

1. [Apache Spark Professional Training with Hands On Lab Sessions](#)
2. [Oreilly Databricks Apache Spark Developer Certification Simulator](#)

SPARK STREAMING PART-2 STATELESS AND STATEFULL TRANSFORMATIONS

By www.HadoopExam.com

**Note: These instructions should be used with the HadoopExam Apache Spark: Professional Trainings.
Where it is executed and you can do hands on with trainer.**

1. Hadoop Training
2. Spark Training
3. HBase Training
4. MapR Developer
5. MapR HBase
6. CCA500 Certification
7. Spark Certification
8. EMC Data Science

Hadoop Specialization offer == 50% + 35% off

Hadoop Expert

~~52000INR~~ == **16900INR Only**
~~\$1150~~ == **\$373 Only**
Hadoop Specialization offer

* @ End of the Offer Prices will increase by 25%

Limited Time Offer (Less Than 5 Days Remain)

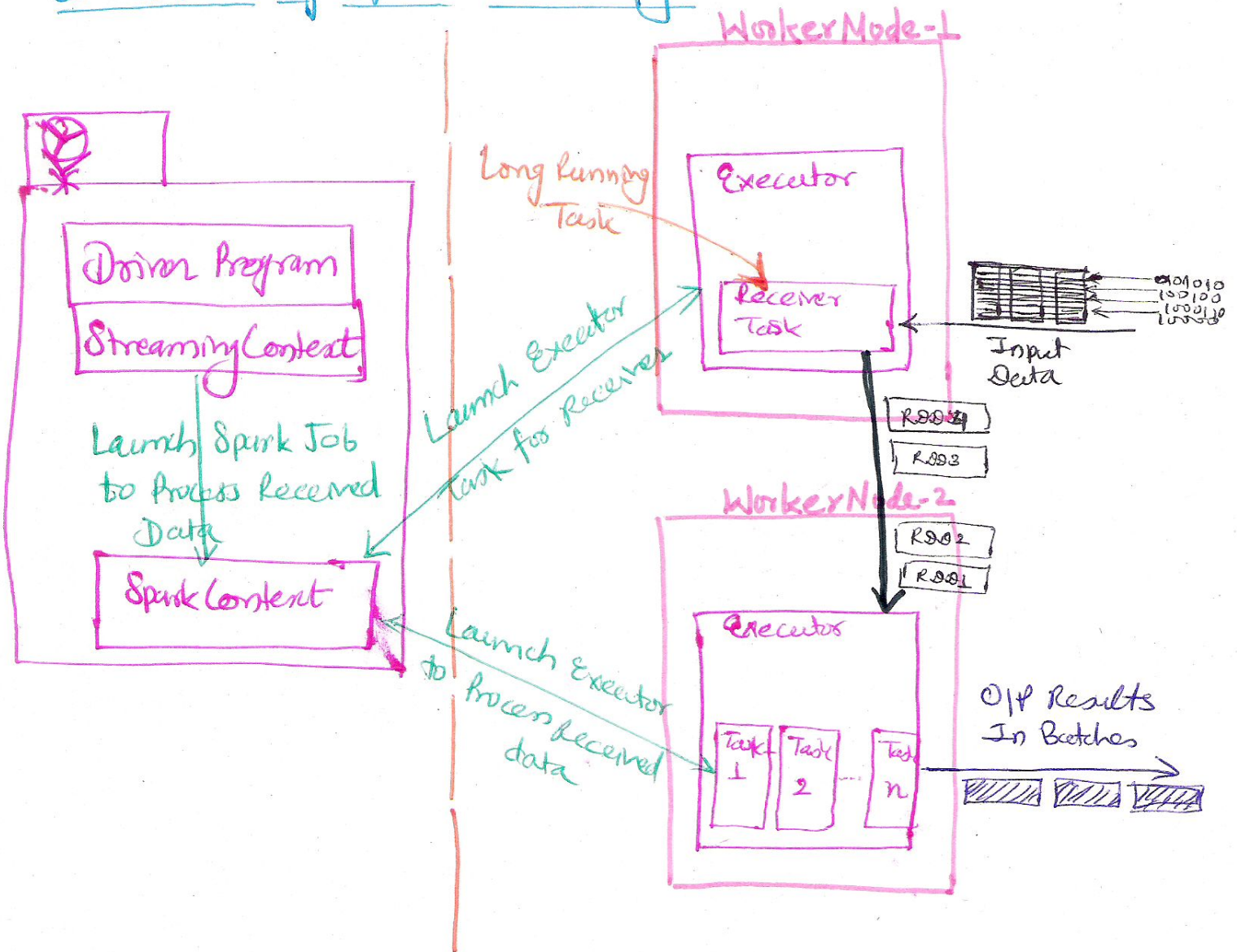


[Cloudera CCA175 \(Hadoop and Spark Developer Hands-on Certification available with total 75 solved problem scenarios. Click for More Detail\)](#)

