

Dstreams

Window

Transformations

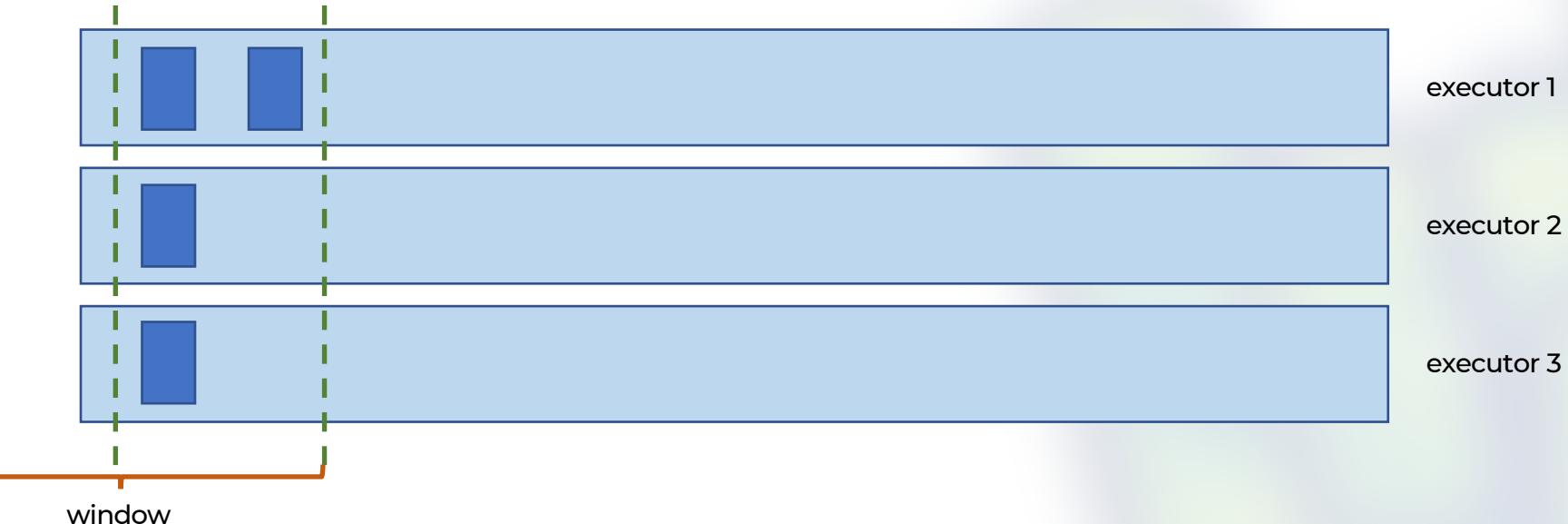


Objective

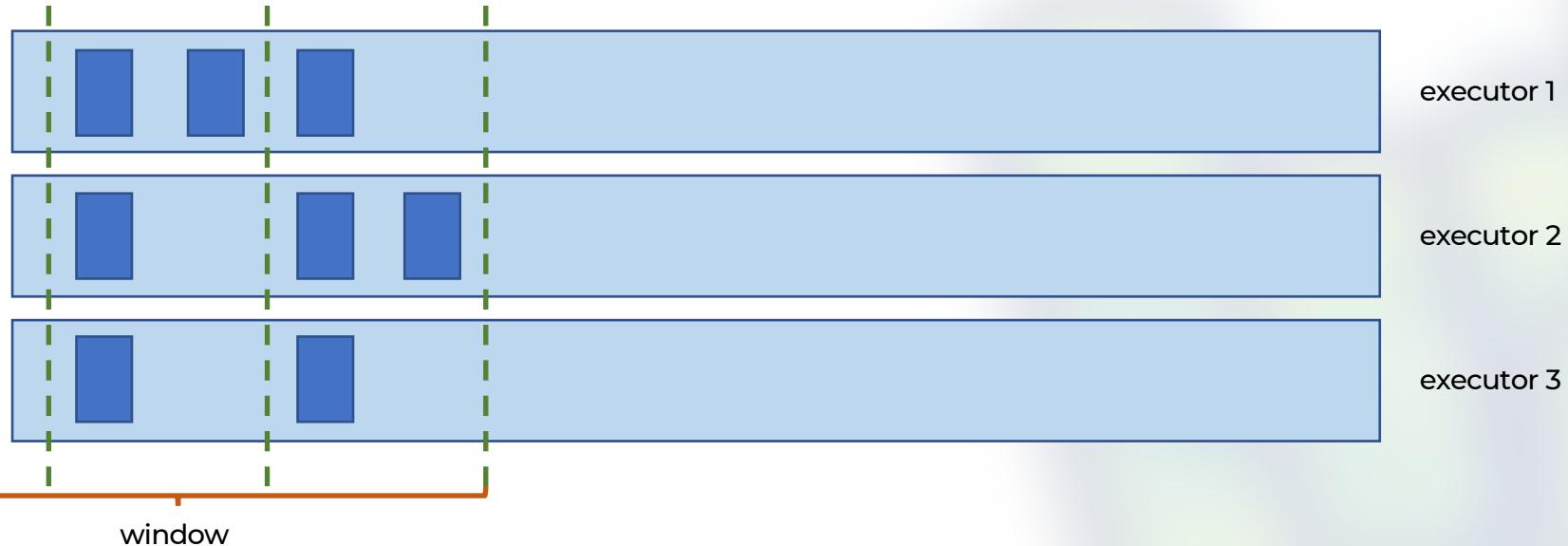
Practice advanced time-based transformations on DStreams



