

## Apache Spark - Interview Questions

(/interview\_questions/bigdata\_spark)

Apache Spark - Core (/interview\_questions/bigdata\_spark/bigdata\_spark\_core)

Apache Spark - Graphx (/interview\_questions/bigdata\_spark/bigdata\_spark\_graphx)

Apache Spark MLlib (/interview\_questions/bigdata\_spark/bigdata\_spark\_mllib)

Apache Spark SQL (/interview\_questions/bigdata\_spark/bigdata\_spark\_sql)

Apache Spark Streaming (/interview\_questions/bigdata\_spark/bigdata\_spark\_streaming)

### MASTER Apache Spark

Top ranked courses to help you master Apache Spark skills.

#### Master the Coding Interview: Data Structures + Algorithms



(<https://click.linksynergy.com/deeplink?>

id=5mEPVSSvmFI&mid=39197&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fmaster-the-coding-interview-data-structures-algorithms%2F)

Offered By - **Andrei Neagoie**

Platform - **Udemy**

Rating - **\*\*\*\*\***

Students Enrolled - **100,000 +**

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies) ☐ Agree

Ad

**Codeworks™ Berlin**

CC

**Algorithms**

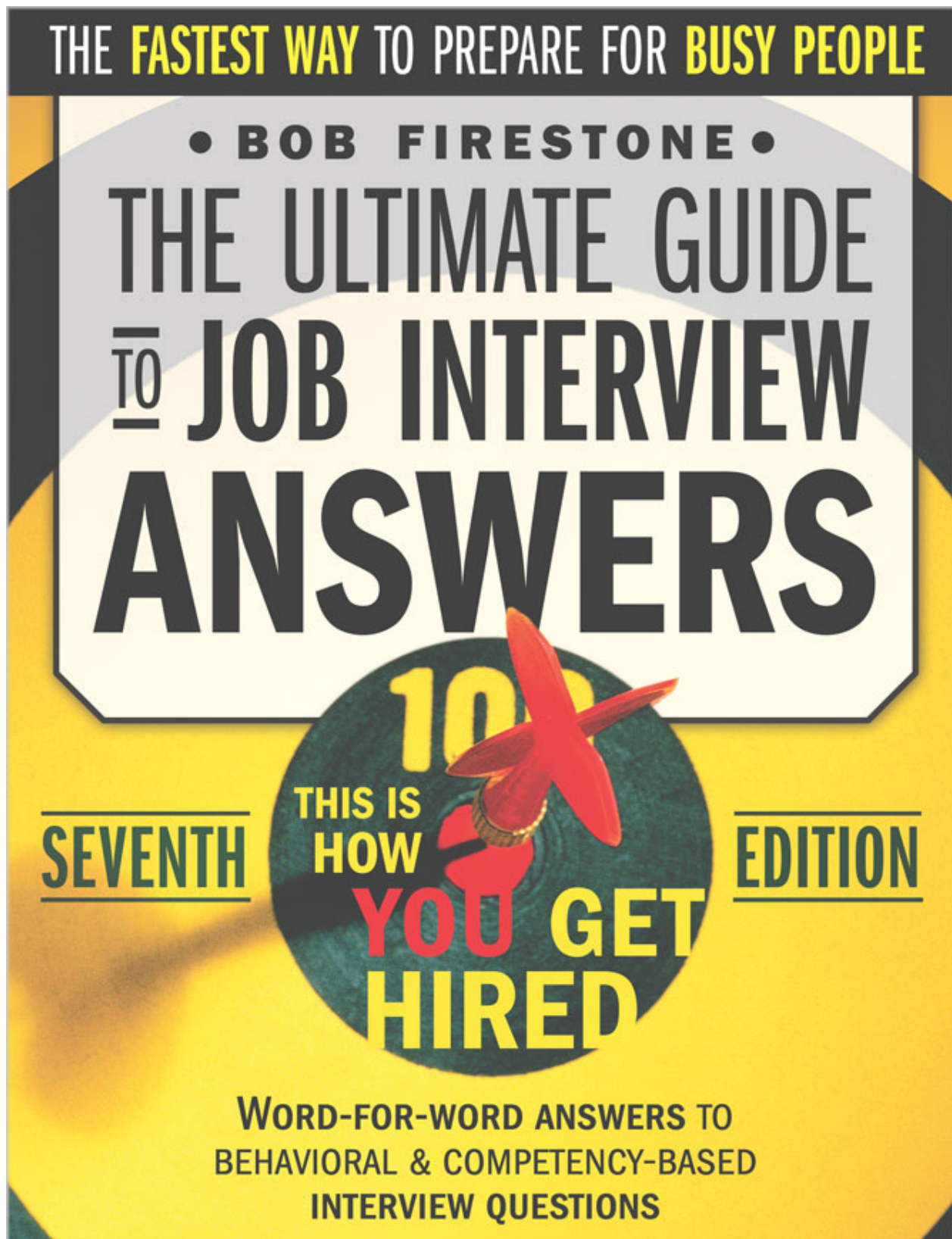
(https://click.linksynergy.com/deeplink?

