

The background image shows a group of students sitting in a lecture hall or classroom. They are viewed from behind, focused on their work. Several students are using laptops, and one student in the foreground is using a tablet. The setting appears to be a modern educational environment with wooden floors and simple wooden chairs.

TDS

v1.0

Système détection de triches

CdP: Anh-Hoang LE

Caractéristiques clés de TDS

Vitesse de traitement:

- automatique (recognition & annotation): 1 user/s *
- manuel (correction & export): 1 user/min *

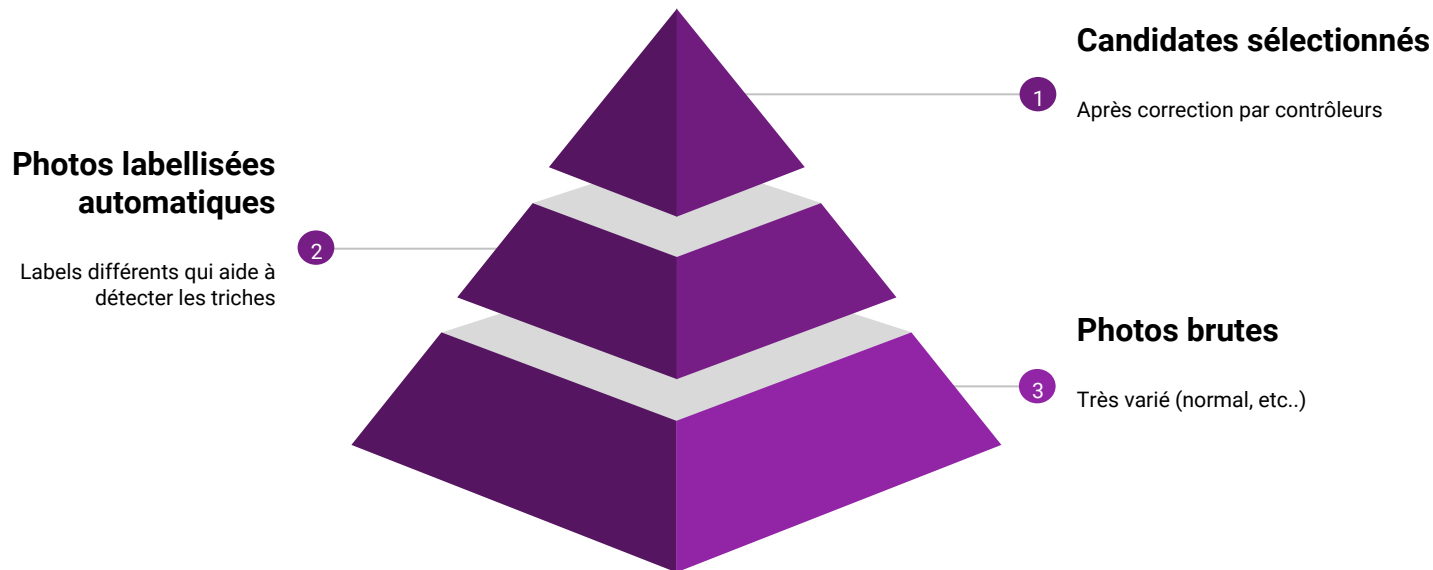
Cost:

- recognition + detect-faces: 0.2€/user **

** 1 user contains ~ 50 photos*

*** basé sur le tarif AWS 2018*

Pyramid de résultats



Comment gérer le Big Data ?

- Volume
 - S3 bucket (common solution)
- Vitesse
 - Queue-Job (common solution)
- Variété
 - Expert system & Machine Learning

Comment détecter les triches ?

