| 🏠 (https://www.edureka.co/blog/all) | Blogs (https://www.edureka.co/blog/) | Videos (https://www.edureka.co/blog/videos/) | Interview Questions (https://www.edureka.co/blog/interview-questions/) |

Top Hadoop Interview Questions On Apache PIG For 2018

🔖 Recommended by 48 users

Shubham Sinha (https://www.edureka.co/blog/author/shubham-sinha/) | Jul 26,2018

Apache Pig Interview Questions

Looking out for Apache Pig Interview Questions that are frequently asked by employers? Here is the fifth blog of Hadoop Interview Questions series, which covers Apache PIG interview questions. I hope you must not have missed the earlier blogs of our **Hadoop Interview Question (https://www.edureka.co/blog/interview-questions/top-50-hadoop-interview-questions-2016/)series (https://www.edureka.co/blog/interview-questions/top-50-hadoop-interview-questions-2016/)**.

After going through the Pig interview questions, you will get an in-depth knowledge of questions that are frequently asked by employers in Hadoop interviews.

In case you have attended Pig interviews previously, we encourage you to add your questions in the comments tab. We will be happy to answer them, and spread the word to the community of fellow job seekers.

Important points to remember about Apache Pig:

♦ Apache Pig is a platform, used to analyze large data sets representing them as data flows. It is designed to provide an abstraction over MapReduce, reducing the complexities of writing a MapReduce task using Java programming. We can perform data manipulation operations very easily in Hadoop using Apache Pig.

♦ Apache Pig has two main components – the **Pig Latin** language and the **Pig Run-time Environment**, in which Pig Latin programs are executed.

♦Apache Pig follows **ETL** (Extract Transform Load) process. It can handle inconsistent schema (in case of unstructured data).

♦ Apache Pig automatically optimizes the tasks before execution, i.e. automatic optimization. Apache Pig handles all kinds of data.

♦ Pig allows programmers to write custom functions which is unavailable in Pig. User Defined Functions (UDF) can be written in different language like Java, Python, Ruby, etc. and embed them in Pig script.

♦ Pig Latin provides various built-in operators like join, sort, filter, etc. to read, write, and process large data sets.

♣ *Tip: Before going through this Apache Pig interview questions, I would suggest you to go through* **Apache Pig Tutorial (https://www.edureka.co/blog/pig-tutorial/)** *to revise your Pig concepts.*

Now moving on, let us look at the Apache Pig interview questions.

Hadoop Interview Questions and Answers | Edureka

1. Highlight the key differences between MapReduce and Apache Pig.

♣ *Tip: In this question, you should explain what were the problems with MapReduce which led to the development of Apache Pig by Yahoo.*

MapReduce vs Apache Pig

MapReduce	Apache Pig
1. It is a low-level data processing paradigm	1. It is a high-level data flow platform
2. Complex Java implementations	2. No complex Java implementations
3. Do not provide nested data types	3. Provides nested data types like tuples, bags, and maps
4. Performing data operations is a humongous task	4. Provides many built-in operators to support data operations

2. What are the use cases of Apache Pig?

Apache Pig is used for analyzing and performing tasks involving ad-hoc processing. Apache Pig is used for:

- Research on large raw data sets like data processing for search platforms. For example, Yahoo uses Apache Pig to analyse data gathered from Yahoo search engines and Yahoo News Feeds.

- Processing huge data sets like Web logs, streaming online data, etc.
- In customer behavior prediction models like e-commerce websites.

3. What is the difference between logical and physical plans?

♣ *Tip: Approach this question by explaining when does the logical and physical plans are created.*

Pig undergoes some steps when a Pig Latin Script is converted into MapReduce jobs by the compiler. Logical and Physical plans are created during the execution of a pig script.

After performing the basic parsing and semantic checking, the parser produces a logical plan and no data processing takes place during the creation of a logical plan. The logical plan describes the logical operators that have to be executed by Pig during execution. For each line in the Pig script, syntax check is performed for operators and a logical plan is created. If an error is encountered, an exception is thrown and the program execution ends.