id=5mEPVSSvmFI&amp;mid=40328&amp;murl=https%3A%2F%2Fwww.coursera.org%2Flearn%2Falgorithms-part1%3F)

Offered By - **Princeton University**Platform - **Coursera**Rating - **\*\*\*\*\***Students Enrolled - **850,000 +****RECOMMENDED RESOURCES****Behaviorial Interview**

Top resource to prepare for behaviorial and situational interview questions.

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies) ☐ Agree



(<https://bd859dq7jkbvbk93unsk19nei.hop.clickbank.net/>)

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies) ☐ Agree



## Quantum ML

### Quantum Processor

IISc Professors & QpiAI Experts. People Are Learning with IISc and QpiAI

qpiai-explorer.tech

OPEN

# Apache Spark Streaming - Interview Questions

## What is Apache Spark Streaming?

FAQ

Spark Streaming is a library provided in Apache Spark for processing live data streams that is scalable, has high-throughput and is fault-tolerant. Spark Streaming can ingest data from multiple sources such as Kafka, Flume, Kinesis or TCP sockets; and process this data using complex algorithms provided in the Spark API including algorithms provided in the Spark MLlib and GraphX libraries. Processed data can be pushed to live dashboards, file systems and databases.

## Describe how Spark Streaming processes data?

FAQ

Apache Spark Streaming component receives live data streams from input sources such as Kafka, Flume, Kinesis etc. and divides them into batches. The Spark engine processes these input batches and produces the final stream of results in batches.

## What are DStreams?

FAQ

DStreams, or discretized streams, are high-level abstractions provided in Spark Streaming that represents a continuous stream of data. DStreams can be either created from input sources such as Kafka, Flume or Kinesis; or by applying high-level operations on existing DStreams.

Internally, a DStream is represented by a continuous series of RDDs. Each RDD in a DStream contains data from a certain interval.

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (cookies) page.

## What is a StreamingContext object?

FAQ

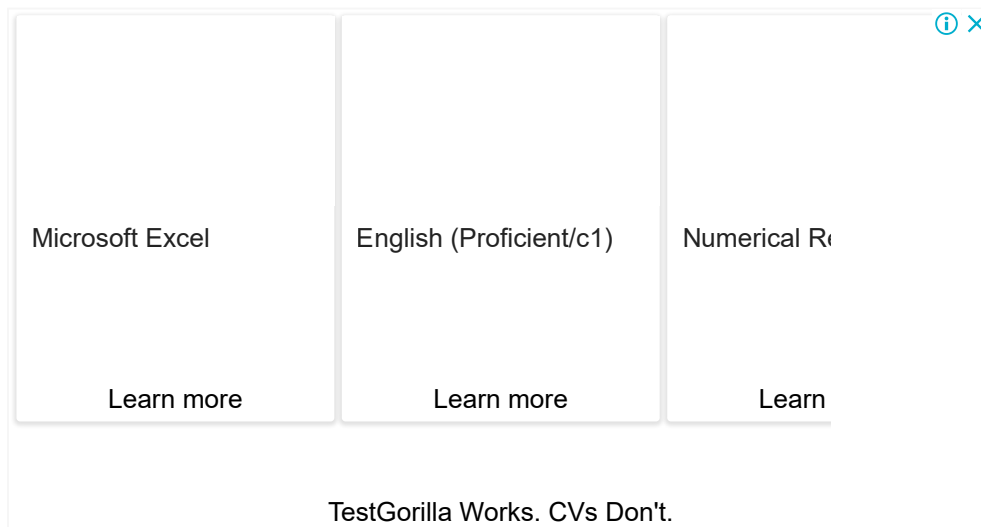
StreamingContext object represents the entry point to a Spark streaming application and contains the batch interval for the Spark streaming program and the SparkConf object. The SparkConf object and the batch interval are set when creating a new instance of the StreamingContext object.

