



will increase by 25%



## APACHE CASSANDRA PROFESSIONAL CERTIFICATION BY DATASTAX: SAMPLE QUESTIONS

There are in total 207 Questions available

Note: These are set of sample questions which is taken from our collection of 600 questions.

1. Hadoop Training

2. Spark Training

3. HBase Training

4. MapR Developer

5. MapR HBase

6. CCA500 Certification

7. Spark Certification Send of the Offer Prices

8. EMC Data Science

<u>Hadoop Specialization offer == 50% + 35% off</u>

52000INR == 16900INR Only pert \$1150 == \$373 Only

Hadoop Specialization offer

I the total Time Office (I are These SDame Parents)

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



## **About Apache Spark Certification**

Apache Cassandra is one of the top NoSQL database which has features of both Google Bitable as well as Amazon DynamoDB. Most of the big organization globally already adopted Cassandra database and running in production you name the industry and they are using it. If you have checked the Cassandra database certification page, it is clearly mentioned that on average professionals get 31% higher salary if they know and have experience working with the Cassandra databases. And one of the recommended way to win first impression in interview is having certification in particular technology. And there is a company who conducts the certification exam for Apache Cassandra and that DataStax.

In the NoSQL world with the Apache Cassandra based solution DataStax is pioneer and they have certification to prove you have valid credentials to work with Cassandra database. Now the point is how to prepare for Apache Cassandra certification. One of the most popular certification for Apache Cassandra is "Professional Certification with Apache Cassandra: Massively NoSQL database" and this certification is for both Developer as well as architects.

This certification has combination of two things fundamentals of Apache Cassandra and DataStax features. Cassandra is among the most 5 popular databases of 2017-2018.

- Redis Popularity Rank No.1
- PostgreSQL Popularity Rank No 2
- MongoDB Popularity Rank No 3
- SQL Server Popularity Rank 4
- Cassandra Popularity Rank 5

There is high demand of Cassandra certified NoSQL developer (BigData really changed IT Technology world). As per DataStax site you can see there are 1000's of company like Apple, Netflix, Sony, Uber etc. are already using Cassandra database in production.

Question 1: Cassandra database is not good for, which of the following datatypes?

- A. large amounts of structured
- B. large amounts of semi-structured
- C. large amounts of unstructured data
- D. large amounts of XML data
- E. None of the above

Ans: E

**Exp**: Apache Cassandra is a massively scalable open source NoSQL database. Cassandra is perfect for managing large amounts of structured, semi-structured, and unstructured data across multiple data centers and the cloud.

Question 2. Select correct statement, which applies for Cassandra NoSQL database.



- A. Cassandra delivers continuous availability
- B. Cassandra delivers linear scalability
- C. Cassandra has operational simplicity across many commodity servers
- D. Cassandra can have single point of failure.

Ans: A, B, C

**Exp**: Apache Cassandra is a massively scalable open source NoSQL database. Cassandra is perfect for managing large amounts of structured, semi-structured, and unstructured data across multiple data centers and the cloud. Cassandra delivers continuous availability, linear scalability, and operational simplicity across many commodity servers with no single point of failure, along with a powerful dynamic data model designed for maximum flexibility and fast response times.

**Question 3**: You have been given a Cassandra cluster (Ring) with 5 nodes, and you want to create new KeySpace and insert data for your mobile application named AcmeApp.

- A. You have to connect primary node in Ring and needs to create Database connection with it.
- B. You can connect any two nodes (Defined by Admin) in Ring and needs to create Database connection any of these two nodes.
- C. You can connect any nodes in Ring and needs to create Database connection any of the node from cluster.
- D. You can connect any nodes in Ring but at a time you can use only one node from ring for your database operations.

Ans: C

**Exp**: Cassandra is a partitioned row store database: Cassandra's architecture allows any authorized user to connect to any node in any data center and access data using the CQL language.

For ease of use, CQL uses a similar syntax to SQL. The most basic way to interact with Cassandra is using the CQL shell, cqlsh.

Using cqlsh, you can create keyspaces and tables, insert and query tables, plus much more. If you prefer a graphical tool, you can use DataStax DevCenter.

