CRT020: Spark Certification Guide (Scala)

Prepared By ©HadoopExam.com

Edition : First

Table of Contents

About book

Apache® Spark is one of the fastest growing technology in BigData computing world. It supports multiple programming languages like Java, Scala, Python and R. Hence, many existing and new framework started to integrate Spark platform as well in their platform e.g. Hadoop, Cassandra, EMR etc. While creating Spark certification material HadoopExam technical team found that there is no proper material and book is available for the Spark (version 2.x) which covers the concepts as well as use of various features and found difficulty in creating the material. Therefore, they decided to create full length book for Spark (Databricks® CRT020 Spark Scala/Python or PySpark Certification) and outcome of that is this book. In this book technical team try to cover both fundamental concepts of Spark 2.x topics which are part of the certification syllabus as well as add as many exercises as possible and in current version we have around 46 hands on exercises added which you can execute on the Databricks community edition, because each of this exercises tested on that platform as well, as this book is focused on the Scala version of the certification, hence all the exercises and their solution provided in the Scala. We have divided the entire book in the 13 chapters, as you move ahead chapter by chapter you would be comfortable with the Databricks Spark Scala certification (CRT020). All the exercises given in this book are written using Scala. However, concepts remain same even if you are using different programming language.

Feedback

This is a full-length book from http://hadoopexam.com and we love the feedback so that we can improve the quality of the book. Please send your feedback on hadoopexam@gmail.com or admin@hadoopexam.com

Restrictions

Copyright© Material

written permission of the author and publisher, except in the case of brief quotations embedded in critical articles or online and off-line reviews. Wherever, you use contents make sure full detail of the book is mentioned.

Author had tried as much as his capacity in preparing of this book so that accuracy can be maintained in the presented material. The material sold using this book does not have any warranty or guaranty either express or implied. Neither of the author, publisher, dealer and distributors will be held liable and responsible (explicit/implicit these all parties mentioned are not liable and responsible) for any damages caused or alleged to be caused directly or indirectly by this book. You should note this material as part of your learning process and as time passes material can be outdated and you should wait or look for that latest material.

Author and publisher has endeavored to provide trademark information about all of the companies and products mentioned in this book. However, we cannot guarantee the accuracy of this information.

Disclaimer:

1. Hortonworks® is a registered trademark of Hortonworks.
2. Cloudera® is a registered trademark of Cloudera Inc
3. Azure® is a registered trademark of Microsoft Inc.
4. Oracle®, Java® are registered trademark of Oracle Inc
5. SAS® is a registered trademark of SAS Inc
6. IBM® is a registered trademark of IBM Inc
7. DataStax ® is a registered trademark of DataStax

8. MapR® is a registered trademark of MapR Inc.
9. Apache® is a registered trademark of Apache Foundation
10. Databricks® is a registered trademark of Databricks Inc

Publication Information

First Version Published: Nov 2019

Edition: 1.0

Piracy

We highly discourage the piracy of copyright material especially it happened online on the internet. Piracy causes the damages to all first of all it damages yourself by not honestly using the correct material, generally pirated material is edited and wrong information is presented which can make big damage as part of your learning process. As well as when you become author and honestly write similar material, piracy will damage your material as well. Hence, don't encourage piracy. If piracy is reduced cost of material will automatically decreases. It also makes damages to author, publisher, dealer and distributors. If you come across any illegal copies of this works in any form on the Internet, then please share the detail with the URL, location or website name immediately on email id hadoopexam@gmail.com we really appreciate your help in protecting author's hard work and also help in reducing the cost of material.

Author/Trainer required

Corporate Trainer: We have many requirements, where our corporate partners need their team to be trained on particular skill sets. If you are already providing corporate trainings for any skills set, then please become our onsite training partner and fill in the form mentioned above and our respective team will contact you soon. You will get very good revenue for sure. However, what we want, you must be able to train our corporate partner resources. What matters to us? Your proficiency in a particular domain/skill and good oral communication skills. You must be able to accessible to learners as well.

