*Nom de la team : RSA*

*Sprint n°0 - Projet N°6 Criminalité à Chicago*

*Scrum Master : Adrien Golebiewski*

*Teammates : Ryme Youb, Sacha Cymermann*

*Début : 06/09/2021*

*Fin : 20/09/2021*

**Sprint 0 - Définition des objectifs et du design de la solution technique**

1. **Définition de la source de données :**

Données criminalité Chicago csv - x3 fichiers

Issue du portail public Chicago Data Portal ( Open Data)

1. **Stockage et traitement des fichiers sources**

Utilisation d’Hadoop/HDFS

Environnement de travail depuis les postes fixes du CNAM pour accès au serveur Hadoop + transférer les fichiers de données CSV dans ce serveur

1. **Requêtage MapReduce (Hadoop/HDFS - YARN) pour calculs d’indicateurs pertinents résumant la complexité des crimes à Chicago.**

Exemple : *Nombre total de crimes par an*

Comparaison des résultats avec les indicateurs calculés avec Spark : temps de calcul, qualité des résultats, etc ...

1. **Prédiction** de la sévérité du crime et de son type

à l’aide de l’API Python dans Spark et de la librairie Python PySpark pour le Machine learning (librairie PySpark)

