**Домашнє завдання до теми «Spark streaming»**

**Виконання завдання:**

1. **Створення даних сенсорів та їх відправлення у Kafka (файл generator.py):**

* Використовуючи модуль **confluent\_kafka**, було створено Kafka Producer для передачі даних сенсорів у Kafka.
* Сенсори генерують дані, які включають ідентифікатор сенсора, температуру, вологість та час вимірювань.
* Дані сенсорів генеруються в рамках циклу, де для кожного сенсора створюється новий набір значень.
* Відображення даних у консолі здійснюється за допомогою бібліотеки **colorama**, що забезпечує кольорове зображення тексту для кращого візуального сприйняття.

A screen shot of a computer

Description automatically generated

1. **Аналіз потокових даних за допомогою PySpark (файл processor.py):**

* Застосування Apache Spark для читання і обробки даних з потоку Kafka, з конфігурацією, що включає механізми SASL для безпеки передачі.
* Потокові дані конвертуються з JSON, аналізуються для отримання середніх значень температури та вологості за визначені періоди часу.
* Ці дані порівнюються з умовами для сповіщень з файлу *alerts\_conditions.csv*, фільтруються згідно з порогами для температури та вологості.
* Результати представлено у консолі у вигляді JSON за кожен часовий інтервал, що дозволяє моніторити значення для сповіщень.
* Отримані результати підготовлені до повторної відправки у Kafka.

1. **Тестування правильності роботи системи:**

* Здійснено випробування шляхом створення декількох записів у Kafka, які підтверджують точність генерації та обробки даних.
* Результати фільтрації відображені через консоль, де можна переглянути, чи відповідають отримані середні значення температури та вологості встановленим критеріям.

A screenshot of a computer program

Description automatically generated