Online Trainer: If you are a working professional and master or proficient in any particular skills and feel that, you are capable of giving online virtual trainings e.g. 2 hrs a day until course contents are completed. Please fill in above form and our respective team will contact you or send an email at admin@hadoopexam.com . You will get a very good revenue share for sure. What matters to us? Your proficiency in a particular domain/skill and good oral communication skills. It will certainly not impact your daily work.

Self-Paced Trainings: Ok, you want to work as per your comfortable time and at the same time sharpen your skills. You can consider this option. You can create self-paced trainings on particular domain/skills. Please fill in above form to connect with us as soon as possible. Before somebody else

connect with us for the same skill set. Your commitment is very important for us. We respect your work and we will not sell your work in just $10 to acquire more resources. As we know, it takes a good amount of time and you will provide quality material, so we charge reasonable on that so, you will feel motivated with your work and effort. We respect you and your skill.

Certification Material: You may be already certified professional or preparing for particular certification in a specific domain/skill. So why not use this to make money as well as sharing your effort with other learners globally. Please connect with us by filling form or send email at admin@hadoopexam.com and our respective team will contact you soon.

Author: Yes, we are also looking for authors. Who can write books on a particular technology and what you can get certainly a very good revenue sharing and you can bring the same on your resume or linked in profile to show your excellence? Yes, we are not in need of very good oral communication skills, but good writing skill. However, team will also help you to get work done. Author can be more than one for a particular book. However, we wanted you to be in long relationship. So that you don't just write a single e book, but can create an entire series for a particular domain or skill. Good royalty for sure...

Trending Skills (Not limited these):

Hadoop Spark AWS Cloud Azure Cloud Google Cloud	EMC NetApp VMWare CISCO HP	Adobe Alfresco Apple AppSense AutoDesk	Data Analysis DJango Docker Drupal Graphics	Infrastructre Automation Internet of Things (IOT) ISO Development Java Java Script
JQuery Kali Linux Laravel Linux Machine Learning	Mobile Application Development NodeJS Android Angular JS Arduino	IBM Watson IBM BPM WebMethod Gemfire Liferay	Scala Python Java SQL/PLSQL Ruby	SAP SAS Salesforce Oracle Cloud Redhat

Chapter-1: CRT020 Databricks® certified Associate Developer using Scala

Access Source code: As this book has around 46 hands on exercises and you wanted to download the same. Link for downloading the source code is provided before the start of each chapter, wherever it is required. From chapter-6 onwards we would be doing hands-on exercises.

Access to Certification Preparation Material
I have already purchased this book printed version from open market, I still wanted to get access for the certification preparation material offered by HadoopExam.com, do you provide any discount for the same.
Answer: First of all, thanks for considering the learning material from HadoopExam.com. Yes, we certainly consider your subscription request and you are eligible for discount as well. What you have to do is that, you can send receipt this book purchase and our sales team can offer you 15% discount on the preparation material. Please send an email to hadoopexam@gmail.com or admin@hadoopexam@gmail.com with the purchase detail and your requirement

If you purchase eBook version of this book from HadoopExam.com website then all future edition of the same book, would be available to you as well. Without any additional fee.

Why Spark framework is so popular

Apache Spark is one of the fastest growing technology for the Data processing, Data Analytics,

Machine Learning, Graph Processing and Data Science. Reason for its adoption in the industry are various for example on the macro levels we can say, it has:

- Big organization which are supporting this in production like Cloudera, Databricks, MapR, Microsoft, IBM, Datastax etc.
- API is very developer friendly, mainly after the release of Spark 2.x
- Spark supports already popular programming languages like Scala (Spark framework, itself written using Scala), Java, Python and R. Hence, industry do not have to train developer for specific programming language if they are already having resources with any of these programming skills and they have to become various other aspects of the Spark framework.
- Support of Structured Query Language, most of the Data Analytics/engineer already well versed with the SQL. And Spark also supports very well the same SQL syntax.
- Frequent releases with the new features and enhancements.
- Much faster processing engine compare to any other available Data processing engine.