For production, DataStax supplies a number drivers so that CQL statements can be passed from client to cluster and back.

**Question 4**: You have been given a Cassandra cluster (Ring) with 5 nodes, and new KeySpace created by Admin named AcmeKeySpace and you want to insert configuration data for your mobile application



named AcmeApp. And administrator had told you that default Replication Factor is 3 on this cluster for your AcmeKeySpace. Now you have inserted 50 records in a table of your AcmeKeySpace.

Select correct statement which apply.

A. After insert query, you have to execute another command named 'replicate-all', using cqlsh. So data can be replicated on three nodes.

B. Before insert query, you have to execute another command named 'replicate-all', using cqlsh, for the session. So data can be replicated on three nodes.

C. You don't have to do anything, replication is taken care by Cassandra itself.

D. Either of 1 and 2

Ans: C

**Exp**: Automatic data distribution: Cassandra provides automatic data distribution across all nodes that participate in a ring or database cluster.

There is nothing programmatic that a developer or administrator needs to do or code to distribute data across a cluster because data is transparently partitioned across all nodes in a cluster.

Built-in and customizable replication: Cassandra also provides built-in and customizable replication, which stores redundant copies of data across nodes that participate in a Cassandra ring.

This means that if any node in a cluster goes down, one or more copies of that node's data is available on other machines in the cluster.

Replication can be configured to work across one data center, many data centers, and multiple cloud availability zones.

**Question 5:** You have been given a Cassandra cluster (Ring) with 10 nodes, spread across two availability zones (AWS: amazon WebService) and new KeySpace created by Admin named AcmeKeySpace with Replication factor 3. Select correct statement.

- A. Whatever data updates are done cannot be replicated to another availability zone.
- B. Whatever data inserts are done cannot be replicated to another availability zone.
- C. Whatever data updates are done, automatically replicated to another availability zone.
- D. Whatever data inserts are done, automatically replicated to another availability zone.

Ans: C,D

**Exp**: Built-in and customizable replication: Cassandra also provides built-in and customizable replication, which stores redundant copies of data across nodes that participate in a Cassandra ring.

This means that if any node in a cluster goes down, one or more copies of that node's data is available on other machines in the cluster.



Replication can be configured to work across one data center, many data centers, and multiple cloud availability zones.

**Question 6:** Select all the features provided by Cassandra, from below list.

A. Cassandra is a partitioned row store database

B. Automatic data distribution: Cassandra provides automatic data distribution across all nodes that participate in a ring or database cluster.

C. Built-in and customizable replication: Cassandra also provides built-in and customizable replication, which stores redundant copies of data across nodes that participate in a Cassandra ring.

D. Cassandra supplies linear scalability

Ans: A,B,C,D

**Exp**: Cassandra's built-for-scale architecture means that it is capable of handling petabytes of information and thousands of concurrent users/operations per second.

Cassandra is a partitioned row store database: Cassandra's architecture allows any authorized user to connect to any node in any data center and access data using the CQL language.

For ease of use, CQL uses a similar syntax to SQL. The most basic way to interact with Cassandra is using the CQL shell, cqlsh. Using cqlsh, you can create keyspaces and tables, insert and query tables, plus much more. If you prefer a graphical tool, you can use DataStax DevCenter. For production, DataStax supplies a number drivers so that CQL statements can be passed from client to cluster and back.

Automatic data distribution: Cassandra provides automatic data distribution across all nodes that participate in a ring or database cluster. There is nothing programmatic that a developer or administrator needs to do or code to distribute data across a cluster because data is transparently partitioned across all nodes in a cluster.

Built-in and customizable replication: Cassandra also provides built-in and customizable replication, which stores redundant copies of data across nodes that participate in a Cassandra ring. This means that if any node in a cluster goes down, one or more copies of that node's data is available on other machines in the cluster. Replication can be configured to work across one data center, many data centers, and multiple cloud availability zones.