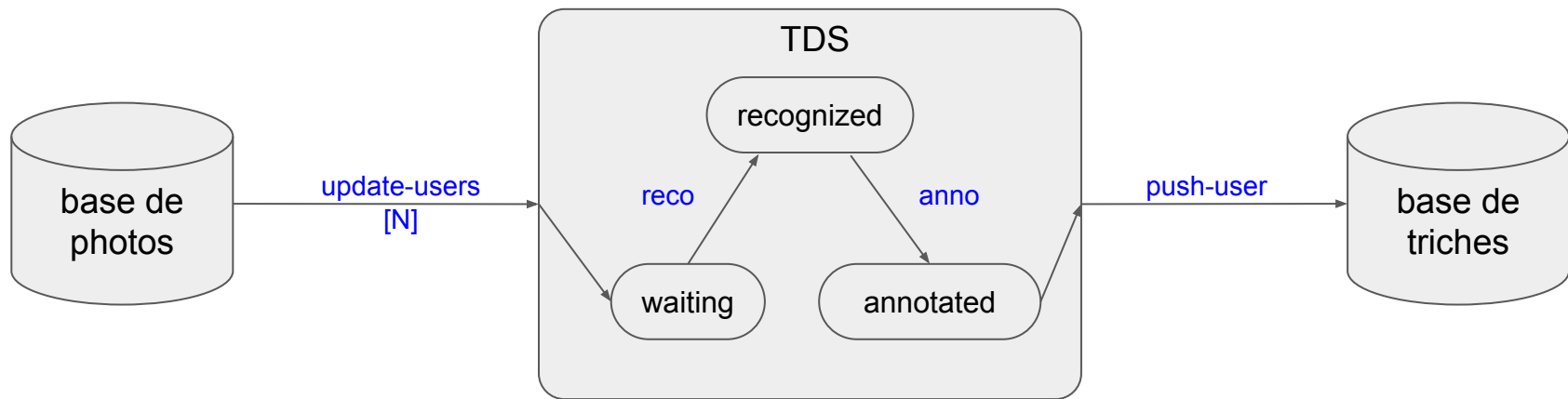
- Classification de triches
 - Règles: strict (évidence) => raisonnable (suspension)
- Bigdata
 - ML

Comment traiter rapidement les mégadonnées ?

- Problem statement
 - 3 jours pour traiter 2500 utilisateurs (1 personne)
- Solution
 - Optimisation
 - Cost function: priorité des triches, omit des cas normaux
 - PLN (Probabilité Logic Network)
 - Decision Making System (basé sur les labels sélectionnés)

