

A photograph of two men in a server room. The man on the left has curly brown hair and a beard, wearing a green button-down shirt over a black t-shirt and a lanyard with an ID badge. The man on the right is bald with a beard, wearing a blue shirt. They are both looking intently at a laptop screen. The background shows rows of black server racks with perforated doors.

CHECKLIST IMPRIMABLE

Projet IA : la checklist données et infrastructure NetApp

Préparez votre projet IA avant de choisir ou de dimensionner l'infrastructure. Cette fiche vous aide à clarifier les données, les performances attendues, les accès et la protection.

Au sommaire

- 1 Le rôle de NetApp, en une minute
- 2 1. Cartographier ce que le projet doit conserver
- 3 2. Distinguer les critères de dimensionnement
- 4 3. Organiser les accès, les flux et la protection
- 5 4. Valider l'architecture par un pilote

Le rôle de NetApp, en une minute

NetApp soutient l'infrastructure de données : stockage, partage, protection, gouvernance et mobilité des données.

La pertinence dépend notamment des volumes, des accès simultanés, des exigences de sécurité et de l'architecture déjà en place.

NetApp ne remplace pas les GPU ou CPU, les modèles d'IA, les outils de machine learning ni les compétences data.

Une grande capacité de stockage ne garantit pas seule la fluidité des entraînements.



Données sources



Stockage et accès



Réseau



Calcul IA



Résultats et modèles



Protection et
restauration

1. Cartographier ce que le projet doit conserver

Recenser les sources de données

Fichiers tabulaires, images, vidéos, textes, données de capteurs ou journaux applicatifs.

Identifier les jeux de données d'entraînement et de validation

Préciser leur origine et leurs transformations.

Prévoir les versions de jeux de données

Chaque nettoyage, transformation ou annotation peut créer une nouvelle version.

Lister les artefacts produits par les expérimentations

Checkpoints, paramètres, résultats d'évaluation et artefacts de modèles.

Repérer les données sensibles

Définir les contraintes de sécurité, de traçabilité et de localisation.

Prévoir les journaux à conserver

Ils participent au suivi du modèle et des traitements en production.

2. Distinguer les critères de dimensionnement

Critère	Question à poser	Point de vigilance
Capacité	Combien de données faut-il conserver ?	Ne pas confondre volume total et performance d'accès.
Débit	Quel volume de données doit être transféré ?	Des données arrivant trop lentement peuvent sous-utiliser le calcul.
Latence	Quel délai de réponse est acceptable ?	À évaluer selon les opérations et les charges.
Accès concurrents	Combien d'utilisateurs, processus ou équipes lisent et écrivent simultanément ?	Les accès multiples peuvent saturer l'infrastructure ou le réseau.
Disponibilité	Quelles données doivent rester accessibles ?	Prévoir les besoins de restauration et de continuité.

3. Organiser les accès, les flux et la protection

Définir les droits d'accès par rôle

Data scientists, ingénieurs ML et équipes IT n'ont pas nécessairement les mêmes besoins.

Documenter les flux entre les environnements

Site interne, cloud privé et cloud public.

Éviter les copies inutiles

Le stockage doit aider les équipes à retrouver et partager les données sans multiplication non maîtrisée.

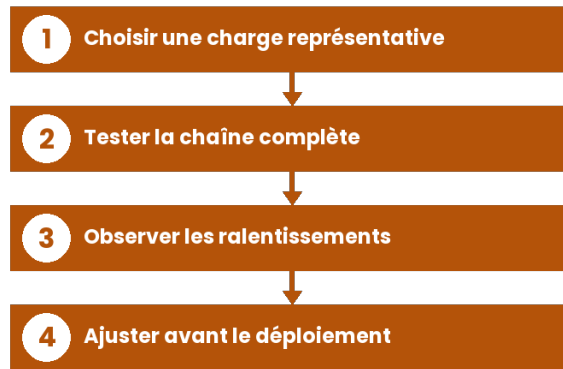
Définir la stratégie de protection

Sauvegarde, restauration, réplication et conservation des versions importantes.

Relier les résultats à leur contexte

Conserver la possibilité de retrouver données, code, paramètres et résultats associés.

4. Valider l'architecture par un pilote



1 Choisir une charge représentative

Utilisez des données, des transformations et des accès proches des conditions réelles du projet.

2 Tester la chaîne complète

Évaluez ensemble le stockage, le réseau, le calcul et les règles d'accès.

3 Observer les ralentissements

Vérifiez si les données sont disponibles assez rapidement pour les charges d'entraînement et de validation.

4 Ajuster avant le déploiement

Utilisez les enseignements du pilote pour adapter l'architecture plutôt que de vous fonder uniquement sur un dimensionnement théorique.

! Point de vigilance

Ne cherchez pas à résoudre un problème de modèle, de pipeline de données ou de ressources de calcul par le seul stockage. Dans un projet IA, les briques données, réseau, calcul, supervision et sécurité doivent être cohérentes.

Cette checklist aide à cadrer une infrastructure de données pour un projet IA ; elle ne remplace pas l'évaluation technique des charges réelles.