

A man with curly hair and glasses is looking down at a laptop in a server room. He is wearing a blue sweater and a lanyard. The background shows server racks with many blue and black cables.

CHECKLIST DE DÉCISION IMPRIMABLE

Votre checklist pour valider un stockage IA NetApp

Avant de dimensionner une infrastructure, validez les flux de données, les contraintes opérationnelles et les performances dans des conditions proches de la production. Cette fiche vous aide à structurer un pilote et à décider sur des mesures concrètes.

Au sommaire

- 1 Comment utiliser cette fiche
- 2 1. Cartographier les données et leur cycle de vie
- 3 2. Relier le profil IA aux exigences de stockage
- 4 3. Préparer un pilote représentatif
- 5 4. Prendre la décision d'architecture

Comment utiliser cette fiche

Cochez les points au fil de l'étude. L'objectif n'est pas de choisir une baie sur sa capacité affichée, mais de vérifier que l'ensemble stockage, calcul, réseau et pipeline de données répond aux besoins réels de votre charge IA.

1. Cartographier les données et leur cycle de vie

- ☐ **Recenser les données sources**
Documents, images, fichiers métier, exports applicatifs et données non structurées.
- ☐ **Identifier les données de travail**
Jeux d'apprentissage et de validation, caches, données transformées et fichiers temporaires.
- ☐ **Isoler les actifs IA critiques**
Modèles, embeddings, checkpoints, résultats d'inférence et journaux techniques.
- ☐ **Qualifier les données de conservation**
Archives, sauvegardes et éléments soumis à des obligations internes ou contractuelles.
- ☐ **Associer à chaque catégorie le niveau requis de performance, de protection et de conservation**
Évitez de placer toutes les données sur le même niveau de stockage.

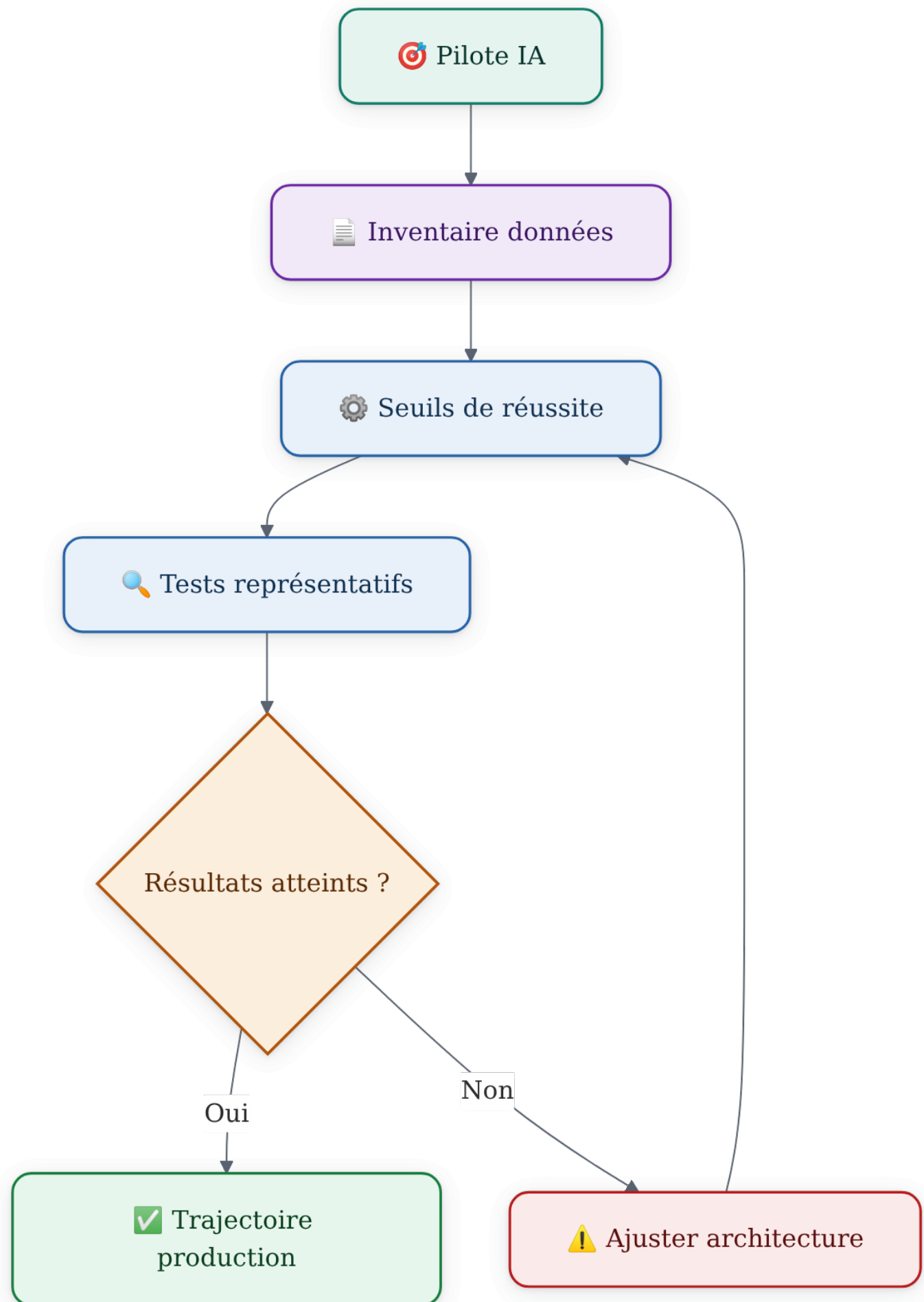
2. Relier le profil IA aux exigences de stockage

Profil de charge	Accès et données à examiner	Point de vigilance
Entraînement	Lectures de gros volumes et écritures régulières de checkpoints	Protéger les checkpoints : leur perte peut imposer de relancer des calculs.
Inférence	Accès réguliers aux modèles, embeddings, journaux et données métier	Vérifier la régularité des accès, pas seulement un pic de débit.
IA générative / RAG	Données métier, documents, embeddings et recherche	Évaluer l'architecture complète et la compatibilité avec les composants retenus.
Analytique	Volumes variables, traitements et données préparées	Distinguer les données actives des archives pour maîtriser les coûts.

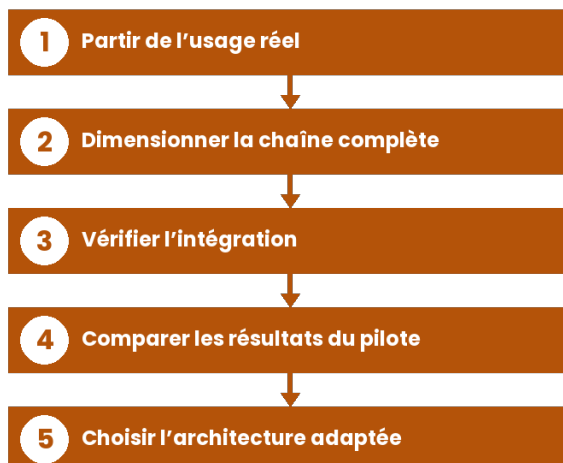
3. Préparer un pilote représentatif

- ☐ **Reproduire des jeux de données, des tailles de fichiers et des accès proches de la production**

- ☐ **Simuler les lectures séquentielles massives, les petits fichiers et les accès concurrents**
- ☐ **Tester les écritures de checkpoints et le comportement des métadonnées**
- ☐ **Mesurer le débit agrégé observé par les nœuds de calcul**
- ☐ **Mesurer le débit par flux, la latence de réponse et le taux d'erreurs**
- ☐ **Consigner le nombre de clients simultanés et la taille moyenne des fichiers**
- ☐ **Tester et chronométrer une restauration**
La performance nominale ne suffit pas si la reprise ne répond pas au besoin opérationnel.



4. Prendre la décision d'architecture



1 Partir de l'usage réel

Définissez si le projet vise l'entraînement, l'inférence, l'IA générative, la recherche sémantique ou plusieurs usages.

2 Dimensionner la chaîne complète

Évaluez ensemble les serveurs, les accélérateurs, le réseau, les protocoles, l'orchestration et les outils de préparation des données.

3 Vérifier l'intégration

Confirmez la compatibilité effective des versions logicielles, des frameworks et des modes d'accès nécessaires : fichiers, blocs ou objets.

4 Comparer les résultats du pilote

Confrontez les mesures de débit, latence, restauration, sécurité et coût total de possession aux exigences définies au départ.

5 Choisir l'architecture adaptée

Envisagez une solution locale telle qu'AIPod Mini ou une architecture plus structurée telle que FlexPod avec Cisco selon le périmètre et les résultats validés.

! Point de contrôle essentiel

Le stockage IA NetApp est une composante de l'infrastructure de données, pas une plateforme d'IA complète. Une baie performante ne compense ni un réseau saturé, ni des GPU sous-dimensionnés, ni un pipeline de données mal conçu.

Décision à fonder sur un pilote représentatif et sur les contraintes propres à votre environnement, vos données et vos logiciels.