Cassandra supplies linear scalability: Cassandra supplies linear scalability, meaning that capacity may be easily added simply by adding new nodes online. For example, if 2 nodes can handle 100,000 transactions per second, 4 nodes will support 200,000 transactions/sec and 8 nodes will tackle 400,000 transactions/sec

Question 7: Select correct statement for Cassandra database, w.r.t. RDBMS

A. Cassandra is designed as a distributed database with peer-to-peer communication.



- B. As a best practice, queries should involve at least two table.
- C. Data should be de-normalized.
- D. client-side joins can be used in applications, after data fetch from Cassandra.

Ans: A,C,D

**Exp**: Cassandra is designed from the ground up as a distributed database with peer-to-peer communication. As a best practice, queries should be one per table. Data is denormalized to make this possible.

For this reason, the concept of JOINs between tables does not exist, although client-side joins can be used in applications.

Question 8: Select correct statement which applies for CQL.

- A. CQL and SQL share the same abstract idea of a table constructed of columns and rows.
- B. Cassandra emphasizes denormalization through CQL features like collections and clustering specified at the schema level
- C. Cassandra does not support joins.
- D. Cassandra does not support subqueries.

Ans: A,B,C,D

**Exp**: Cassandra Query Language (CQL) is the primary interface into the Cassandra DBMS. Using CQL is similar to using SQL (Structured Query Language). CQL and SQL share the same abstract idea of a table constructed of columns and rows. The main difference from SQL is that Cassandra does not support joins or subqueries. Instead, Cassandra emphasizes denormalization through CQL features like collections and clustering specified at the schema level.

CQL is the recommended way to interact with Cassandra. Performance and the simplicity of reading and using CQL is an advantage of modern Cassandra over older Cassandra APIs.

Question 9: Which all are possible using cqlsh

A. you can create keyspaces

B. you can create tables

C. you can insert and query tables

Ans: A,B,C

Exp: The most basic way to interact with Cassandra is using the CQL shell, cqlsh. Using cqlsh, you can create keyspaces and tables, insert and query tables, plus much more. If you prefer a graphical tool, you can use DevCenter. For production, DataStax supplies a number of drivers in various programming languages, so that CQL statements can be passed from client to cluster and back.



**Question 10:** You are working in an Investment Bank, and they have subscribed data feeds from Bloomberg, which delivers data in various different format e.g. CSV file, Direst database access either RDBMS or NOSQL(Cassandra) and daily incremental data. Which of the following you can use in Cassandra to load data or migrate data from one Cassandra cluster to another.

A. CQL INSERT command

B. correct option

C. correct option.

D. correct option

**Ans**: A,B,C,D

**Exp**: Data is inserted using the CQL INSERT command, the CQL COPY command and CSV files, or sstableloader. But in reality, you need to consider how your client application will query the tables, and do data modeling first. The paradigm shift between relational and NoSQL means that a straight move of data from an RDBMS database to Cassandra will be doomed to failure.

**Question 11**. A file to which Cassandra get full question from here hardware failure is called.

A. correct option

B. memtable

C. SSTable

D. correct option

Ans: A

**Exp**: A sequentially Get full explanation here activity to ensure data durability.

Question 12. Cassandra get full question from here employing

A. Master-slave Architecture

B. correct option

C. peer-to-peer TCP communication protocol

D. correct option

Ans: B

**Exp**: Cassandra is designed <u>Get full explanation here</u> other nodes across the cluster using peer-to-peer gossip communication protocol.

**Question 13:** Which of the get full question from here that resembles a write-back cache.

A. commit log



- B. correct option
- C. correct option
- D. redo log

Ans: B

Exp: A sequentially written commit log on each node captures Get full explanation here cache

## Question 14: Fill in blanks

Each time the memory structure is full, the data is written to disk in an SSTable data file. All writes are automatically partitioned and replicated throughout the cluster.

Cassandra get full question from here.

- A. correct option
- B. correct option
- C. Serialization, compaction
- D. repair, compaction
- E. tombstone, repair

Ans: A

**Exp**: Cassandra is designed to handle big data workloads across multiple nodes with no single point of failure. Its architecture is based on the understanding that system and hardware failures can and do occur. Cassandra addresses the problem of failures by employing a peer-to-peer distributed system across homogeneous nodes <u>Get full explanation here</u>.