A logical plan contains a collection of operators in the script, but does not contain the edges between the operators.

After the logical plan is generated, the script execution moves to the physical plan where there is a description about the physical operators, Apache Pig will use, to execute the Pig script. A physical plan is like a series of MapReduce jobs, but the physical plan does not have any reference on how it will be executed in MapReduce.

4. How Pig programming gets converted into MapReduce jobs?

Pig is a high-level platform that makes many Hadoop data analysis issues easier to execute. A program written in Pig Latin is a data flow language, which need an execution engine to execute the query. So, when a program is written in Pig Latin, Pig compiler converts the program into MapReduce jobs.

5. What are the components of Pig Execution Environment?

The components of Apache Pig Execution Environment are:

- **Pig Scripts**: Pig scripts are submitted to the Apache Pig execution environment which can be written in Pig Latin using built-in operators and UDFs can be embedded in it.
- **Parser**: The Parser does the type checking and checks the syntax of the script. The parser outputs a DAG (directed acyclic graph). DAG represents the Pig Latin statements and logical operators.
- **Optimizer:** The Optimizer performs the optimization activities like split, merge, transform, reorder operators, etc. The optimizer provides the automatic optimization feature to Apache Pig. The optimizer basically aims to reduce the amount of data in the pipeline.
- **Compiler**: The Apache Pig compiler converts the optimized code into MapReduce jobs automatically.
- **Execution Engine**: Finally, the MapReduce jobs are submitted to the execution engine. Then, the MapReduce jobs are executed and the required result is produced.

6. What are the different ways of executing Pig script?

There are three ways to execute the Pig script:

- **Grunt Shell**: This is Pig's interactive shell provided to execute all Pig Scripts.
- **Script File**: Write all the Pig commands in a script file and execute the Pig script file. This is executed by the Pig Server.
- **Embedded Script**: If some functions are unavailable in built-in operators, we can programmatically create User Defined Functions (UDF) to bring that functionality using other languages like Java, Python, Ruby, etc. and embed it in the Pig Latin Script file. Then, execute that script file.

7. What are the data types of Pig Latin?

Pig Latin can handle both atomic data types like int, float, long, double etc. and complex data types like tuple, bag and map.

Atomic or scalar data types are the basic data types which are used in all the languages like string, int, float, long, double, char[], byte[]. These are also called the primitive data types.

The complex data types supported by Pig Latin are:

- **Tuple**: Tuple is an ordered set of fields which may contain different data types for each field.
- **Bag**: A bag is a collection of a set of tuples and these tuples are a subset of rows or entire rows of a table.
- **Map**: A map is key-value pairs used to represent data elements. The key must be a chararray [] and should be unique like column name, so it can be indexed and value associated with it can be accessed on the basis of the keys. The value can be of any data type.

♣ *Tip: Complex Data Types of Pig Latin are very important to understand, so you can go through **Apache Pig Tutorial (https://www.edureka.co/blog/pig-tutorial/)** blog and understand them in-depth.*

8. What is a bag in Pig Latin?

A bag is one of the data models present in Pig. It is an unordered collection of tuples with possible duplicates. Bags are used to store collections of tuples while grouping. The size of bag is the size of the local disk, this means that the size of the bag is limited. When the bag is full, then Pig will spill this bag into local disk and keep only some parts of the bag in memory. There is no necessity that the complete bag should fit into memory. We represent bags with "{}".

♣ *Tip:You can also explain the two types of bag in Pig Latin i.e. outer bag and inner bag, which may impress your employers.*

9. What do you understand by an inner bag and outer bag in Pig?

Outer bag or relation is nothing but a bag of tuples. Here relations are similar as relations in relational databases. For example:

   {(Linkin Park, California), (Metallica, Los Angeles), (Mega Death, Los Angeles)}

An inner bag contains a bag inside a tuple. For Example:

   (Los Angeles, *{(Metallica, Los Angeles), (Mega Death, Los Angeles)}*)

   (California, *{(Linkin Park, California)}*)

10. How Apache Pig deals with the schema and schema-less data?

♣ *Tip: Apache Pig deals with both schema and schema-less data. Thus, this is an important question to focus on.*

The Apache Pig handles both, schema as well as schema-less data.

