파이스파크로 패키지 옵션을 줘서 sql -kafka로 열기(뒤에 다운로드한 플러그인이랑 맞춰야함)

|  |
| --- |
| pyspark --packages org.apache.spark:spark-sql-kafka-0-10\_2.12:3.1.1 |

필요한 모듈을 먼저 불러오기

|  |
| --- |
| import pyspark  import pyspark.sql  from pyspark.sql import \*  import time  from pyspark.sql.functions import \*  from pyspark.sql.types import \*  from pyspark.sql import SparkSession |

불러온 후에

스파크세션 객체를 불러옴

|  |
| --- |
| spark = SparkSession \  .builder \  .master("local")\  .appName("api")\  .getOrCreate() |

데이타 프레임을 만들어줌

format에 카프카로 쓰고 option에 카프카 토픽명을 넣어줌

|  |
| --- |
| df = spark\  .read \  .format("kafka")\  .option("kafka.bootstrap.servers","localhost:9092")\  .option("subscribe","people2019")\  .load()  df.show() |

value값에

|  |
| --- |
| df\_new = df.withColumn("value", df["value"].cast(StringType()))  df\_new.show() |