**Question 15:** What get full question from here node?

- A. correct option.
- B. Node which contain master copy of any data.
- C. correct option
- D. Node which contain configuration for entire cluster.

Ans: C

**Exp**: Cassandra is a partitioned row store database, where rows are organized into tables with a required primary key. Cassandra's architecture allows any authorized user to connect to any node in any data center and Get full explanation here.

Question 16: Map the below, components



A. correct option
B. Data center
C. <u>correct option</u>
1. It can span Physical locations
2. <u>correct option</u>
3. A collection of related machines and can be physical as well as virtual
Ans:
A-2
B-3
C-1
Exp: Node
Where you store your data. It is the basic infrastructure component of Cassandra.
Data center : <u>Get full explanation here</u> .
Question 17: Map the followings
A. <u>correct option</u>
B. <u>correct option</u>
C. CQL Table
1. Consists of columns and has a primary key
2. <u>correct option</u> .
3. Maintained for each Cassandra table
Ans.
A-2
B-3
C-1
<b>Exp</b> : Commit log: All data is written first to the commit log for durability.
Get full explanation here.



Question 18: Select get full question from here protocol.

A. Using Gossip protocol nodes periodically exchange state information about themselves and about other nodes they know about.

B. The gossip process runs every 5 seconds and exchanges state messages with as many other nodes in the cluster as possible.

C. correct option.

D. correct option

Ans: A,C,D

**Exp**: Gossip is a peer-to-peer communication protocol in which nodes periodically exchange state information about themselves and about other nodes they know about. Get full explanation here

Question 19: Select all get full question from here nodes

A. Use the multiple list of seed nodes in a cluster, so all nodes quickly learn about all other nodes in the cluster

B. correct option

C. correct option

D. Seed nodes are a single point of failure

Ans: B,C

Exp: To prevent problems in gossip communications, use the same list of seed nodes for all nodes in a cluster. This is most critical the first time a node starts up <u>Get full explanation here</u>.

Question 20: Select all get full question from here nodes

A. correct option

B. correct option

C. Making every node a seed node is recommended for fault tolerance

D. It is recommended to use a small seed list (approximately three nodes per data center).

Ans: A,B,D

**Exp**: In multiple data-center clusters, the seed list should include at least one node from each data center (replication group). Get full explanation here.









Click Here







Click Here





Click Here

AWS Contribed Systèms Administrator Associate Level AWS Cartified Solutions Architect Associate Level

AWS Certified Solutions Architest Professional Level

Click Here for AWS Package Deal

All Premium Training Access Annual Subscription (You will get early access to under development training and early edition books): Used By More than 20000 subscribers

Spark Professional Training: HandsOn

**CLICK HERE** 

HadoopExam.com

32 Modules

NiFi : Hortoneworks DataFlow (HDF) Hands On Training

CLICK HERE

HadoopExam.com

16 Modules



Click Here

Hadoop Training With HandsOn



Cloudera Hadoop Admin Training Course-1





Click Here

e BASE Training Click Here

HBase Professional Training 35 Hands On Sessions 20 Module

OOzie Professional Training: HandsOn

CLICK HERE

HadoopExam.com

22 Modules

AWS Solution Architect : Associate Training

CLICK HERE

HadoopExam.com













Apache Spark Training & Certifications: Apache Spark is new and fastest data processing engine for Big Data world, after Hadoop it's becoming more popular in Industry (recently demand increased a lot). Now using power of Hadoop and Spark. Hence, data processing speed has dramatically increased. So if you wish to work in/with Big Data then Learning Spark is a must even for becoming data scientist., HadoopExam Learning Resources launched low cost material for in depth learning of Spark in the form of Spark Professional Training with Hands on practice sessions and helping you to get certified with most popular Apache Spark Certification conducted by Oreilly and Databricks only. So without delaying start preparing or prove your skills of Apache Spark, subscribe to our trainings and certification material with special discount of unbeatable price. You can request free updates as well, whenever it is done.