- If the schema only includes the field name, the data type of field is considered as a byte array.
- If you assign a name to the field you can access the field by both, the field name and the positional notation, whereas if field name is missing we can only access it by the positional notation i.e. $ followed by the index number.
- If you perform any operation which is a combination of relations (like JOIN, COGROUP, etc.) and if any of the relation is missing schema, the resulting relation will have null schema.
- If the schema is null, Pig will consider it as a byte array and the real data type of field will be determined dynamically.

11. How do users interact with the shell in Apache Pig?

Using Grunt i.e. Apache Pig's interactive shell, users can interact with HDFS or the local file system.

To start Grunt, users should use *pig –x local* command . This command will prompt Grunt shell. To exit from grunt shell, press CTRL+D or just type exit.

12. What is UDF?

If some functions are unavailable in built-in operators, we can programmatically create User Defined Functions (UDF) to bring that functionality using other languages like Java, Python, Ruby, etc. and embed it in the Pig Latin Script file.

♣ *Tip: To understand how to create and work with UDF, go through this blog –* **creating UDF in Apache Pig (https://www.edureka.co/blog/creating-udf-in-apache-pig/).**

♣ *Tip: Important points about UDF to focus on:*

- LoadFunc abstract class has three main methods for loading data and for most use cases it would suffice to extend it.
- LoadPush has methods to push operations from Pig runtime into loader implementations.
- setUdfContextSignature() method will be called by Pig both in the front end and back end to pass a unique signature to the Loader.
- The load/store UDFs control how data goes into Pig and comes out of Pig.
- The meaning of getNext() is called by Pig runtime to get the next tuple in the data.
- The loader should use setLocation() method to communicate the load information to the underlying InputFormat.
- prepareToRead method enables the RecordReader associated with the InputFormat provided by the LoadFunc is passed to the LoadFunc. The RecordReader can then be used by the implementation in getNext() to return a tuple representing a record of data back to pig.
- pushProjection() method tells LoadFunc which fields are required in the Pig script. Pig will use the column index requiredField.index to communicate with the LoadFunc about the fields required by the Pig script.
- LoadCaster has methods to convert byte arrays to specific types.
- A loader implementation should implement LoadCaster() if casts (implicit or explicit) from DataByteArray fields to other types need to be supported. LoadCaster has methods to convert byte arrays to specific types.

13. List the diagnostic operators in Pig.

Pig supports a number of diagnostic operators that you can use to debug Pig scripts.

- **DUMP:** Displays the contents of a relation to the screen.
- **DESCRIBE:** Return the schema of a relation.
- **EXPLAIN:** Display the logical, physical, and MapReduce execution plans.
- **ILLUSTRATE:** Gives the step-by-step execution of a sequence of statements.

♣ *Tip: Go through this blog on* **diagnostic operators** *(https://www.edureka.co/blog/operators-in-apache-pig-diagnostic-operators/), to understand them and see their implementations.*

14. Does 'ILLUSTRATE' run a MapReduce job?

No, illustrate will not pull any MapReduce, it will pull the internal data. On the console, illustrate will not do any job. It just shows the output of each stage and not the final output.

ILLUSTRATE operator is used to review how data is transformed through a sequence of Pig Latin statements. ILLUSTRATE command is your best friend when it comes to debugging a script. This command alone might be a good reason for choosing Pig over something else.

Syntax: *illustrate relation_name;*

15. What does illustrate do in Apache Pig?

Executing Pig scripts on large data sets, usually takes a long time. To tackle this, developers run Pig scripts on sample data, but there is possibility that the sample data selected, might not execute your Pig script properly. For instance, if the script has a join operator there should be at least a few records in the sample data that have the same key, otherwise the join operation will not return any results.

To tackle these kind of issues, illustrate is used. Illustrate takes a sample of the data and whenever it comes across operators like join or filter that remove data, it ensures that only some records pass through and some do not, by making modifications to the records such that they meet the condition. Illustrate just shows the output of each stage but does not run any MapReduce task.

Learn Apache Pig from Industry Experts

(https://www.edureka.co/big-data-and-

hadoop)

16. List the relational operators in Pig.

All Pig Latin statements operate on relations (and operators are called relational operators). Different relational operators in Pig Latin are:

- **COGROUP**: Joins two or more tables and then perform GROUP operation on the joined table result.
- **CROSS**: CROSS operator is used to compute the cross product (Cartesian product) of two or more relations.
- **DISTINCT**: Removes duplicate tuples in a relation.
- **FILTER**: Select a set of tuples from a relation based on a condition.
- **FOREACH**: Iterate the tuples of a relation, generating a data transformation.
- **GROUP**: Group the data in one or more relations.
- **JOIN**: Join two or more relations (inner or outer join).
- **LIMIT**: Limit the number of output tuples.
- **LOAD**: Load data from the file system.
- **ORDER**: Sort a relation based on one or more fields.
- **SPLIT**: Partition a relation into two or more relations.
- **STORE**: Store data in the file system.
- **UNION**: Merge the content of two relations. To perform a UNION operation on two relations, their columns and domains must be identical.

♣ Tip: Go through this blog on **relational operators** (https://www.edureka.co/blog/operators-in-apache-pig/), to understand them and see their implementations.

17. Is the keyword 'DEFINE' like a function name?

Yes, the keyword 'DEFINE' is like a function name.

DEFINE statement is used to assign a name (alias) to a UDF function or to a streaming command.

- The function has a long package name that you don't want to include in a script, especially if you call the function several times in that script. The constructor for the function takes string parameters. If you need to use different constructor parameters for different calls to the function you will need to create multiple defines – one for each parameter set.
- The streaming command specification is complex. The streaming command specification requires additional parameters (input, output, and so on). So, assigning an alias makes it easier to access.

18. What is the function of co-group in Pig?

COGROUP takes members of different relations, binds them by similar fields, and creates a bag that contains a single instance of both relations where those relations have common fields. Co-group operation joins the data set by grouping one particular data set only.

It groups the elements by their common field and then returns a set of records containing two separate bags. The first bag consists of the first data set record with the common data set and the second bag consists of the second data set records with the common data set.

19. Can we say co-group is a group of more than 1 data set?

Co-group is a group of data sets. More than one data set, co-group will group all the data sets and join them based on the common field. Hence, we can say that co-group is a group of more than one data set and join of that data set as well.

20. The difference between GROUP and COGROUP operators in Pig?

Group and Cogroup operators are identical. For readability, GROUP is used in statements involving one relation and COGROUP is used in statements involving two or more relations. Group operator collects all records with the same key. Cogroup is a combination of group and join, it is a generalization of a group instead of collecting records of one input depends on a key, it collects records of n inputs based on a key. At a time, we can Cogroup up to 127 relations.

21. You have a file personal_data.txt in the HDFS directory with 100 records. You want to see only the first 5 records from the employee.txt file. How will you do this?

For getting only 5 records from 100 records we use limit operator.

First load the data in Pig:

personal_data = LOAD "/personal_data.txt" USING PigStorage(',') as (parameter1, Parameter2, …);

Then Limit the data to 5 records:

limit_data = LIMIT personal_data 5;

22. What is a MapFile?

MapFile is a class which serves file-based map from keys to values.

A map is a directory containing two files, the data file, containing all keys and values in the map, and a smaller index file, containing a fraction of the keys. The fraction is determined by *MapFile.Writer.getIndexInterval()*.

The index file is read entirely into memory. Thus, key implementations should try to keep themselves small. Map files are created by adding entries in-order.

23. What is BloomMapFile used for?

The BloomMapFile is a class that extends MapFile. So its functionality is similar to MapFile. BloomMapFile uses dynamic Bloom filters to provide quick membership test for the keys. It is used in Hbase table format.

24. What are the different execution modes available in Pig?

The execution modes in Apache Pig are:

- **MapReduce Mode**: This is the default mode, which requires access to a Hadoop cluster and HDFS installation. Since, this is a default mode, it is not necessary to specify -x flag (you can execute pig OR pig -x mapreduce). The input and output in this mode are present on HDFS.
- **Local Mode:** With access to a single machine, all files are installed and run using a local host and file system. Here the local mode is specified using '-x flag' (pig -x

local). The input and output in this mode are present on local file system.

25. Is Pig script case sensitive?

♣ *Tip: Explain the both aspects of Apache Pig i.e. case-sensitive as well as case-insensitive aspect.*

Pig script is both case sensitive and case insensitive.

User defined functions, the field name, and relations are case sensitive i.e. EMPLOYEE is not same as employee or M=LOAD 'data' is not same as M=LOAD 'Data'.

Whereas Pig script keywords are case insensitive i.e. LOAD is same as load.

It is difficult to say whether Apache Pig is case sensitive or case insensitive. For instance, user defined functions, relations and field names in Pig are case sensitive. On the other hand, keywords in Apache Pig are case insensitive.

26. What does Flatten do in Pig?

Sometimes there is data in a tuple or a bag and if we want to remove the level of nesting from that data, then Flatten modifier in Pig can be used. Flatten un-nests bags and tuples. For tuples, the Flatten operator will substitute the fields of a tuple in place of a tuple, whereas un-nesting bags is a little complex because it requires creating new tuples.

27. What is Pig Statistics? What are all stats classes in the Java API package available?

Pig Statistics is a framework for collecting and storing script-level statistics for Pig Latin. Characteristics of Pig Latin scripts and the resulting MapReduce jobs are collected while the script is executed. These statistics are then available for Pig users and tools using Pig (such as Oozie) to retrieve after the job is completed.

The stats classes are in the package *org.apache.pig.tools.pigstats*:

- PigStats
- JobStats
- OutputStats
- InputStats.

28. What are the limitations of the Pig?

Limitations of the Apache Pig are:

1. As the Pig platform is designed for ETL-type use cases, it's not a better choice for real-time scenarios.
2. Apache Pig is not a good choice for pinpointing a single record in huge data sets.
3. Apache Pig is built on top of MapReduce, which is batch processing oriented.

Check out our Hadoop Course

(https://www.edureka.co/big-data-and-

hadoop)

Conclusion:

I hope these Apache Pig Interview Questions were helpful for you. I would suggest you to go through the whole series, to get in-depth knowledge on Hadoop Interview Questions. Learn Hadoop from industry experts while working with real-life use cases.

Kindly, refer to the links given below and enjoy the reading:

- *Top 50 Hadoop Interview Questions* (https://www.edureka.co/blog/interview-questions/top-50-hadoop-interview-questions-2016/)
- *Hadoop Cluster Interview Questions* (https://www.edureka.co/blog/interview-questions/hadoop-interview-questions-hadoop-cluster/)
- *Hadoop HDFS Interview Questions (https://www.edureka.co/blog/interview-questions/hadoop-interview-questions-hdfs-2/)*
- *Hadoop MapReduce Interview Questions* (https://www.edureka.co/blog/interview-questions/hadoop-interview-questions-mapreduce/)
- *Hive Interview Questions* (https://www.edureka.co/blog/interview-questions/hive-interview-questions/)
- *HBase Interview Questions (https://www.edureka.co/blog/interview-questions/hbase-interview-questions/)*

Got a question for us? Mention them in the comments section and we will get back to you.

About Shubham Sinha (26 Posts (https://www.edureka.co/blog/author/shubham-sinha/))

Shubham Sinha is a Big Data and Hadoop expert working as a Research Analyst at Edureka. He is keen to work with Big Data related technologies such as Hadoop, Spark, Flink and Storm and web development technologies including Angular, Node.js & PHP.

Share on

(https://www.facebook.com/sharer.php?
u=https://www.edureka.co/blog/interview-
questions/hadoop-
interview-
questions-
pig/)

(https://twitter.com/intent/tweet?
url=https://www.edureka.co/blog/interview-
questions/hadoop-
interview-
questions-
pig/)

(https://plus.google.com/share?
url=https://www.edureka.co/blog/interview-
questions/hadoop-
interview-
questions-
pig/)

(https://www.linkedin.com/shareArticle?
mini=true&url=https://www.edureka.co/blog/interview-
questions/hadoop-
interview-
questions-
pig/)

‹ PREVIOUS

NEXT ›

Got your brain cells running?
Stay tuned to latest technology updates
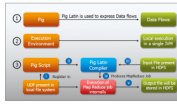
Enter your Email Address

SUBSCRIBE

Related Posts

Pig Tutorial: Apache Pig Architecture & Twitter Case Study
👁 23.7K

Pig Programming: Apache Pig Script with UDF in HDFS Mode
👁 14.2K

(https://www.edureka.co/blog/pig-tutorial/)

(https://www.edureka.co/blog/pig-programming-apache-pig-script-with-udf-in-hdfs-mode/)

Browse Categories

Big Data NoSQL (https://www.edureka.co/blog/category/big-data-nosql/)   Blockchain (https://www.edureka.co/blog/category/blockchain/)

Business Intelligence (https://www.edureka.co/blog/category/business-intelligence/)   Cloud Computing (https://www.edureka.co/blog/category/cloud-computing/)

Cyber Security (https://www.edureka.co/blog/category/cyber-security/)   Deep Learning (https://www.edureka.co/blog/category/deep-learning/)

Finance (https://www.edureka.co/blog/category/finance/)   Frameworks (https://www.edureka.co/blog/category/frameworks/)

Marketing (https://www.edureka.co/blog/category/marketing/)   Mobile Development (https://www.edureka.co/blog/category/mobile-development/)

Operations (https://www.edureka.co/blog/category/operations/)   Programming (https://www.edureka.co/blog/category/programming/)

Project Management (https://www.edureka.co/blog/category/project-management/)   Robotic Process Automation (https://www.edureka.co/blog/category/robotic-process-automation/)

Success Story (https://www.edureka.co/blog/category/success-story/)   Systems & Architecture (https://www.edureka.co/blog/category/systems-architecture/)

Systems Engineering (https://www.edureka.co/blog/category/systems-engineering/)   Testing (https://www.edureka.co/blog/category/testing/)

Comments   **7 Comments**

7 Comments   **https://www.edureka.co/blog/**   ● **Rajiv Chaudhuri** ▾

♡ **Recommend**   🐦 **Tweet**   f **Share**   Sort by Best ▾

Join the discussion…

Uday KArnam • 2 years ago
can we visualized data in graphs, chart, plot digrams by using apcahe pig
△ | ▽ • Reply • Share ›

> **EdurekaSupport** Mod ➔ Uday KArnam • 2 years ago
> Hey Uday, thanks for checking out our blog.
> No. The data in PIG can be transformed into some graphical representation (for example charts) only by using some visualisation tools. You have to use some visualisation tools for representation of your data.There are data visualisation tools available in the market, for example some of them are Tableau, Worlform Alpha,Excel,Many eyes,Talend.
> Hope this helps. Cheers!
> △ | ▽ • Reply • Share ›
>
> > **Uday KArnam** ➔ EdurekaSupport • 2 years ago
> > thanks for Edureka.....giving a solutions for my query
> > △ | ▽ • Reply • Share ›
> >
> > > **EdurekaSupport** Mod ➔ Uday KArnam • 2 years ago

You're welcome, Uday. :) Do follow our blog to stay posted. Cheers!

∧ | ∨ • Reply • Share ›

Uday KArnam ➜ EdurekaSupport • 2 years ago

sure:-

∧ | ∨ • Reply • Share ›

Pavan Hadoop • 3 years ago

There is a mistake in answers to one of the questions -
What co-group does in Pig?
Co-group joins the data set by grouping one particular data set only.

The answer has to be one or more than one dataset

∧ | ∨ • Reply • Share ›

Abhishek Srivastava ➜ Pavan Hadoop • 2 years ago

You might have missed to read the whole para of Co-group, It has mentioned in the end that "co-group is a group of more than one data set and join of that data set as well."

∧ | ∨ • Reply • Share ›

ALSO ON HTTPS://WWW.EDUREKA.CO/BLOG/

How Blockchain Technology Works? Step by Step Guide for Beginners

2 comments • 6 months ago

ICO Development — Hey, great write up this is exactly true. Thanks for sharing blog, I think most people unaware about blockchain and its …

Top 55 Blockchain Interview Questions You Must Prepare In 2018

4 comments • 7 months ago

Maximilian Fischer — A blockchain (database) only contains a record of all changes that were ever made to it. Value is ascribed to the fact of actually being …

How To Install Kubernetes Cluster On Ubuntu 16.04

9 comments • 4 months ago

Pramod Lawate — I got below error while executing below command# kubeadm init --apiserver-advertise-address=198.168.56.100 …

Top 10 Reasons Why You Should Learn Selenium

1 comment • 6 months ago

Partho — Since there is no Global Certification in Selenium what is the benchmark then.....

✉ **Subscribe** ⓓ **Add Disqus to your siteAdd DisqusAdd** 🔒 **Disqus' Privacy PolicyPrivacy PolicyPrivacy**

Subscribe
to our newsletter

Enter your Email Address

SUBSCRIBE

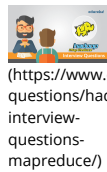## Related Blogs

Top 50 Hadoop Interview Questions You Must Prepare In 2018 (https://www.edureka.co/blog/interview-questions/top-50-hadoop-interview-questions-2016/)

(https://www.e
questions/top
50-hadoop-
interview-
questions-
2016/)

Top Hadoop Interview Questions To Prepare In 2018 – HDFS (https://www.edureka.co/blog/interview-questions/hadoop-interview-questions-hdfs-2/)

(https://www.e
questions/had
interview-
questions-
hdfs-2/)

Hadoop MapReduce Interview Questions In 2018 (https://www.edureka.co/blog/interview-questions/hadoop-interview-questions-mapreduce/)
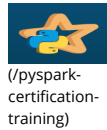
(https://www.
questions/hac
interview-
questions-
mapreduce/)

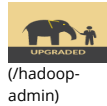Pig Tutorial: Apache Pig Architecture & Twitter Case Study (https://www.edureka.co/blog/pig-tutorial/)

(https://www.
tutorial/)

Big Data Analytics Courses

Big Data Hadoop Certification Training (/big-data-and-hadoop)

(/big-data-
and-hadoop)

Python Spark Certification Training using PySpark (/pyspark-certification-training)

(/pyspark-
certification-
training)

Hadoop Administration Certification Training (/hadoop-admin)

(/hadoop-
admin)

Apache Kafka Certification Training (/apache-kafka)

(/apache-

Edureka

About us (https://www.edureka.co/about-us)

News & Media (https://www.edureka.co/allmedia)

Contact us (https://www.edureka.co/contact-us)

Blog (https://www.edureka.co/blog/all/)

Community (https://www.edureka.co/community)

Work with us

Careers (https://www.edureka.co/careers)

Become an Instructor (https://www.edureka.co/instructors/add)

Become an Affiliate (https://www.edureka.co/affiliate-program)

Hire from Edureka (https://www.edureka.co/hire-from-edureka)

Useful Links

Reviews (https://www.edureka.co/reviews)

Terms & conditions (https://www.edureka.co/terms-and-conditions)

Privacy policy (https://www.edureka.co/privacy-policy)

Sitemap (https://www.edureka.co/sitemap)

Follow us on

f   🐦   in   ▶

(https://www.facebook.com/edurekaIN)(https://twitter.com/edurekaIN)(https://www.linkedin.com/company/edureka)(https://www.youtube.com/user/edurekaIN)

Learn on the GO!

(https://itunes.apple.com/in/app/edureka/id1033145415?mt=8)

(https://play.google.com/store/apps/details?id=co.edureka.app)

"PMP®","PMI®", "PMI-ACP®" and "PMBOK®" are registered marks of the Project Management Institute, Inc.
MongoDB®, Mongo and the leaf logo are the registered trademarks of MongoDB, Inc.