1. Real/Near real time data processing
2. Streaming Sources and Sinks
3. DStream (Discretized Stream)
4. Dstream Concepts
5. Stock Visualization Example (How Streaming Helpful)

Spark Streaming Part - 2

Execution of Spark Streaming

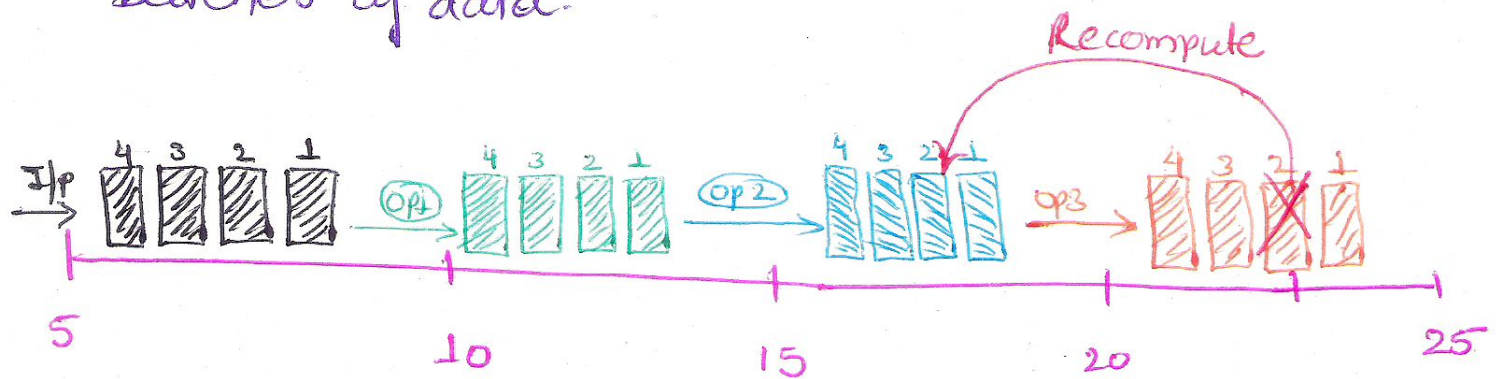


① For each input source, Spark Streaming Launches Receivers.

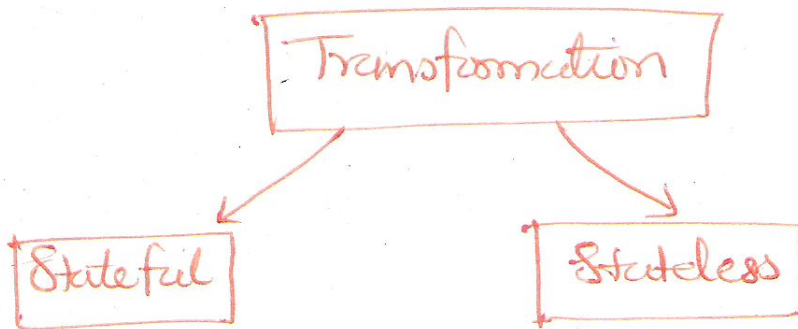
② Receiver: - (Longer Task): - Running within application executors that collect data from the i/p source and save it as RDD.

- ③ This created RDD now replicated to another node for fault tolerance.
- ④ This Data (RDD) is stored in memory of the executors in some way as a cached RDD's.
- ⑤ The StreamingContext in the Driver Program then periodically runs Spark jobs to process this data and combine it with RDD's from previous time steps.
- ⑥ Lineage: - Same as core Spark, it can also derive or recompute lost RDD.
- ⑦ Time Consuming: - using lineage, recomputation could take a long time for data that has been built up since beginning of the program.
- ⑧ Checkpointing: - Thus Spark streaming also includes a mechanism called checkpointing that saves state periodically to a reliable filesystem e.g. HDFS, S3.