## What are the two different types of built-in streaming sources provided by Spark Streaming?

FAQ

Spark Streaming provides two different categories of built-in streaming sources.

**Basic Sources** - These sources are directly available in the StreamingContext API. Example: file systems and socket connections.



**Advanced Sources** - These are sources such as Kafka, Flume etc. that are provided in the Spark Streaming library through extra utility classes.

## Big Data Interview Guide has over 150+ interview questions and answers. Get the guide for \$44.95 only.

BUY EBOOK (<https://gumroad.com/l/MChNX>)

## What are some of the common transformations on DStreams supported by Spark Streaming ?

FAQ

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies) ☐ Agree

**map(func)** - map() transformation returns a new distributed dataset from a source dataset formed by passing each element of the source through a function func.

**filter()** - filter() transformation returns a new distributed dataset from a source dataset formed by selecting the elements of the source on which func returns true.

Microsoft Excel

Learn more

English (Proficient/c1)

Learn more

Numerical R

Learn

TestGorilla Works. CVs Don't.

**flatMap()** - flatMap() transformation is similar to map() function. In flatMap() each input item can be mapped to 0 or more output items.

Microsoft Excel

Learn more

English (Proficient/c1)

Learn more

Numerical R

Learn

TestGorilla Works. CVs Don't.

**union()** - union() transformation returns a new dataset that contains the union of the elements in the source dataset and the dataset that is passed as argument to the function.

**intersection()** - intersection() transformation returns a new distributed dataset that contains the intersection of elements in the source dataset and the dataset that is passed as argument to the function.

**distinct()** - distinct() transformation returns a new distributed dataset that contains the distinct elements of the source dataset.

**AED 12,000,000 could be yours**

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies).

1 ticket, 5 opportunities to win: 12 Million, 1 Million and AED 500,000 could be yours

Agree

**groupByKey()** - groupByKey(func) transformation called on a dataset of (K, V) pairs and returns a dataset of (K, Iterable) pairs.

**reduceByKey(func)** - reduceByKey(func) transformation is called on a dataset of (K, V) pairs, and returns a dataset of (K, V) pairs where the values for each key are aggregated using the given reduce function func.

**aggregateByKey(func)** - aggregateByKey() transformation is called on a dataset of (K, V) pairs, returns a dataset of (K, U) pairs where the values for each key are aggregated using the given combine functions and a neutral zero value.

**sortByKey()** - sortByKey() transformation is called on a dataset of (K, V) pairs which returns a dataset of (K, V) pairs sorted by keys in ascending or descending order, as specified in the boolean ascending argument.

## What are the output operations that can be performed on DStreams?

### FAQ

Output operations on DStreams pushes the DStream's data to external systems like a database or a file system. Following are the key operations that can be performed on DStreams.