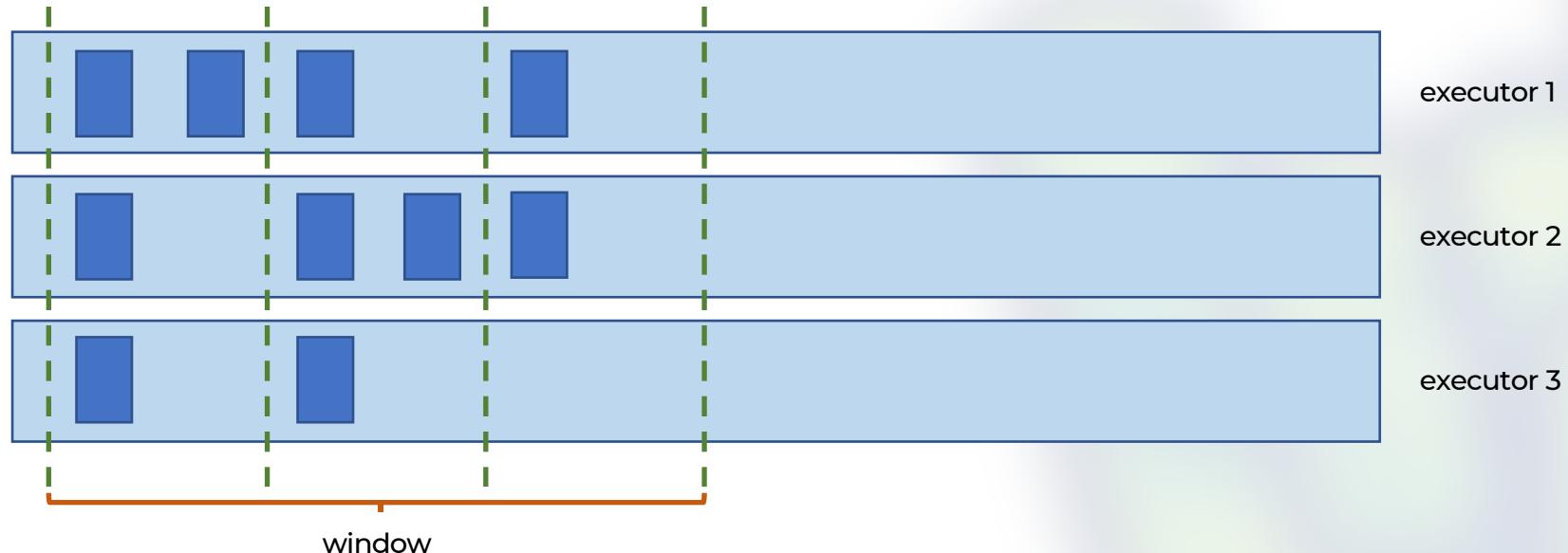
Sliding Window



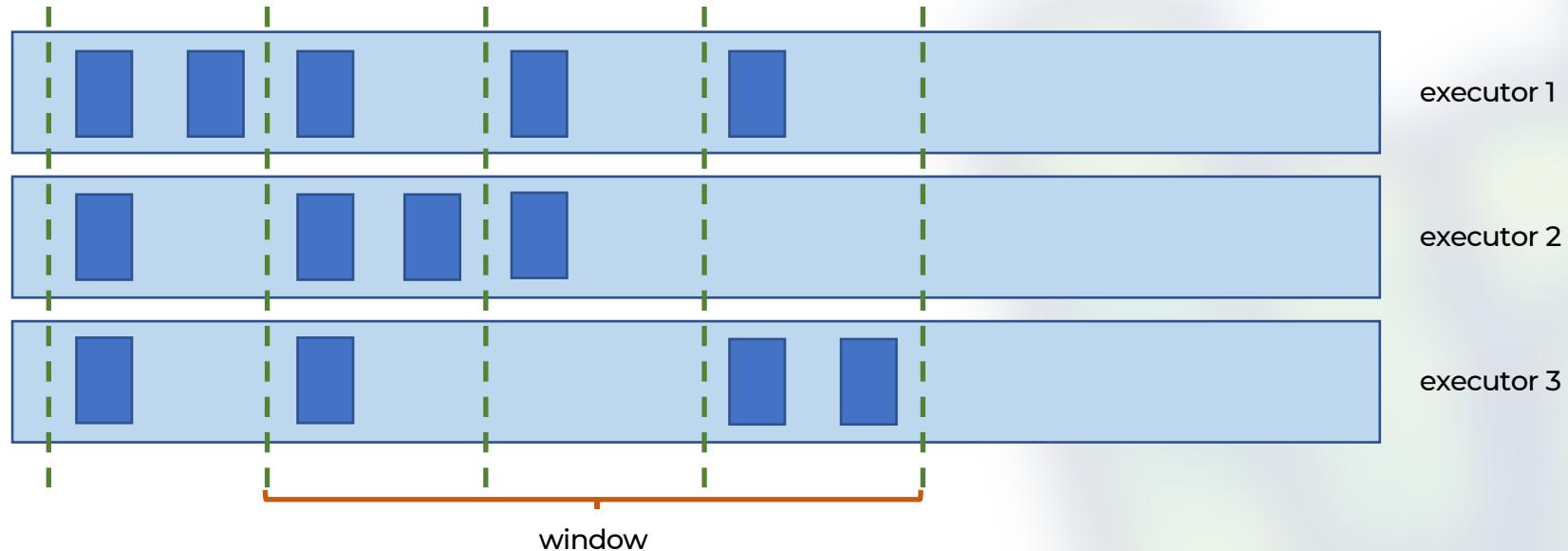
Sliding Window



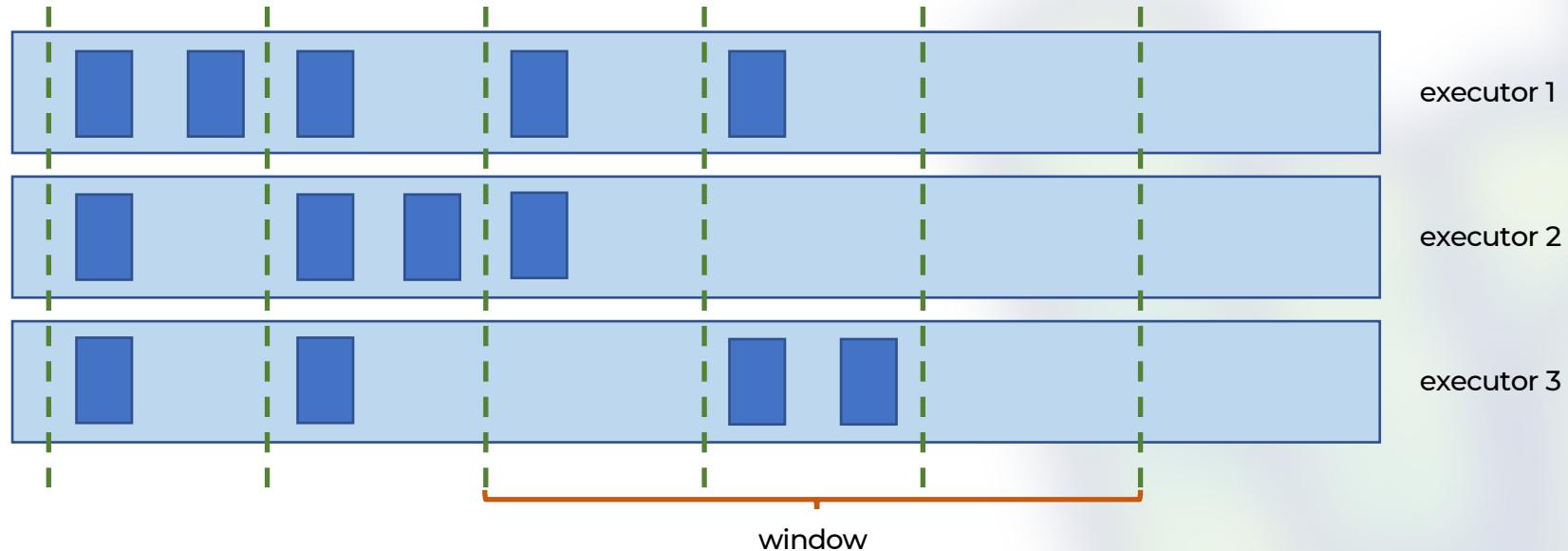
Sliding Window



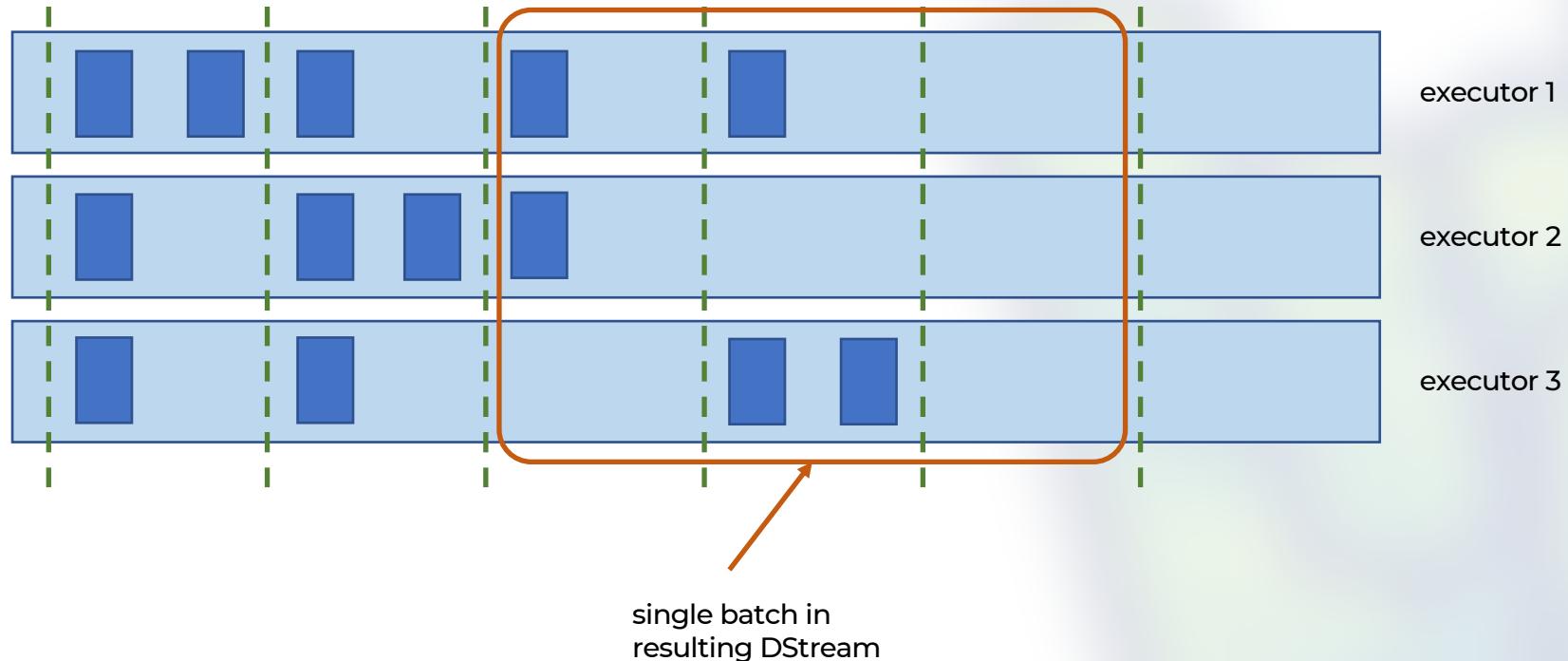
Sliding Window



Sliding Window



Sliding Window



Takeaways

Window transformations

```
dataStream.window(Seconds(5))
```

keeps all the values emitted between now and 5s ago
interval updated with every batch

```
dataStream.window(Seconds(10), Seconds(5))
```

interval updated every 5s

```
dataStream.countByWindow(Seconds(10), Seconds(5))
```

```
dataStream.reduceByWindow(_ + _, Seconds(10), Seconds(5))
```

each RDD has a single element obtained reducing past 10 seconds with this function

```
dataStream.reduceByKeyAndWindow(_ + _, _ - _ Seconds(10), Seconds(5))
```

only on DStream[(K, V)]

reduction function

function to remove old elements from the sliding window

Spark rocks

