

Challenge - data ingé

Pipeline du journal d'erreur du robot

Objectif:

Votre objectif est d'ingérer et transformer un journal simulé de l'état opérationnel d'un robot.

Ces données sont des successions d'états d'erreurs.

Vous devez:

- parser ces données et les ingérer dans une base de donnée
- les transformers (à votre guise) puis:
 - Calculer la contribution par types de pannes dans le temps
 - Calculer un taux de dispo en % (hypothèse service qui tourne 24/24)
 - Le robot est considéré comme indisponible à partir du timestamp d'un log contenant au moins une erreur
- Afficher ces données dans grafana (format libre)

Données d'Entrée:

Un fichier CSV nommé

`simulated_error_logs.csv` peut être généré par le script `generate.py`.

Contraintes:

- Le process/script d'ingestion doit être différent de celui de transformation
 - Le script de transformation doit s'appuyer sur le 1er stockage issue de l'ingestion
- L'ingestion doit être robuste: que faire si une ligne n'est pas valide ?

- Les fichiers CSV peuvent être très volumineux

Livrables:

Vous fournirez dans une archive (.tar.gz) les éléments suivants:

1. **Code (python ou autre)**
2. **Outils utilisés (databases)**
 - a. scripts de lancement / amorçage (via docker)
3. **Grafana**
 - a. (Le script de lancement vous ai fourni (via docker)
 - b. Dashboard + Datasource (le dossier du volume devrait suffir)
4. **Un document d'analyse sur:**
 - a. Une alternative au format d'entrée (argumentée)
 - b. Les limites du système actuel
 - c. Une description des étapes en vue d'industrialiser cette pipeline (c.a.d lever les limites)
5. **Bonus**
 - a. Makefile
 - b. Script ou process d'orchestration / sheduling

Critères d'Évaluation:

- Précision des calculs du temps de dispo et de la contribution de chaque type d'erreur.
- Code et documentation bien structurés, commentés et compréhensibles.
- La démarche