

Top Apache Spark Interview Questions You Should Prepare In 2022

Last updated on Mar 29,2022 280.3K Views

[Share](#)**Sandeep Dayananda**

Sandeep Dayananda is a Research Analyst at Edureka. He has expertise in...

Hadoop Certification Training

-  CloudLab Access
-  Real-life Projects
-  Lifetime Access
-  24x7 Tech Support

 **edureka!**

[EXPLORE COURSE](#)

2018 has been the year of Big Data – the year when big data and analytics made tremendous progress through innovative technologies, data-driven decision making and outcome-centric analytics. Worldwide revenues for big data and business analytics (BDA) will grow from \$130.1 billion in 2016 to more than \$203 billion in 2021 (source IDC). Prepare with these top **Apache Spark Interview Questions** to get an edge in the burgeoning Big Data market where global and local enterprises, big or small, are looking for a quality **Big Data and Hadoop experts**.

Want to Upskill yourself to get ahead in Career? Check out the [Top Trending Technologies Article](#).

As a big data professional, it is essential to know the right buzzwords, learn the right technologies and prepare the right answers to commonly asked Spark interview questions. With questions and answers around *Spark Core, Spark Streaming, Spark SQL, GraphX, MLlib* among others, this blog is your gateway to your next Spark job.

Apache Spark Interview Questions And Answers

1. Compare Hadoop and Spark.

We will compare Hadoop MapReduce and Spark based on the following aspects:

Feature Criteria	Apache Spark	Hadoop
Speed	100 times faster than Hadoop	Decent speed
Processing	Real-time & Batch processing	Batch processing only
Difficulty	Easy because of high level modules	Tough to learn
Recovery	Allows recovery of partitions	Fault-tolerant
Interactivity	Has interactive modes	No interactive mode except Pig & Hive

Apache Spark vs. Hadoop**Table: Apache Spark versus Hadoop**

Let us understand the same using an interesting analogy.

"Single cook cooking an entree is regular computing. Hadoop is multiple cooks cooking an entree into pieces and letting each cook her piece."

Each cook has a separate stove and a food shelf. The first cook cooks the meat, the second cook cooks the sauce. This phase is called "Map". At the end the main cook assembles the complete entree. This is called "Reduce". For Hadoop, the cooks are not allowed to keep things on the stove between operations. Each time you make a particular operation, the cook puts results on the shelf. This slows things down.

For Spark, the cooks are allowed to keep things on the stove between operations. This speeds things up. Finally, for Hadoop the recipes are written in a language which is illogical and hard to understand. For Spark, the

Kladko, Galactic Exchange.io



Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

Figure: Real Time Processing In Spark

Figure: Data Parallelism In Spark

Spark is one of the most successful projects in the Apache Software Foundation. Spark has clearly evolved as the market leader for Big Data processing. Many organizations run Spark on clusters with thousands of nodes. Today, Spark is being adopted by major players like Amazon, eBay, and Yahoo!

3. Explain the key features of Apache Spark.

The following are the key features of Apache Spark:

1. **Polyglot**
2. **Speed**
3. **Multiple Format Support**
4. **Lazy Evaluation**
5. **Real Time Computation**
6. **Hadoop Integration**
7. **Machine Learning**

Let us look at these features in detail:

1. **Polyglot:** Spark provides high-level APIs in Java, Scala, Python and R. Spark code can be written in any of these four languages. It provides a shell in Scala and Python. The Scala shell can be accessed through `./bin/spark-shell` and Python shell through `./bin/pyspark` from the installed directory.
2. **Speed:** Spark runs upto 100 times faster than Hadoop MapReduce for large-scale data processing. Spark is able to achieve this speed through controlled partitioning. It manages data using partitions that help parallelize distributed data processing with minimal network traffic.
3. **Multiple Formats:** Spark supports multiple data sources such as Parquet, JSON, Hive and Cassandra. The Data Sources API provides a pluggable mechanism for accessing structured data though Spark SQL. Data sources can be more than just simple pipes that convert data and pull it into Spark.
4. **Lazy Evaluation:** Apache Spark delays its evaluation till it is absolutely necessary. This is one of the key factors contributing to its speed. For transformations, Spark adds them to a DAG of computation and only when the driver requests some data, does this DAG actually gets executed.
5. **Real Time Computation:** Spark's computation is real-time and has less latency because of its in-memory computation. Spark is designed for massive scalability and the Spark team has documented users of the system running production clusters with thousands of nodes and supports several computational models.
6. **Hadoop Integration:** Apache Spark provides smooth compatibility with Hadoop. This is a great boon for all the Big Data engineers who started their careers with Hadoop. Spark is a potential replacement for the MapReduce functions of Hadoop.

e! Top Apache Spark Interview Questions You Should Prepare In 2022
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

Copy Link

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

Copy Link

 FREE WEBINAR

Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

Copy Link

 FREE WEBINAR

Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

Copy Link

 FREE WEBINAR

Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

e! Top Apache Spark Interview Questions You Should Prepare In 2022
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

Copy Link

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

e! Top Apache Spark Interview Questions You Should Prepare In 2022
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

e! Top Apache Spark Interview Questions You Should Prepare In 2022
edureka.co



Whatsapp

Linkedin

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

Copy Link

 FREE WEBINAR

Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

e! Top Apache Spark Interview Questions You Should Prepare In 2022
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

e! Top Apache Spark Interview Questions You Should Prepare In 2022
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

e! Top Apache Spark Interview Questions You Should Prepare In 2022
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

Copy Link

 FREE WEBINAR

Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

e! Top Apache Spark Interview Questions You Should Prepare In 2022
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

e! Top Apache Spark Interview Questions You Should Prepare In 2022
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR
Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

!

Reduce Side Joins With
MapReduce

[Watch Now](#)

stion



Hadoop Architecture – Hadoop
Tutorial on HDFS Architecture

[Watch Now](#)

Link

5 Scenarios: When To Use &
When Not to Use Hadoop

[Watch Now](#)

What is Apache Stor
about?

[Watch Now](#)

[Copy Link](#)



 FREE WEBINAR

Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^



Sign up with Google



Signup with Facebook

Already have an account? [Sign in.](#)

[Read Article](#)

[Read Article](#)

[Read Article](#)

[Read Article](#)

»

e!

stions

x

[Big Data Hadoop Certification Training Course](#)

• 160k Enrolled Learners
Weekend
Live Class

[Copy Link](#)

5 (67250)

[Apache Spark and Scala Certification Training ...](#)

LinkedIn • 29k Enrolled Learners
Weekend
Live Class

[Reviews](#)

5 (11500)

[PySpark Certification Training Course](#)

• 8k Enrolled Learners
Weekend
Live Class

[Reviews](#)

5 (3100)

[Splunk Certification Power User and ...](#)

• 9k Enrolled Learners
Weekend
Live Class



FREE WEBINAR

Apache Pig Tutorial Explained in 60 ...

Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR

Apache Pig Tutorial Explained in 60 ...



Become a Certified Professional →

Subscribe to our Newsletter, and get personalized recommendations. ^

 Sign up with Google

 Signup with Facebook

Already have an account? [Sign in.](#)

 **Top Apache Spark Interview Questions You Should Prepare In 2022**
edureka.co



Whatsapp

LinkedIn

Twitter

Facebook

Reddit

[Copy Link](#)

 FREE WEBINAR

Apache Pig Tutorial Explained in 60 ...

