kafka에서 읽어온 데이터를 spark에서는 binary로 읽어옴 우리는 이거를 string으로!!

필요한 모듈 임포트

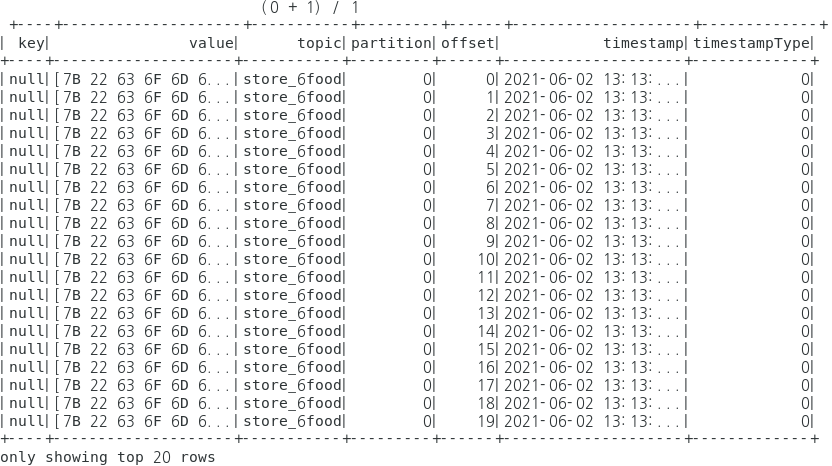
|  |
| --- |
| import pyspark  import pyspark.sql  from pyspark.sql import \*  import time  from pyspark.sql.functions import \*  from pyspark.sql.types import \*  from pyspark.sql import SparkSession |

Spark 객체 생성

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

dataFrame 생성해서 데이터 읽어옴

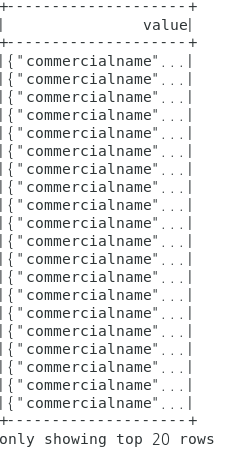
|  |
| --- |
| df = spark\  .read\  .format(“kafka”)\  .option(“kafka.bootstrap.servers”,”localhost:9092)\  .option(“subscribe”, “store\_6food”)\  .load()  df.show() |



우리가 원하는 값이 binary로 들어왔음 -> string으로 타입 변환

우리가 원하는 값은 “value”에 있음 그것만 골라서 string으로

|  |
| --- |
| dfstr = df.selectExpr(“CAST(value AS STRING)”) |



우리가 원하는 구조로 만들기 위해 struct 설정

|  |
| --- |
| df\_struct = StructType()\  .add(“year”, StringType())\  .add(“quarter”, StringType())\  .add(“divisioncode”, StringType())\  .add(“divisionname”, StringType())\  .add(“commercialcode”, StringType())\  .add(“commercialname”, StringType())\  .add(“servicecode”, StringType())\  .add(“servicename”, StringType())\  .add(“storecnt”, StringType())\  .add(“opencnt”, StringType())\  .add(“closecnt”, StringType())\ |

이제 우리가 만든 구조에다가 값을 넣을 것임

|  |
| --- |
| df\_struct = dfstr.select(from\_json(col(“value”), df\_struct).alias(“df”)  dfnew = df\_struct.select(“df.\*”)  dfnew.show() |