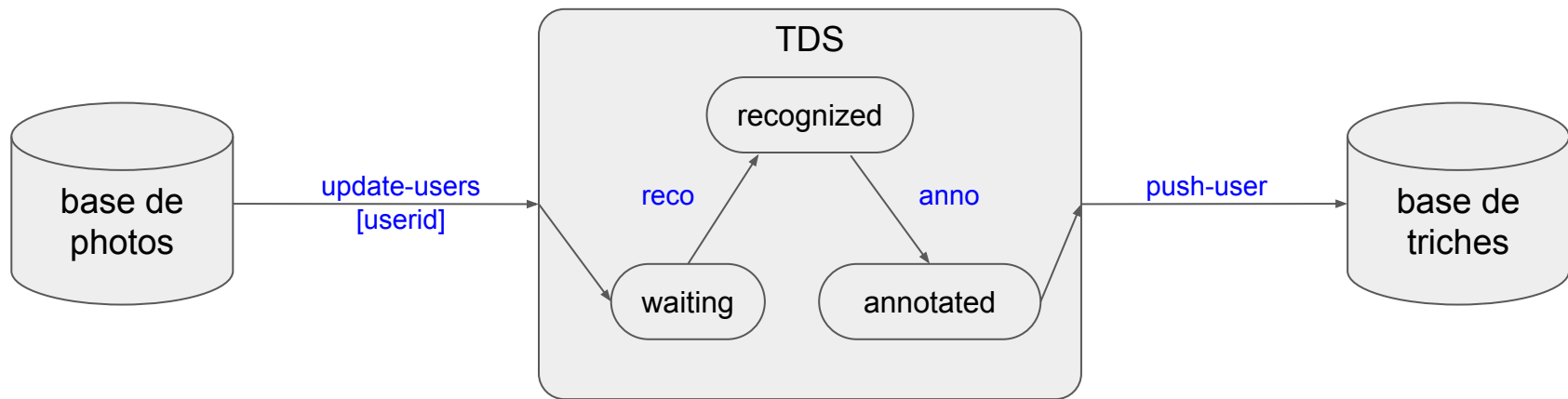
TDS APIs

- Batch mode: 50 users/run



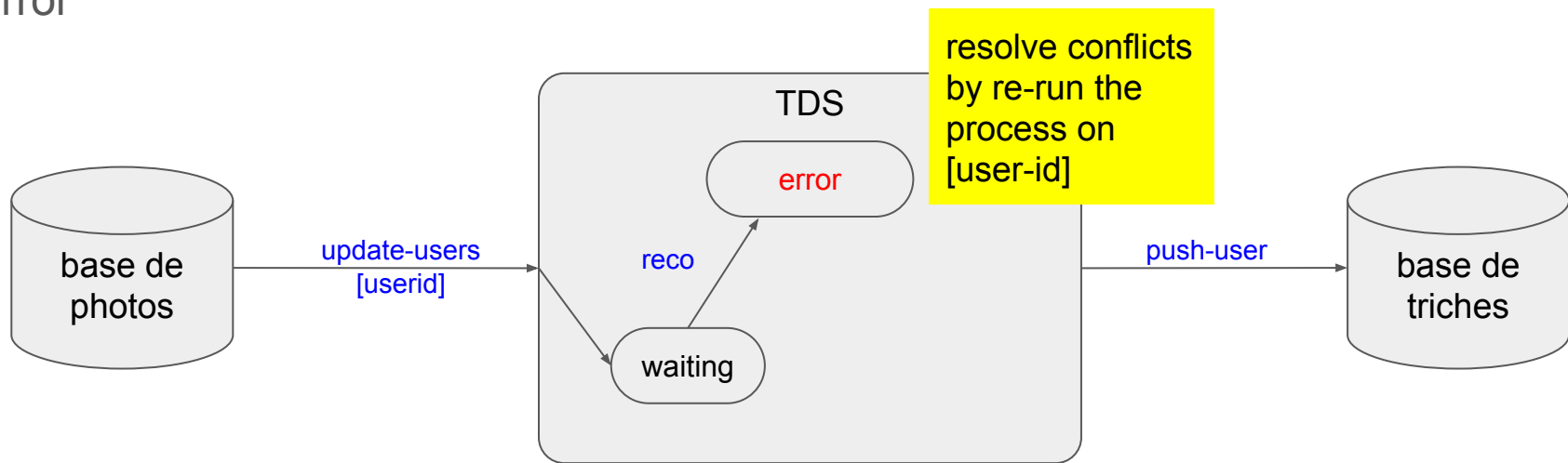
TDS APIs

- User mode: 1 user/run



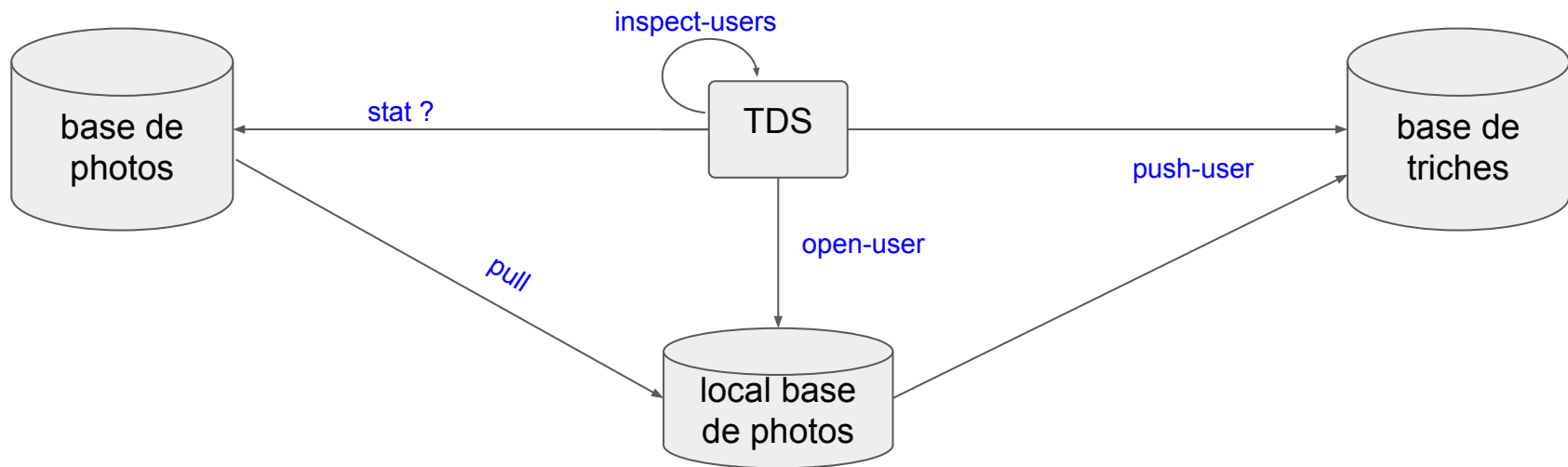
TDS APIs

Error



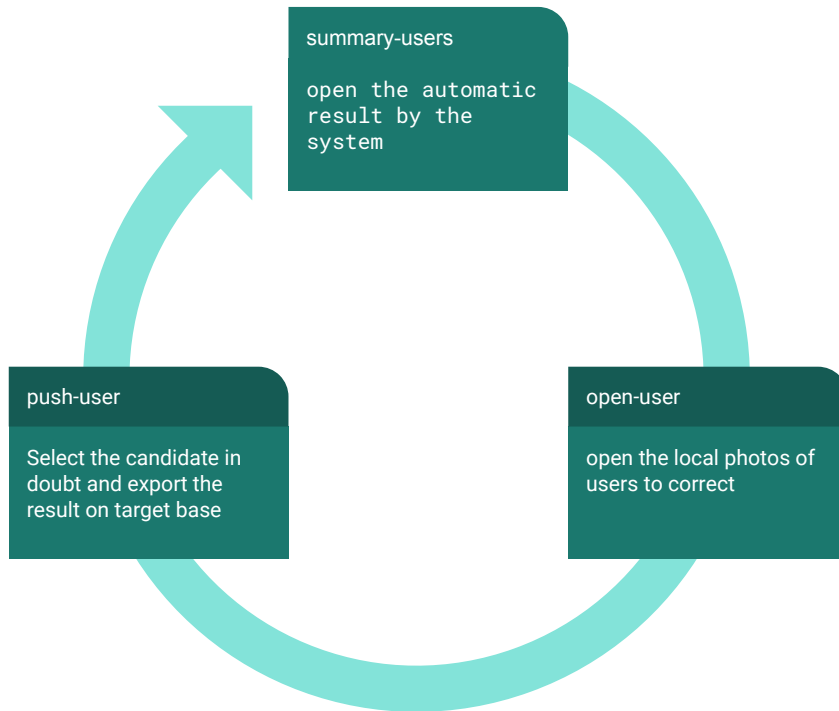
TDS APIs

Query, Sync, Correction



TDS APIs

Correction



TDS APIs

Priorité de traitement manuel, les candidates avec :

N1. Labels rouges (Many Persons, Different player)

N2. Les photos mauvaise qualités (Unknown)

N3. Les joueurs ne sont pas stables (High variance)

Si N (maximum candidates traités dans le délai 3 jours) $< N1 + N2$:

Seul N1 va être traité

Conclusion & Perspective

- Pros

- TDS a capacité de traiter les mégadonnées (big data)
 - vitesse, volume, et variété
- Précision suffit pour classification
- Document, portabilité (NodeJS)

- Cons

- Tags sont limités
- Méthods de traitement sont encore limités (dépend le service AWS ML)

- Future work

- Ajoute les nouveaux methods pour augmenter la précision du système
- Tester avec plus que donnés (> 10 000 users au lieu de 2 000 users)

Planning for future work

Voici les travaux intéressants pour améliorer ce travail (v2):

- Convert les APIs CLI en RESTful (2 - 3 jours)
- Ajouter les méthodes de traitement des images supplémentaires, OpenCV (? jours)

Merci à tous