- 1. Apache Spark Professional Training with Hands On Lab Sessions
- 2. Oreilly Databricks Apache Spark Developer Certification Simulator
- 3. Hortonworks Spark Developer Certification
- 4. Cloudera CCA175 Hadoop and Spark Developer Certification

**Cloudera® Certifications Preparation Kits and Trainings**: Cloudera is a pioneer for Hadoop Big Data framework and they have grown a lot since last a decade. Cloudera® solutions is being used a lot in industry. They had also converted all their certification exam from multiple choice to Hands-on exam.



HadoopExam was the first one, who launched Cloudera certification material 5 years back and since than we have also grown and keeping in pace with Cloudera new certifications. We also provide industry class training used by more than 10000 learners across the globe. Check all the products below for more detail.



- 1. CCA 175 : Cloudera® Hadoop & Spark Developer : 95 Solved Scenarios
- 2. CCA159: Cloudera® Data Analyst Certification: 73 Solved Scenarios
- 3. CCA131 : Cloudera Hadoop Administrator Certification : 92 Solved Scenarios
- 4. CCP:DE 575 : Cloudera Hadoop Data Engineer : 79 Solved Scenarios
- 5. CCA 500 : Hadoop Admin Certification : 250+ Practice Questions
- 6. Training: CDH: Cloudera Hadoop Admin Beginner Course-1: 30 Training Modules
- 7. Hadoop Professional Training
- 8. HBase Professional Training
- 9. Cloudera Hadoop Developer (CCD410) Certification : Retired
- 10. Cloudera HBase Specialist (CCB400) Certification: Retired
- 11. Hadoop Package Deal

**About Hortonworks® Training & Certifications:** Hortonworks is one of the leader in providing Big Data solution through their own HDP platform. To check candidate's proficiency or skills for HDP platform they have various certification exams. HDPs most of the exam are Hands-on exam other than HCA (Hortonworks Certified Associate). All the exam aspirant has to solve given tasks on HDP cluster. In each exam there are approx. 10-12 problem scenario would be given and needs to be solved in 2 Hrs. Being



an Hands-on exam, these certifications has high value in industry, because it require real hands on experience to solve given scenario. Hence to help you, HadoopExam is providing from scratch how to setup environment to practice scenarios. HadoopExam also provides the complementary videos, where we guide you how to solve problems and setup the environment. Currently we have following certification preparation material available.



- 1. HDPCD: Hadoop (HDP) No Java Certification: 74 Solved Scenarios
- 2. HDPCD-Spark: HDP Certified Developer: 65 Solved Scenarios
- 3. HDPCA: HDP Certified Administrator: 57 Solved Scenarios
- 4. Hortonworks Certification Package Deal

**Data Science & Machine Learning**: Data Science is one of the most demanding field, currently and we are providing following products to become a data scientist from one of the popular organization in the data world EMC



- 1. Data Science Certification EMC® E20-007 (Data Science Associate)
- 2. EMC® Data Science Specialist (E20-065)
- 3. Cloudera Data Scinece DS-200 (235 Questions + 150 Page Study Notes): Retired



**MapR® Training & Certifications**: MapR is another most popular BigData solution provider based on Hadoop. These are the following certifications, which HadoopExam is providing currently.



- 1. MCSD: MapR Spark (Scala) Certified Developer
- 2. MapR Hadoop Developer Certification
- 3. MapR HBase NoSQL Certification
- 4. MapR Package Deal

**AWS Training & Certifications**: In the Cloud computing world, Amazon is a pioneer and most used Cloud Computing solutions. Currently there are following products are provided bt HadoopExam for the AWS trainings and certifications preparation. We have been providing this matrial since last approx 5 years and many 1000s of learners already using our material to grow in their career.