There are many other things which make the Spark very popular technology. These are few of the reason, for which you have selected this book and CRT020 certification exam.

Introduction to CRT020 Certification

As demand is growing day by day for the Spark Developer and industry wants easy access to Spark professional and for searching right candidates, they don't have to spend so much time. To find the resources which are good or have some knowledge of the Spark framework and from the candidate side, it should also be easy to prove by showing that he isalready a certified professional in Spark technology. There are various Spark certifications but CRT020 became very popular recently because this is conducted by the company called Databricks, who heavily spend their time on the Spark framework development as well as they have their own enterprise version of the Spark framework with the additional enterprise feature. Databricks is conducting Spark certification since many years and they have different certifications for Python and Scala programming language, and to pass this certification you have to have fundamental knowledge about how Spark works and similarly have good experience for doing hands on with the

Spark. Hence, it is recommended you complete all the exercises given in this book as well as in the certification preparation material provided by HadoopExam.com . In next few sections we would be discussing about the frequently asked questions about the Spark CRT020 certification.

Where and How to get Databricks Spark CRT020 Certification Sample Questions

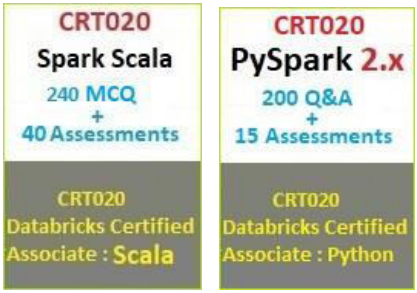
There are various Spark certification exams available and this particularly this one "CRT020 : Databricks Certified Associate Developer for Apache Spark 2.4 and Scala 2.11 - Assessment Certification Exam " is the latest available Spark exam from the Databricks and similarly Python version. This certification became popular in very short span of time and within the launch on HadoopExam.com , more than 100 learners have subscribed in a week. This prove that, how popular is this certification exam. And also this is based and tested on the Databricks Enterprise version of the platform.

Even it uses the Databricks Enterprise version but its underline engine is same as Apache Spark, hence, the same code you can run on the Apache

Spark as well as on Databricks Spark platform.

However, it is recommended that you practice very well before you appear in the real exam. Because without practice, you would not be able to complete the exam on time. CRT020 exam is divided in two major section as below.

- Multiple Choice Questions (Get access to all 240 Multiple Choice Questions from Here Scala , PySpark)
- Assessment (Hands On Section) : Get access to all 40+ assessment Questions and Answer (Including Videos) Scala or PySpark



If you want to check the Sample Questions and Answer then use the below link or watch the below video to understand more.

- Scala : http://hadoopexam.com/spark/databricks/SparkScalaCRT020DatabricksAssessment.html
- PySpark : http://hadoopexam.com/spark/databricks/PySparkCRT020DatabricksAssessment.html
- Sample Assessment PySpark: http://learn.hadoopexam.com/PySparkCRT020/SampleAssessment/index.html
- Sample Assessment Scala : http://learn.hadoopexam.com/SparkScalaCRT020/SampleAssessment/index.html
- Multiple Choice:

  http://learn.hadoopexam.com/SparkScalaCRT020/Sample/index.html

How you should prepare for CRT020 Spark Scala/Python (Databricks) Certification Exam?

Databricks is the leader for Apache Spark technology, they support the open source version of Apache Spark framework.

Based on the open source Apache Spark, Databricks created enterprise version of Spark Framework. And this newly created framework also works on

the Cloud platform like AWS, Azure, Google cloud etc.