**saveAsTextFiles()** - Saves the DStream's data as text file.

**saveAsObjectFiles()** - Saves the DStreams data as serialized Java objects.

**saveAsHadoopFiles()** - Saves the DStream's data as Hadoop files.

**foreachRDD()** - A generic output operator that applies a function, func, to each RDD generated from the DStream.

## Important Keywords to Remember

- ✓ Apache Spark Streaming
- ✓ DStreams
- ✓ StreamingContext
- ✓ Transformations
- ✓ Output operations - saveAsTextFiles(), saveAsObjectFiles(), saveAsHadoopFiles(), foreachRDD()

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies) ☐ Agree

Microsoft Excel

Learn more

English  
(Proficient/c1)

Learn more

16

l

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies)     Agree





www.interviewgrid.com

# Big Data

## INTERVIEW GUIDE

**A Comprehensive Interview Preparation Companion**

✓ **150+ Interview Questions And Answers**

**Get the Edge**

- Hadoop - MapReduce | HDFS | YARN
- Flume | Kafka | Hive | Hue | Oozie
- Spark SQL | Spark Streaming
- Spark MLlib | Spark GraphX

TEKEX  
PRESS

*Kiran K. Gopisetty*

**\$49.95**

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies) <https://gumroad.com/l/MChNX>



SSL Secure Payment

[www.interviewgrid.com](http://www.interviewgrid.com)

# Big Data

## INTERVIEW GUIDE

**A Comprehensive Interview Preparation Companion**

✓ **150+ Interview Questions And Answers**

**Get the Edge**

- Hadoop - MapReduce | HDFS | YARN
- Flume | Kafka | Hive | Hue | Oozie
- Spark SQL | Spark Streaming
- Spark MLlib | Spark GraphX


TEKEX  
PRESS

*Kiran K. Gopisetty*

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies) ☐ Agree

# \$49.95

BUY EBOOK (<https://gumroad.com/l/MChNX>)

 SSL Secure Payment



Email Address

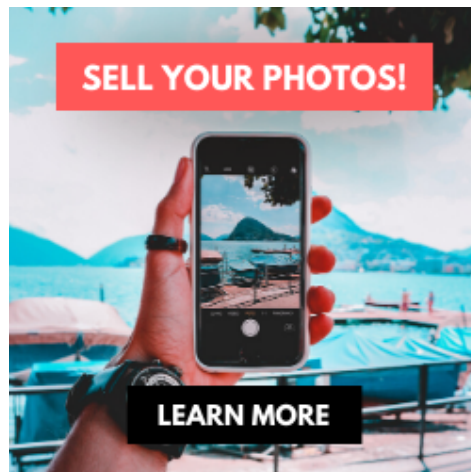
DOWNLOAD FREE CHAPTER

You're also subscribing to the [interviewgrid.com](https://interviewgrid.com) email newsletter for tips, updates & promotions. Unsubscribe any time.

Ad

## EARN SIDE INCOME

Earn income using your skills and hobbies that you are passionate about.



(<https://fbc87qgjl8lfz4dodl3p8tw8c.hop.clickbank.net/>)

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [privacy policy \(/privacy\)](#) & [cookie policy \(/cookies\)](#) ☐ Agree



 WritingPaychecks

CLICK HERE

(<https://b519dhfjft8ubt3kzaulctcv16.hop.clickbank.net/>)

Ad

### FINANCIAL INDEPENDENCE

Transition out of your full-time job by creating businesses that you are passionate about.

START HERE (<https://cece1suiijfqdnf75hiz6taqaf.hop.clickbank.net/?cbpage=sp>)

 (<https://www.facebook.com/InterviewGrid/>)  (<https://twitter.com/interviewgrid>)  
 (<https://www.pinterest.com/interviewgrid/>)  
 ([https://www.youtube.com/channel/UCF5BuT2\\_HnloqnGKB7rnJTw](https://www.youtube.com/channel/UCF5BuT2_HnloqnGKB7rnJTw))

About Us (<http://www.interviewgrid.com/about>) | Privacy Policy (<http://www.interviewgrid.com/privacy>) |  
Terms of Use (<http://www.interviewgrid.com/terms>) | Contact Us (<http://www.interviewgrid.com/Contact>)

© Interview Grid 2021

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our privacy policy (/privacy) & cookie policy (/cookies) ☐ Agree