266 Q & A Click Here



300 + Q & A Click Here



474 + Q & A Click Here



Click Here



AW6 Certified SysOps Administrator Associate Level

A W/S Cartified
Solutions
Architect
Associate Level

# WS Contified Solutions Architect Professional Level



Click Here

Click Here for AWS Package Deal

- AWS Solution Architect Associate : Training
- 2. AWS Solution Architect Associate Certification Preparation
- 3. AWS Solution Architect Professional Certification Preparation
- 4. AWS Sysops Certification Preparation
- 5. AWS Developer Certification Preparation

**IBM® BigData Architect**: This is a multiple choice exam conducted by IBM for a BigData Architect. IBM also has Hadoop framework known as BigInsight and they will be asking Question based on BigInsight, however it is very similar to Hadoop only, because they are using Apache Hadoop framework only. As you know, IBM is the oldest and one of the matured software vendor and they have more penetration in the Industry, compare to any other BigData vendor. Hence, certifying yourself as a BigData Architect for IBM, certainly have high value in industry.



IBM C2090-102 Big Data Architect



• IBM C2090-102: IBM Big Data Architect: Total 240 Questions: Highest number of Questions: 95% Questions with explanations

DataStax® Apache Cassandra Certification: This is a multiple choice exam conducted by DataStax for Apache Cassandra. DataStax is one of the leader in providing Apache Cassandra based solutions. Apache Cassandra is one of the most demanding and used NoSQL database across the industry. Cassandra has been used in Finance, HealthCare, Aviation, Retail, e-commerce and many more. It has proved itself with high degree of performance. However, it's a different database and RDBMS principals does not fit with Cassandra. You certainly need to learn Cassandra Data Modeling to design database perfectly and this certification is designed towards this only. And HadoopExam had put lot of effort to create this material to help in clearing this certification exam.



<u>Professional Certification Apache Cassandra(Datastax)</u>: <u>Total 207 Questions</u>: <u>Highest number of</u>
 Questions: 95% Questions with explanations

**SAS**®: One of the most used commercial solutions for analytics, Data science, mathematical and statistical modeling. In analytics world no other solution is close to SAS. Its leader in its field and mostly used across industry. Below are the all products provided by HadoopExam.





- 1. SAS Base Certification Professional Training
- 2. SAS Base Programming Certification(A00-211)
- 3. SAS Certified Advanced Programmer for SAS 9 Credential
- 4. SAS Certified Statistical Business Analyst Using SAS 9: Regression and Modeling Credential
- 5. SAS Certified Platform Administrator 9 (A00-250) Certification Practice Questions
- 6. SAS Package Deal

**HBase Training & Certifications**: HBase is a NoSQL solution based on Hadoop framework. Hence, is very well compitible with the Hadoop based solution. You should certainly learn HBase, if you are working in BigData world using HadoopExam. Following are the products provided by HadoopExam for HBase.





- 1. HBase professional Training with HandsOn Sessions
- 2. MapR HBase certification preparations

**Microsoft® Azure:** Microsoft Azure is another provider for Cloud computing solutions and also heavily used in the industry. If you are planning to make your career in Cloud computing than you should have very good understanding of the Microsoft Azure. Please find all the products and solution provided by HadoopExam for the Azure.



- 1. Microsoft Azure 70-532 Developing Azure Solution Certification
- 2. Microsoft Azure 70-533 Implementing Microsoft Azure Infrastructure Solutions

Oracle Cloud, Java and Other Programing Trainings and Certifications: There is no development without a programming skills. We provide trainings and certification material which will make you developer who can work in well developed IT industry, with the most demanding programming skills. So start learning Java, Scala, Python and complete its certifications as well. Please check all the available products below.

## HadoopExam









- 1. Oracle 1Z0-337 Oracle Oracle Infrastructure as a Service Certified Implementation Specialist
- 2. Full length HandsOn Step By Step Training for Java 1z0-808)
- 3. Scala Professional Trainings with HandsOn Session
- 4. Python Professional Trainings with HandsOn Session
- 5. Java SE-8 Programmer-1 (1z0-808) Certification
- 6. Java SE-8 Programmer-2 (1z0-809)
- 7. JAVA EE Web Services Developer (1z0-897)
- 8. Oracle® 1Z0-060: Upgrade to Oracle Database 12c Administrator
- 9. Questions for Oracle 1Z0-061 : Oracle Database 12c: SQL Fundamentals