Since last few years Databricks platform became very popular because they are capable of deploying Spark in the production env. Enterprise or companies who all are using Databricks platform in production or planning to have in production in need of Databricks certified professionals. Databricks has following two popular certifications as of today. They might come more in future for different solutions like Machine Learning, Graph and Structure Streaming etc. Let's go through below two links for the currently available certifications.

- CRT020 : Databricks Certified Associate Developer Apache Spark 2.4 with Scala 2.11 : Assessment Certification (Newly launched & Active)
- CRT020 : Databricks Certified Associate Developer for Apache Spark 2.4 with Python 3.0 - Assessment Certification (Newly launched & Active)

Both the above certification exam has the same pattern and syllabus. Only difference is, which programming language you prefer.

Exam format: In each certification exam there are two sections as below

- **Multiple choice** questions and answers (which include single and multiple correct answers, fill in the blanks questions and answers etc.)
- **Assessment Exam:** You need to write complete solution for given problem statements. Also, link would be provided for downloading or accessing the data.

However, it is not mentioned on the certification detail page that how many questions they would be asking in each section. HadoopExam.com experience shows that there would be around 20 multiple choice questions and 20+ assessment exercises would be given and difficulty level would increases Question by Question, Same is provided on HadoopExam online Spark Certification Simulator. It is clearly mentioned that the exam would be 3 hrs long and include both the above section. Hence, please note that

- In multiple choice 20 questions would be covered. In that they would be asking various concepts, internal Architecture, API and SQL functions-based questions.

- Around 20 assessment questions would be asked, in this you would be given problem statement for each question and you need to write or implement the solutions either using PySpark or Spark Scala.
- You need to write problem solution in online version of Databricks Enterprise platform.
- **How the Scoring would be done?** Databricks have not mentioned, whether you need to pass separately each exam section or aggregate score from both the section would be considered. What HadoopExam.com experience again says here that you need to score 75% marks in each section at least so that your overall aggregated score remains 75% as well and you can clear the exam. Whether Databricks consider individual section or aggregated marks.

Timeline for CRT020 Spark Certification preparation

Preparations and timeline depend on the how good you are in Spark technology as well as what is your strength in Scala and Python programming language. As per HadoopExam.com experience following timeline you can consider for preparing this certifications, if you spend 2-3 hrs. 5 days a week.

- **6 months:** If you are completely new to Spark.
- **3-4 month** : If you know one of the programming language like Java, Scala, or Python etc.
- **1-2 month**: If you already know Spark technology.
- Above timeline is not perfect these are derived based on HadoopExam.com previous experience with other certifications.

How to prepare for CRT020 Spark Certification

To prepare for the Spark certification you need to have right material, and also you need to properly planned and have properly drafted material, which can save your lot of time. Otherwise, you would be going for material here and there and lose lot of time and it may take much longer to complete the exam even without having full confidence in the real exam. Also, remember if things are not properly planned and drafted or organized, it does not matter how good you are in Spark.

To make your life simple and easy for the Spark CRT020 certification preparation HadoopExam.com have created cool material. You should consider the following material for preparing Spark Certification

1. CRT020 : Databricks Certified Associate Developer Apache Spark 2.4 with Scala 2.11 : Assessment Certification (Newly launched & Active) : Include 200+ multiple choice questions and more than 40 assessments.
2. CRT020 : Databricks Certified Associate Developer for Apache Spark 2.4 with Python 3.0 - Assessment Certification (Newly launched & Active) : Include 200+ multiple choice questions and more than 30 assessments. More would be added soon.
3. Apache Spark Professional Training with Hands On Lab Sessions (Active)
4. Spark 2.X SQL (Using Scala) Professional Training with Hands On Sessions
5. PySpark 2.X (Using Python) Professional Training with Hands On Sessions
6. Scala Professional Training with HandsOn Session
7. Python Professional Training with HandsOn Session

All the required questions come with the full explanation of the questions and answer. To justify the correctness of the questions and answers.