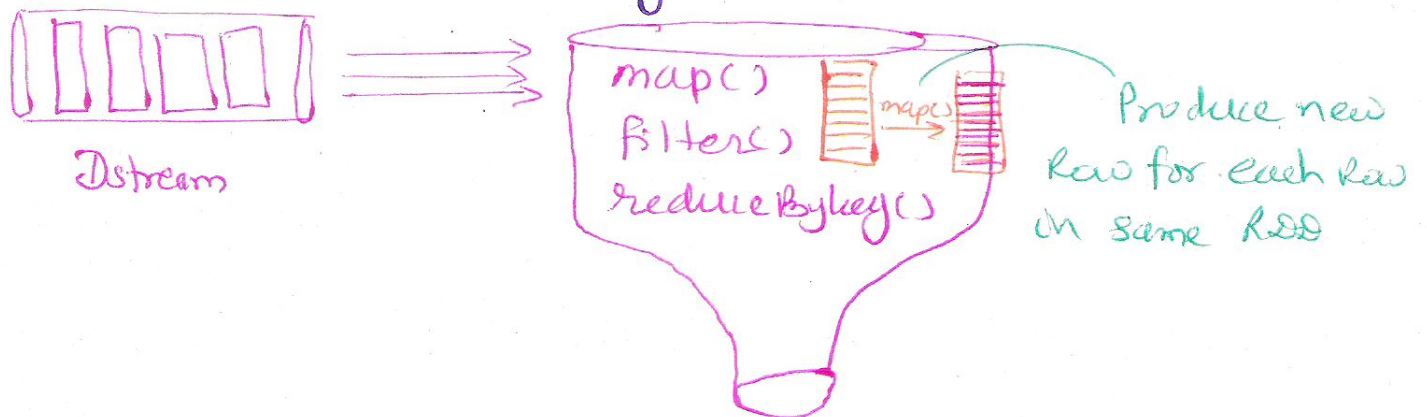
⇒ Typically checkpointing is done every 5-10 batches of data.



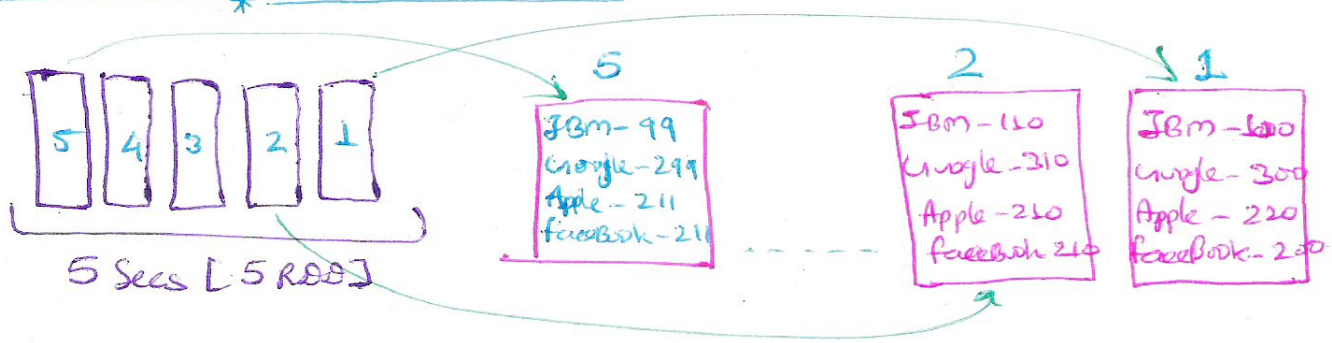
Transformation



① Stateless: — Processing of each batch does not depend on the data of its previous batches.



Stateful Transformation



- ① Each RDD contains stock name and its current price (key, value)
- ② Now calculate Average price for every 5 seconds.
- ③ Hence all the RDDs needs to be processed together
(They depend on each other)
- ④ They include transformations based on sliding windows.

