

BIG DATA

6 - Mini projet



SOMMAIRE

- [Présentation](#)
- [Part 0 - Setup](#)
- [Part 1 - Download files](#)
- [Part 2 - Work count example](#)
- [Part 3 - Processing data](#)
- [Part 4 - Exposing data](#)
- [Rendu](#)

PRÉSENTATION

Objectifs :

- Prise en main : Python, PySpark
- Ingestion : Récupération du jeux de données
- Préparation : Transformation, Filtrage, Analyse des données
- Exposition : Création d'une vue Hive pour exposer les données

RENDU ATTENDU

Objectifs :

- Scripts
- Rapport PDF