- It covers the entire syllabus for both Python and Scala version of certification exam. You can attempt their questions and answers as many times as you want.
- All multiple-choice question and answer, you can access from any device where browsers are supported like Desktop, Macbook/IOS, iPhone, mobile, tablet etc.
- There are no separate installations are required.
- Most of our learners are happy that because while travelling or during free time they can access the certification preparation material as well as interview questions audio cum video book.
- You can check some sample questions and answers as below

    - Check Sample Assessment Paper **(Scala)**
    - Check Multiple Choice Sample Paper(Scala)
    - Check Sample Assessment Paper **(Python)**

- Check Multiple Choice Sample Paper (Python)
- Video Cum Audio Book: Spark 2.x Interview Preparations (Total 185+ Interview Questions): Video + Audio + PDF

More detail on assessment exam

HadoopExam.com give capability to you for accessing problem statement and assessment solutions which can be accessed from mobile and tablet and that you can understand the same in detail. Once you understand the problem statement, then in the next tab, you would be given instructions to access or download the data which you need to use for solving the problem statement.

Videos: Possibly for selected assessment would have videos as well as, HadoopExam.com would explain the entire problem statements and its solution. However, it is not guaranteed that each assessment would be having the videos.

Assessment Solution: We are providing step by step solution for the given problem in multiple steps. Each step would be written with the detailed

comments as well. So that you can easily understand what is being done in the solution.

Training: HadoopExam.com has very popular training for Apache Spark, Spark SQL, Structured Streaming in Python and Scala. As well as interview Questions Audio cum video books. These all are On-Demand training access which you can access anytime anywhere using mobile, desktop, MacBook, iPhone etc. Check all below and more material would be added soon.

Spark Interview Preparation

By going through certification exam and training, your ultimate target is to join the companies which are using these new platforms or if you are already working in the organization then you are looking for vertical growth or increase on pay package and salary. Hence, HadoopExam.com prepared almost 185+ Interview Questions and answers which you can access in these two formats EBook and Video cum Audio Book format. This material if you want

to read you can read, you want to watch you can watch and if you want to listen then you can listen as well anytime-anywhere. Check more detail as below

Chapter-2: FAQ for Spark CRT020 Certification

Spark CRT020 Certifications FAQ (43 FAQs)

Question-1: I am a Java programmer, which language I have to choose for this CRT020 Spark certification?

Answer: As you know currently there is no specific certification in Java programming language for Spark. But Spark fully support Java programming language. Spark framework is written using the Scala framework and the Scala itself uses Java Run time environment. Hence, you should be quite comfortable with the Spark CRT020 certification using Scala framework.

Question-2: I don't know Scala programming language, is it required to be an expert in Scala to work on the Spark Framework using Scala?

Answer: No, we don't think so. If you know the basics of Scala and you are fluent in one of the programming languages then it is good enough. You can attend crash courses for Scala and Python on HadoopExam to learn the same.

Question-3: I prefer Python, do you have material specific the Python or PySpark?

Answer: Yes, this book you are reading has both the version Spark Certification using Scala and Using PySpark. So, you can choose as per your requirement. Similarly, all the practice material created on the HadoopExam.com are also segregated based on the programming language.

Question-4: How many questions are expected in the real exam, as I see HadoopExam has around 200+ practice questions and around 40 assessments?

Answer: We are providing practice questions which are based on the feedback provided by the learners and expertise of our technical and engineering team. And we want you practice as much as possible before your real exam. In real exam, based on our learners' feedback you will get

- Around 20 multiple choice questions (included fundamentals concepts as well as some programming questions)
- Around 20 assessments, which you need to complete on the Databricks community edition provided in the Cloud env.

Question-5: What is your recommendation regarding spending time on multiple choice questions and assessment questions?

Answer: HadoopExam recommend that you should be able to complete all multiple-choice questions and answer in less than 25 mins.

As assessment questions can take more time, so spend around 2 hr 30 mins on assessment. You can see there are around 20 assessment questions.

