un site distant à un réseau privé.
IaaS	Infrastructure as a Service	Fourniture de ressources brutes (VM, réseaux, stockage) par un fournisseur cloud.
PaaS	Platform as a Service	Fourniture de plateformes applicatives managées (BDD, runtimes...) sans gestion des VM.
SaaS	Software as a Service	Application clé en main consommée via Internet, stockage inclus chez l''éditeur.
RGPD	Règlement Général sur la Protection des Données	Règlement européen sur la protection des données personnelles.

Résumé du module (Jour 1 + Jour 2)

Thème	Points clés
Données	ADN de l'entreprise ; gouvernance et cycle de vie essentiels
Types	Fichier à plat (documents, data lake) ; BDD (transactionnel, BI) ; bloc / fichier / objet
On Premise	Baies, RAID (5/6/10), SAN (bloc, zoning, LUN masking), NAS (fichier)
Online	IaaS (volumes + objet), PaaS (BDD managées, objet), SaaS (stockage inclus)
Accès	Protocoles bloc (iSCSI, FC), fichier (NFS, SMB), objet (S3) ; annuaire pour identité et droits
Sécurité	Chiffrement, sauvegardes, conformité, contrôle d'accès et audit

Ce module donne les bases pour **identifier** les types de stockage en entreprise, les **positionner** selon les usages et les contraintes, et participer à la **sécurisation** et à la **gouvernance** des données.