

```
In [ ]: from pyspark.sql import SparkSession
import pyspark.sql.functions as F
from pyspark.sql.types import *

spark = SparkSession\
    .builder\
    .appName("chapter-23-stream-in-prod")\
    .getOrCreate()

import os
SPARK_BOOK_DATA_PATH = os.environ['SPARK_BOOK_DATA_PATH']
```

```
In [ ]: static = spark.read.json(SPARK_BOOK_DATA_PATH + "/data/activity-data")
streaming = spark\
    .readStream\
    .schema(static.schema)\
    .option("maxFilesPerTrigger", 10)\
    .json("/data/activity-data")\
    .groupBy("gt")\
    .count()
```

```
In [ ]: query = streaming\
    .writeStream\
    .outputMode("complete")\
    .option("checkpointLocation", SPARK_BOOK_DATA_PATH)\
    .queryName("test_python_stream")\
    .format("memory")\
    .start()
```

```
# COMMAND -----
```