

Apache Kafka

IUT informatique de Annecy - R5.A.05 FA

1 Introduction

Objectif: mettre en pratique les connaissances théoriques vues en cours et en TD papier.

Prérequis:

- Connaître les services rendus par Kafka
- Savoir coder en Python

Connaissances à acquérir:

- Comprendre l'intérêt de Kafka pour la transmission de données
- Être capable d'utiliser Kafka dans une application réelle

Instructions: Réaliser ces exercices seul.

2 Installation de Kafka

- Installez la librairie `kafka-python`
- La documentation : <https://kafka-python.readthedocs.io/en/master/>

3 Téléchargement du dataset

- Téléchargez le dataset suivant : <https://www.kaggle.com/datasets/uciml/sms-spamcollectiondataset/>

Ce jeu de données est un ensemble d'environ 5000 sms spam ou légitimes (ham).

4 Créer une application de communication

- A l'aide du package `kafka-python` et du dataset présenté ci-dessus, créez une application contenant deux *topics* un pour les sms de type *spam* et un pour les sms de type *ham*.
- Affichez tous les messages *spam* reçus du côté sur un des consommateurs, et tous les messages *ham* sur l'autre.

5 Afficher les données

Utilisez Grafana comme un consommateur des données récupérées sur votre broker :

- Affichez un graphique montrant le nombre de messages *spam* et le nombre de messages *ham*.
- Affichez les statistiques de performance de Kafka (usage CPU, nombre de messages par topic, nombre de messages par broker, *etc.*)