In [18]:
```
from pyspark.sql import SparkSession

spark = (
    SparkSession
    .builder
    .master("local[*]")
    .appName("Spark Streaming")
    .getOrCreate()
)
```

In [19]:
```
spark
```

Out[19]:

SparkSession - in-memory

SparkContext

Spark UI

Version	v3.3.0
Master	local[*]
AppName	Spark Streaming

In [33]:
```
df_raw = spark.readStream.format("socket").option("host", "localhost").optio
df_raw.printSchema()
```

```
root
 |-- value: string (nullable = true)
```

In [22]:
```
from pyspark.sql.functions import split

df_words = df_raw.withColumn("words", split("value", " "))
```

In [23]:
```
from pyspark.sql.functions import explode

df_explode = df_words.withColumn("word", explode("words"))
```

In [24]:
```
df_explode = df_explode.drop("value", "words")
```

In [25]:
```
from pyspark.sql.functions import count, lit

df_counted = df_explode.groupBy("word").count()
```

In [ ]:
```
df_counted.writeStream.format("console").outputMode("complete").start().awai
```