كلية العلوم و التقنيات
FACULTÉ DES SCIENCES ET TECHNIQUES

# Rapport

## Extraction des Amis Communs avec PySpark

**Nom :** *Cheikhani Lemrabett*
**Filière :** *Machine Learning et Data Science*
**Date :** 1er juillet 2025

# Table des matières

# 1. Présentation du Projet

**Lightbulb Objectif Principal**

Développer une solution distribuée pour identifier les amis communs entre utilisateurs d'un réseau social en utilisant Apache Spark et PySpark.

## 1.1 Contexte

L'analyse des réseaux sociaux est devenue cruciale dans le monde numérique actuel. Ce projet vise à explorer les relations d'amitié dans un graphe social et à identifier efficacement les connexions communes entre utilisateurs.

## 1.2 Objectifs Spécifiques

- ✓ Maîtriser les concepts de **MapReduce** appliqués aux graphes sociaux
- ✓ Utiliser **Apache Spark** pour le traitement parallèle de données volumineuses
- ✓ Implémenter un algorithme d'identification d'amis communs
- ✓ Analyser les performances du traitement distribué

# 2. Architecture du Projet

```
spark_friends_project/
 data/
    friends.txt        ← Données des relations d'amitié
 src/
    main.py            ← Code source principal
 run.sh                ← Script d'exécution
 README.md             ← Documentation
```

# 3. Installation et Configuration

**ExclamationTriangle Prérequis**

**Système requis :** Linux/Unix, Java 11+, Python 3.6+

## 3.1 Téléchargement de Spark

```
# T l chargement de Spark
wget https://archive.apache.org/dist/spark/spark-3.5.1/spark-3.5.1-bin-hadoop3.
    tgz
```

```
# Extraction et installation
tar -xvf spark-3.5.1-bin-hadoop3.tgz
mv spark-3.5.1-bin-hadoop3 ~/spark
```

Listing 1 – Installation d'Apache Spark

### 3.2 Configuration des Variables d'Environnement

```
# Configuration des variables Spark
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export SPARK_HOME=$HOME/spark
export PATH=$PATH:$SPARK_HOME/bin:$SPARK_HOME/sbin
export PYSPARK_PYTHON=python3
export PYSPARK_DRIVER_PYTHON=python3
```

Listing 2 – Configuration de l'environnement

### 3.3 Installation de PySpark

```
# Installation via pip
pip3 install pyspark

# V rification de l'installation
pip3 show pyspark
```

Listing 3 – Installation des dépendances Python

### 3.4 Test de Configuration

```
from pyspark.sql import SparkSession

# Cr ation d'une session Spark
spark = SparkSession.builder \
    .appName("ConfigurationTest") \
    .getOrCreate()

print("Spark configur  avec succ s!")
spark.stop()
```

Listing 4 – Test de fonctionnement

## 4. Algorithme et Implémentation

### 4.1 Principe de l'Algorithme

**Étapes de traitement :**

1. **Map :** Transformation des relations en paires (utilisateur, liste$_a mis$)
   **Combinaison :**$Gnrationdetouteslespairespossiblesd'utilisateurs$

2. **Reduce :** Calcul des intersections pour identifier les amis communs
3. **Filtrage :** Extraction des résultats pertinents

### 4.2 Structure des Données

Format du fichier `friends.txt`

```
1 Mohamed 2,3,4
2 Sidi 1,4,5
3 Ahmed 1,4
4 Fatima 1,2,3
5 Mariam 2
```

## 5. Résultats et Analyse

### 5.1 Exemple de Sortie

CheckCircle Résultat Obtenu

```
1<Sidi>2<Mohamed> ['Fatima']
```

**Interprétation :** Fatima est identifiée comme l'amie commune entre Mohamed (utilisateur 1) et Sidi (utilisateur 2).

### 5.2 Analyse des Performances

ChartLine **Complexité :** $O(n^2)$ pour n utilisateurs

ChartLine **Parallélisation :** Traitement distribué efficace

ChartLine **Scalabilité :** Adapté aux grandes bases de données

## 6. Exécution du Projet

Guide d'Exécution

```
# Navigation vers le projet
cd ~/spark_friends_project

# Rechargement de l'environnement
source ~/.bashrc

# Ex cution du script
./run.sh
```

### 6.1 Interface Web Spark

Une fois l'application lancée, accédez à l'interface de monitoring :

Globe `http://localhost:4040`

## 7. Conclusion et Perspectives

Ce projet démontre l'efficacité d'Apache Spark pour l'analyse de graphes sociaux. La solution développée permet de traiter efficacement les relations d'amitié et d'identifier les connexions communes dans un réseau social.

**Applications futures :**

- Recommandations d'amis
- Détection de communautés
- Analyse de l'influence sociale
- Optimisation des algorithmes de graphes