Question-6: Do you know how is marking done between multiple choice and assessment questions?

Answer: No, as of now we don't know. But it seems assessment questions would have more weightage. However, we still don't recommend skipping multiple choice questions at all.

Question-7: What is your recommendation during the real exam for attempting the questions?

Answer: It is similar to any other exam which you have appeared till now. Always attempt easier questions first and then do all the tough questions once you are done with easy questions. If you got stuck on a particular question then don't spend too much time on it and try to attempt another easy question. This is universally known strategy. But yes, for this CRT020 certification exam complete all multiple-choice questions first in less than 25 mins.

Question-8: Is multiple choice questions had more than one answer correct?

Answer: Till now, whatever feedback we have received. All multiple-choice questions are having only single correct answer, but it is not guaranteed for future exam. Because Databricks has not explicitly mentioned it.

Question-9: Is it mandatory to attempt multiple choice question first and then coding question?

Answer: No, it is not mandatory in CRT020 certification exam. But we recommend you spend your initial time on multiple choice. Because once you start assessment question and then coming back to multiple choice question is little hard. However, it is allowed to switch between these two sections.

Question-10: Is there any specific section from which multiple-choice questions are being asked?

Answer: Again, this is not mentioned specifically on the exam guide. But we have seen more questions are being asked to check your understanding of the Spark fundamentals and the topic mentioned on the first 4 section are frequently being asked in the multiple-choice questions.

Question-11: Still can you specify which section; we need to prepare specifically for multiple choice questions?

Answer: Ok, for that you should consider the following sections

- What is the use of Spark Driver component?
- What is the relation between core and executor?
- How executor and tasks are related
- What do you mean by partitioning and how Spark parallel processing affected by partitioning?
- Understand these three components working in detail
    1. Jobs
    2. Stages
    3. Tasks
- And how all these are related to each other.
- What is the caching, and how it can be implemented?
- You would certainly get multiple choice question based on caching and memory management.
- Understand the Spark architecture
- Make yourself well aware about wide and narrow transformation (discussed in this book in depth)

Question-12: How complex or tough to resolve assessment question and answer?

Answer: We have seen that out of 20, around 6-7 questions are quite easy. And 3-4 questions are time consuming and little hard as well. And rest are medium level. If you have completed all the exercises from this book and Spark practice material then you will feel this exam is quite easy to crack. Even after completing all the assessment, we are sure that you are quite comfortable working using the Spark framework.

Question-13: Do you thing we should cover each individual topic mentioned in the syllabus for the CRT020 certification?

Answer: As you can see syllabus is quite wide compare to any other certification. And you should not skip any section from the syllabus before the real exam. In some situation, if you have not done 1 or 2 section from the entire syllabus that is ok. But dont skip more than 1 or 2 topics mentioned in the syllabus. We are also doing hard work for your success then why do you want to skip any section, lets complete all before your real exam.

Question-14: Can you please provide the detail, what kind of questions are being asked for the assessments?

Answer: Regarding the kind of assessment questions, you would be asked questions like below but not limited, again complete all the questions

and answer from this book as well as practice material provide by HadoopExam.com

- Load the data from file (most frequently asked parquet, JSON) and possibly other format as well like text, csv. Each exam attempt has different questions and answer.
- Create DataFrame and extract the data from it by applying projection or filter
- De-duplicate the data
- Find the distinct records from the DataFrame
- Transform the DataFrame by applying Lambda functions.
- Finally write the data to the file store like in Parquet, JSON or text format.
- Make yourself comfortable with the following file formats in order of priority
    1. Parquet
    2. JSON
    3. CSV
    4. Text

Question-15: I am already certified with Spark 1.6, what is your recommendation for this certification preparation?

Answer: Its good, then for preparing for this certification is even easier for you. Because the API in Spark 2.x is much easier to use compare to RDD API.

Question-16: I am already certified in Spark 1.6, why should I go for Spark CRT020 certification?