⇒ More on Stateless Transformation :-

API — map(), flatMap(), filter()

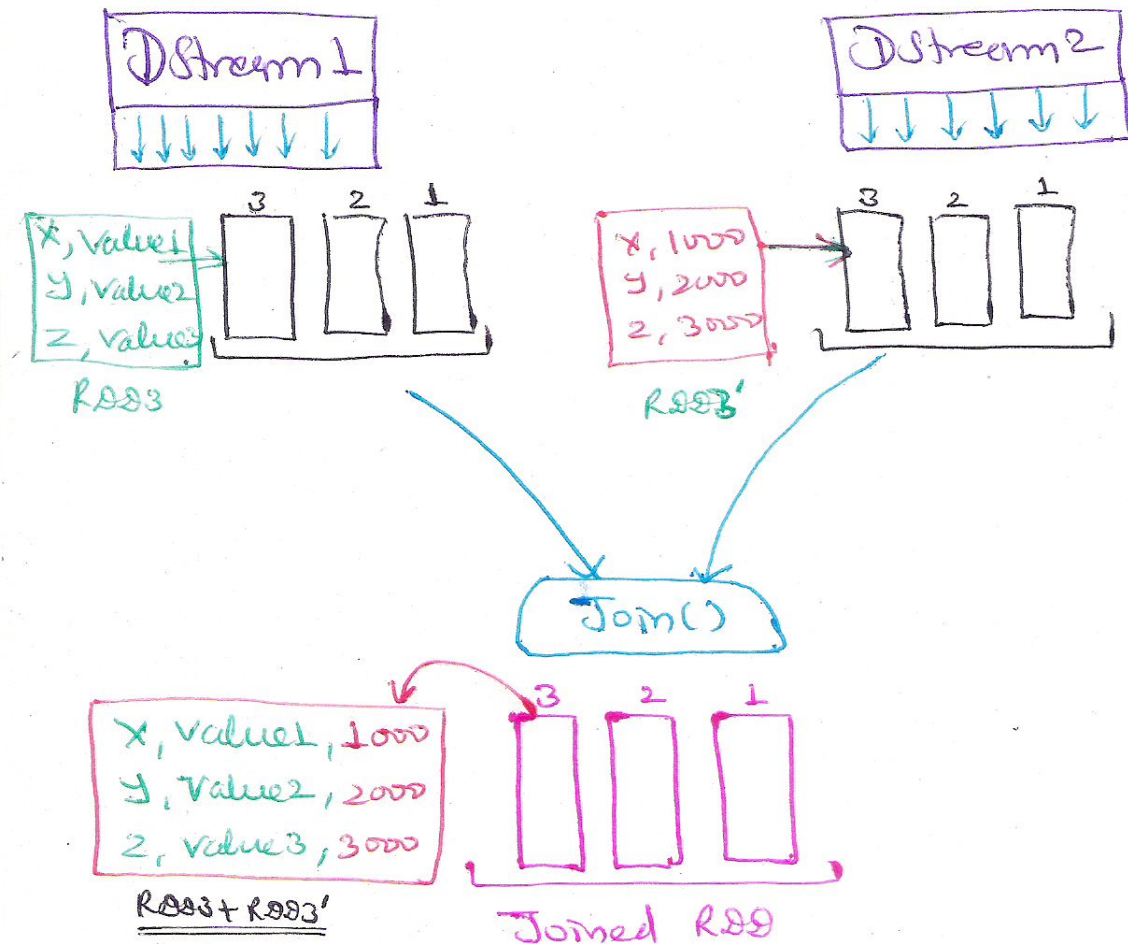
repartition() → change the number of partitions of the streams.

⇒ ReduceByKey() — to

⇒ GroupByKey

⇒ Each stateless transformation applies separately to each RDD.

⇒ Stateless Transformation & Combining Multiple DStream



⇒ In stateless transformations data can be combined from multiple DStream as shown above, within each timestep. for instance

`join(), cogroup(), leftOuterJoin()`

⇒ Similarly we can also merge the contents of two different DStream using union operator.

Applying Custom transformation

Using transform() operator

- ⇒ transform() allows you to operate directly on the RDDs inside them.
- ⇒ You have, your transformation (Custom logic) written which you have regularly using for Spark batch processing.
- ⇒ This same transformations you want to use in Dstream. How would you do that

```
Val customRDD = dataStream.transform(
```

```
    rdd ⇒ CustomTransformation(rdd)
```

```
)
```


CCA Spark and Hadoop Developer Certification

HadoopExam Learning Resource provides the following material for the Advanced Technologies.
Please visit www.HadoopExam.com for more detail this is just a few products from portfolio.

Price start for training with Just \$79/3500INR



Apache Spark
Professional
Training with
HandsOn Session

+ Certification
Material



Hadoop Professional
Training with
HandsOn Session

+ Certification
Material



HBase Professional
Training with
HandsOn Session

+ Certification
Material



Certification
Material



Certification
Material



Certification
Material



Certification
Material



Certification
Material



Microsoft Azure

Certification
Material



Certification
Material