

SPARK STREAMING 22.2

Problem Statement

STATEFUL STREAMING IN SPARK

Input Commands:

//Initiating the spark-shell to run spark

```
[acadgild@localhost Downloads]$ spark-shell --master local[4]
```

//importing the spark packages

```
import org.apache.spark._
```

```
import org.apache.spark.streaming._
```

```
import org.apache.spark.streaming.StreamingContext._
```

//Declaring the Streaming Context Object with the help of SPARK context object sc with batch interval 5 secs

```
val ssc = new StreamingContext(sc, Seconds(5))
```

//Setting the checkpoint to current home directory

```
ssc.checkpoint(".")
```

//setting the host details for source input stream

```
val lines = ssc.socketTextStream("localhost", 9999)
```

//converting the data into words via flatmap operation

```
val words = lines.flatMap(_.split(" "))
```

//Map operation for converting each word as key and 1 as value

```
val wordDstream = words.map(word => (word, 1))
```

//Parallelize operation for entire data received in a batch interval for the string in form of List

```
val initialRDD = ssc.sparkContext.parallelize(List[(String, Int)]())
```

//Mapping function to get the count from system variable STATE and adding the same to current state of batch interval

```
val mappingFunc = (word: String, one: Option[Int], state: State[Int]) => {  
    val sum = one.getOrElse(0) + state.getOption.getOrElse(0)  
    val output = (word, sum)  
    state.update(sum)  
    output  
}
```

//Applying the mapWithState function to current stream for stateful streaming

```
val stateDstream =  
wordDstream.mapWithState(StateSpec.function(mappingFunc).initialState(initialRDD))
```

//Printing the data stream with word count

```
stateDstream.print()
```

//Starting the SPARK STREAMING session

```
ssc.start()
```

//Keep awaiting for the streamed data until streaming session is terminated with CTRL+C

```
ssc.awaitTermination()
```

Streaming input Source Screenshots:

```
[acadgild@localhost ~]$ nc -lk 9999
Hi This is Acadgild
Hi This is Acadgild
This is a beautiful day
This is a beautiful day
This is a beautiful day
This is a beautiful day
█
```

Ouput Screenshot:

```
(beautiful,2)
(This,4)
(is,4)

17/11/29 21:12:44 WARN BlockManager: Block input-0-1511970164000 replicated to only 0 peer(s) instead of 1 peers
-----
Time: 1511970165000 ms
-----

(day,3)
(a,3)
(beautiful,3)
(This,5)
(is,5)

-----
Time: 1511970170000 ms
-----

-----
Time: 1511970175000 ms
-----

17/11/29 21:12:57 WARN BlockManager: Block input-0-1511970177600 replicated to only 0 peer(s) instead of 1 peers
-----
Time: 1511970180000 ms
-----

(day,4)
(a,4)
(beautiful,4)
(This,6)
(is,6)

-----
```