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 |

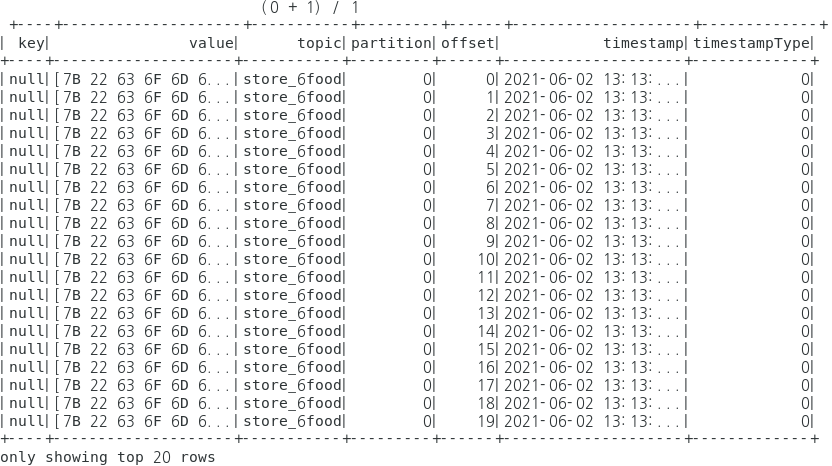
Spark 객체 생성

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

▣ store\_6food 불러오는거 (토픽이름: store\_6food / csv명 : store\_6food.csv)

dataFrame 생성해서 데이터 읽어옴

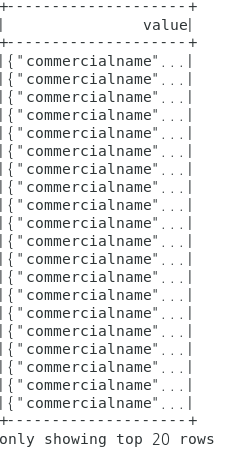
|  |
| --- |
| df1 = spark\  .read\  .format(“kafka”)\  .option(“kafka.bootstrap.servers”,”localhost:9092)\  .option(“subscribe”, “store\_6food”)\  .load()  df1.show() |



우리가 원하는 값이 binary로 들어왔음 -> string으로 타입 변환

우리가 원하는 값은 “value”에 있음 그것만 골라서 string으로

|  |
| --- |
| dfstr1 = df1.selectExpr(“CAST(value AS STRING)”) |

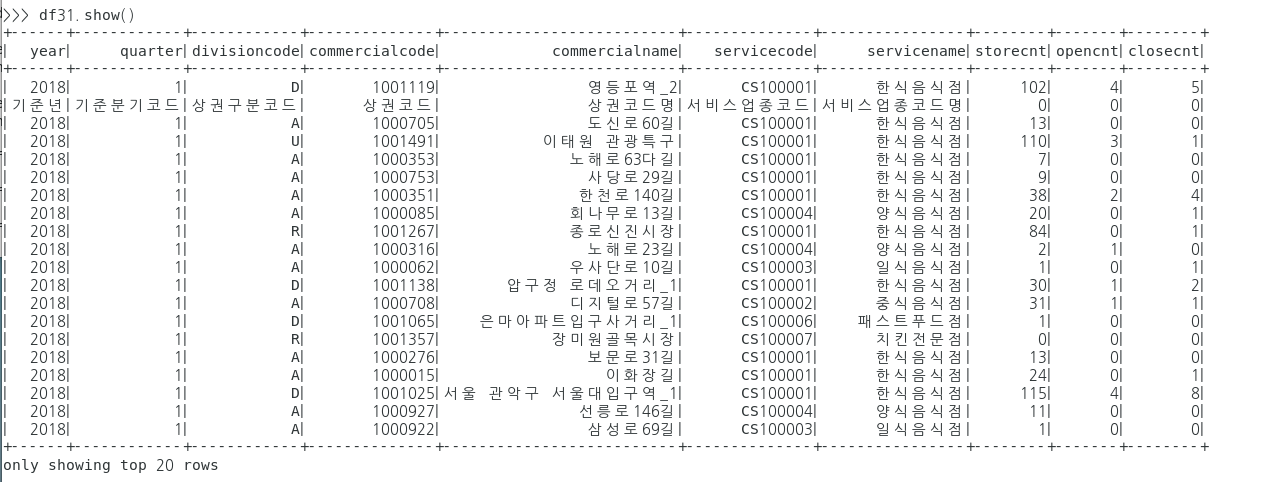


우리가 원하는 구조로 만들기 위해 struct 설정

|  |
| --- |
| df\_struct1 = 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(“openpercent”, StringType())\  .add(“closepercent”, StringType()) |

이제 우리가 만든 구조에다가 값을 넣을 것임

|  |
| --- |
| df\_struct1 = dfstr1.select(from\_json(col(“value”), df\_struct1).alias(“df1”))  dfnew1 = df\_struct1.select(“df1.\*”)  dfnew1.show() |



▣ sales\_6food 불러오는거 (토픽이름: sales\_6food / csv명 : sales\_6food.csv)

dataFrame 생성해서 데이터 읽어옴

|  |
| --- |
| df2 = spark\  .read\  .format(“kafka”)\  .option(“kafka.bootstrap.servers”,”localhost:9092)\  .option(“subscribe”, “sales\_6food”)\  .load()  df2.show() |

우리가 원하는 값이 binary로 들어왔음 -> string으로 타입 변환

우리가 원하는 값은 “value”에 있음 그것만 골라서 string으로

|  |
| --- |
| dfstr2 = df2.selectExpr(“CAST(value AS STRING)”) |

우리가 원하는 구조로 만들기 위해 struct 설정

|  |
| --- |
| df\_struct2 = 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("monthprice", StringType())\  .add("monthcnt",StringType())\  .add("weekdayprice",StringType())\  .add("weekendprice",StringType())\  .add("maleprice",StringType())\  .add("femaleprice",StringType())\  .add("10sprice",StringType())\  .add("20sprice",StringType())\  .add("30sprice",StringType())\  .add("40sprice",StringType())\  .add("50sprice",StringType())\  .add("60sprice",StringType())\  .add("weekdaycnt",StringType())\  .add("weekendcnt",StringType())\  .add("malecnt",StringType())\  .add("femalecnt",StringType())\  .add("storecnt",StringType()) |

이제 우리가 만든 구조에다가 값을 넣을 것임

|  |
| --- |
| df\_struct2 = dfstr2.select(from\_json(col(“value”), df\_struct2).alias(“df2”))  dfnew2 = df\_struct2.select(“df2.\*”)  dfnew2.show() |

▣ people1820 불러오는거 (토픽이름: people1820 / csv명 : people1820.csv)

dataFrame 생성해서 데이터 읽어옴

|  |
| --- |
| df3 = spark\  .read\  .format(“kafka”)\  .option(“kafka.bootstrap.servers”,”localhost:9092)\  .option(“subscribe”, “people1820”)\  .load()  df3.show() |

우리가 원하는 값이 binary로 들어왔음 -> string으로 타입 변환

우리가 원하는 값은 “value”에 있음 그것만 골라서 string으로

|  |
| --- |
| dfstr3 = df3.selectExpr(“CAST(value AS STRING)”) |

우리가 원하는 구조로 만들기 위해 struct 설정

|  |
| --- |
| df\_struct3 = StructType()\  .add(“year”, StringType())\  .add(“quarter”, StringType())\  .add(“divisioncode”, StringType())\  .add(“divisionname”, StringType())\  .add(“commercialcode”, StringType())\  .add(“commercialname”, StringType())\  .add("totalpeople", StringType())\  .add("male", StringType())\  .add("female", StringType())\  .add(“10s”, StringType())\  .add(“20s”, StringType())\  .add(“30s”, StringType())\  .add(“40s”, StringType())\  .add(“50s”, StringType())\  .add(“60s”, StringType())\ |

이제 우리가 만든 구조에다가 값을 넣을 것임

|  |
| --- |
| df\_struct3 = dfstr3.select(from\_json(col(“value”), df\_struct3).alias(“df3”))  dfnew3 = df\_struct3.select(“df3.\*”)  dfnew3.show() |

▣ sigungu 불러오는거 (토픽이름: sigungu / csv명 : sigungu.csv)

dataFrame 생성해서 데이터 읽어옴

|  |
| --- |
| df4 = spark\  .read\  .format(“kafka”)\  .option(“kafka.bootstrap.servers”,”localhost:9092)\  .option(“subscribe”, “sigungu”)\  .load()  df4.show() |

우리가 원하는 값이 binary로 들어왔음 -> string으로 타입 변환

우리가 원하는 값은 “value”에 있음 그것만 골라서 string으로

|  |
| --- |
| dfstr4 = df4.selectExpr(“CAST(value AS STRING)”) |

우리가 원하는 구조로 만들기 위해 struct 설정

|  |
| --- |
| df\_struct4 = StructType()\  .add(“commercialcode”, StringType())\  .add(“sigungucode”, StringType()) |

이제 우리가 만든 구조에다가 값을 넣을 것임

|  |
| --- |
| df\_struct4 = dfstr4.select(from\_json(col(“value”), df\_struct4).alias(“df4”))  dfnew4 = df\_struct4.select(“df4.\*”)  dfnew4.show() |

▣ mapping 불러오는거 (토픽이름: mapping / csv명 : mapping.csv)

dataFrame 생성해서 데이터 읽어옴

|  |
| --- |
| df5 = spark\  .read\  .format(“kafka”)\  .option(“kafka.bootstrap.servers”,”localhost:9092)\  .option(“subscribe”, “mapping”)\  .load()  df5.show() |

우리가 원하는 값이 binary로 들어왔음 -> string으로 타입 변환

우리가 원하는 값은 “value”에 있음 그것만 골라서 string으로

|  |
| --- |
| dfstr5 = df5.selectExpr(“CAST(value AS STRING)”) |

우리가 원하는 구조로 만들기 위해 struct 설정

|  |
| --- |
| df\_struct5 = StructType()\  .add(“commercialcode”, StringType())\  .add(“sigungucode”, StringType()) |

이제 우리가 만든 구조에다가 값을 넣을 것임

|  |
| --- |
| df\_struct5 = dfstr5.select(from\_json(col(“value”), df\_struct5).alias(“df5”))  dfnew5 = df\_struct5.select(“df5.\*”)  dfnew5.show() |