

```
In [1]: import findspark
        from pyspark.streaming import StreamingContext
        from pyspark.sql import SparkSession
        from pyspark.sql import functions as F
        import math

        findspark.init()
```

```
In [10]: try: query.stop(True, True)
        except: pass
        try: spark.stop()
        except: pass
```

```
In [11]: spark=SparkSession.builder.appName("SparkStreaming-02-DF").master('local[1]').getOrCreate()
        sc=spark.sparkContext
        # .config("spark.driver.allowMultipleContexts","true")
        spark
```

Out[11]: **SparkSession - in-memory**

**SparkContext**

[Spark UI](#)

<b>Version</b>	v2.4.8
<b>Master</b>	local[1]
<b>AppName</b>	SparkStreaming-02-DF

```
In [12]: spark.conf.set("spark.sql.shuffle.partitions", 5)
        spark.conf.get("spark.sql.shuffle.partitions")
```

Out[12]: '5'

```
In [13]: lines=spark.readStream.text("hdfs://localhost:9000/data")
```

```
In [14]: words=lines.select(F.explode(F.split(lines.value, " ")).alias("word"))
```

```
In [15]: wordc=words.groupBy("word").count()
```

```
In [16]: # query=wordc.writeStream.outputMode("complete").format("console").start()
```

In [ ]: