

[Apache Pig Interview Questions and Answers](#)

Practice Best Apache Pig Interview Questions and Answers

Apache Pig is a platform for analyzing large data sets that include a high-level language. It is used for expressing data analysis programs. It is coupled with infrastructure for evaluating these programs. The structure of Apache Pig is amenable to substantial parallelization, which in turn enables them to handle very large data sets.

Q1. [Explain what is Apache Pig?](#)

Apache Pig is a platform for creating programs that run on Apache Hadoop. It uses the Pig Latin language. It also executes its Hadoop jobs in MapReduce, Apache Tez, or Apache Spark.

Q2. [What is difference between Apache Pig and Hadoop?](#)

The difference between Apache Pig and Hadoop are as follows:

Topics	Apache Pig	Hadoop
Data Processing	It is used to analyze large sets of data representing them as data flows.	All the data manipulation operations in Hadoop are performed using Apache Pig.
Processing Speed	Apache Pig is faster than Hadoop.	Apache Pig is used in Hadoop.
Definition	Apache Pig is a platform for creating programs that run on Apache Hadoop.	Hadoop is a framework to process/query large data.
Operations	Apache Pig is a tool/platform which is used to analyze large sets of data representing them as data flows.	Hadoop is used for analytical and BIG data processing.
Operates On	Apache Pig operates on the Client-side of the cluster	Apache hive operates on the Server side of the Cluster
File Format	Apache Pig Supports Avro file format.	Hadoop also provides support for binary file format.

Q3. [What is BloomMapFile in Apache Pig?](#)

BloomMapFile in Apache Pig is a class that is used to provide a quick membership test for the keys using dynamic bloom filters. It extends the MapFile class.

Q4. What is Pig Latin?

Pig Latin is a language used in Apache Pig.

Q5. List some inbuilt Eval Functions of Apache Pig?

Some inbuilt Eval Functions of Apache Pig is listed below:

- AVG
- BagToString
- BagToTuple
- Bloom
- CONCAT
- COUNT
- COUNT_STAR
- DIFF
- IsEmpty
- MAX
- MIN
- PluckTuple
- SIZE
- SUBTRACT
- SUM
- IN
- TOKENIZE

Q6. What is use of PigDump and PigStorage functions?

PigDump used to Stores data in UTF-8 format, while **PigStorage** is used to Loads and store data as structured text files.

Q7. List some major differences between Apache pig and sql?

Some major differences between Apache Pig and SQL are listed below:

Pig	SQL
Pig Latin is a procedural language used in Apache Pig.	SQL is a declarative language.
Pig Latin data model is fully nested and can treat both atomic like integer, float, and non-atomic complex data types such as Map and tuple.	SQL data models are database dependent.

Apache Pig provides limited opportunity for **Query optimization**.

SQL provides more opportunities for query optimization.

Q8. What are scalar datatypes in Apache Pig?

Scalar/Primitive Types specify the type of data that a variable can contain. Generally, It consists of predefined data types.

- Int
- Long
- Float
- Double
- Chararray
- Bytearray
- Boolean
- Datetime
- Biginteger

Q9. Define different execution mode available in Apache Pig?

The three different execution modes are defined below:

Interactive Mode (Grunt shell) in Apache Pig includes the Grunt shell in which users can enter the Pig Latin statements and get the output (using Dump operator).

Batch Mode (Script) in Apache Pig allows writing the Pig Latin script in a single file with .pig extension.

Embedded Mode (UDF) in Apache Pig has the provision of defining User Defined Functions in programming languages such as Java and using them in our script

Q10. What is use of Grunt Shell?

Grunt shell is a shell of Apache pig to write commands that uses pig Latin scripts.

Q11. List out some Relational Operators available in Pig language?

Some Relational Operators available in Pig language is listed below:

- LOAD:
- FOREACH:
- FOREACH Result:
- FILTER:

- FILTER Result:
- JOIN:
- JOIN Result:
- ORDER BY:

Q12. List data models in Apache Pig?

The four data models in Apache Pig are listed below:

- Atom is an atomic data value that is used to store as a string.
- The tuple is an ordered set of the fields.
- The bag is a collection of tuples.
- The map is a set of key/value pairs.

Q13. What are Dynamic Invokers in Apache Pig?

In Apache Pig, Dynamic Invokers can be used to call a built-in static Java function that accepts a combination of strings, ints, longs, doubles, floats, or arrays, sometimes no arguments.

Q14. List some utility commands available in Apache Pig?

Some utility commands available in Apache Pig are listed below:

- Clear Command.
- Help Command.
- History Command.
- Set command.
- exec command.
- Kill Command.
- Run command.
- Quit Command.

Q15. How one can disable a Pig command and operator?

An admin feature provides the ability to blacklist or/and whitelist certain commands and operations that could be not very safe in a multitenant environment.

Blacklisting assigns "pig.blacklist" to a comma-delimited set of operators and commands. For instance, pig.blacklist=rm,killcross would disable users from executing any of "rm", "kill" commands and "cross" operator.

Whitelist disables all commands and operators that are not a safer part of the whitelist environment. For instance, pig.whitelist=load,filter,store will disallow every command and operator other than "load", "filter" and "store".

Q16. List some Diagnostic Operators available in Apache Pig?

Four Diagnostic operators available in Apache Pig are listed below:

- Dump operator.
- Describe operator.
- Explain the operator.
- Illustration operator.

Please Visit OnlineInterviewquestions.com to download more pdfs