Answer: Spark had done a major change in Spark 2.x and most of the API is re-written to support for

- Project Tungsten
- Catalyst optimizer

And you should know all this, if you are building your career with the Spark technology. And there are many more new things, we highly recommend that you always update your certification credentials. As in Spark 1.6 major focus was on RDD, DStream and this is not at all recommended in Spark 2.x for programming but rather you should use Spark SQL framework heavily for ETL, Data analytics workload.

Question-17: Is Structured Streaming and Machine learning being asked in the real exam?

Answer: No, only the things which are explicitly mentioned for the syllabus, is being asked in the exam.

Question-18: I see in HadoopExam practice questions has the questions related to RDD API, is CRT020 asks questions based on RDD API?

Answer: No, in the real exam, you don't have RDD questions. But as of now, we kept it and updating the exam regularly. Soon you would see more

questions would be added and all RDD API questions would be removed from practice exam. Please ignore those questions as of now.

Question-19: Should we memorize the Spark API for CRT020 exam?

Answer: As on the exam instructions it is mentioned that you would be provided with the API doc and you can search the same during you real exam. But HadoopExam highly recommend that all frequently used API and packages you remember & memorize, so that you don't have to waste your time on finding the required methods from the docs. However, make yourself comfortable with the API doc as well, before the exam.

Question-20: Is HadoopExam providing any specific notes for memorizing the API for this certification exam?

Answer: As of now we don't have, but in sometime we would have. So, if you have subscription on the HadoopExam.com for the certification preparation or annual subscription then you can get the access for the same, once it is released. You can keep visiting release and update tab on the HadoopExam.com website.

Question-21: Do you recommend which pages we should have try and make myself comfortable.

Answer: for **Python** use below

1. https://spark.apache.org/docs/latest/api/python/pyspark.sql.html
2. https://docs.databricks.com/

For **Scala** use below

1. https://spark.apache.org/docs/latest/api/scala/#org.apache.spark.sql.package

In this check API related to below components

- Dataset/DataFrame
- Row
- DataFrameReader
- DataFrameWriter
- Column

Question-22: I am good at Spark SQL; can I avoid using DataFrame at all in the exam?

Answer: In the exam you would see most of the questions are based on the DataFrame and initial code snippet also they are giving using DataFrame. So, try to solve using DataFrame first, if not comfortable then switch the Spark SQL API. It may eat some of your time.

Question-23: Where should I practice this exam. HadoopExam provide any environment for practicing the questions?

Answer: No, HadoopExam does not provide any environment for practicing coding question. You can use the Databricks community edition for the same. (We would be providing the videos, how you can use the same). If it is currently not available then soon it would be released.

Question-24: How long does Databricks take to announce the result?

Answer: Initially, user was complaining that they are not getting the result until one week. But we have seen recently learners are getting their result on the same day. If not same day, then within 2-3 days they are announcing the result.

Question-25: Are you sure 20-25 mins are good enough for multiple choice questions?

Answer: Most of our learners who had practiced well, completing multiple choice section in less than 20 mins. So please read contents from the book provided by HadoopExam.com carefully before your real exam.

Question-26: My friend was saying that coding questions in the CRT020 exam are tough?

Answer: These questions are not very tough really, few questions you may feel tough. If you have not practiced well, if you know the stuff then questions are not that tough. Yes, that is possible that data

processing or understanding the data may take more time for you.

Question-27: Why people say, keep Spark API by heart for this CRT020 exam?

Answer: As we suggested before, because we have seen learners are not able to complete the assessment exam on time. Because they spend more time on the documentation. We are again suggesting please memorize the API as much as possible, specially the things which are frequently used. Like Row, DataFrame, Select, filter, distinct, foreach, take, persist, format, load, StructType, StructField etc.

Memorize how to set the properties like "spark.sql.shuffle.partions" how it is set on SparkSession or SparkContext. Soon HadoopExam would be creating quick reference or revision notes the same.

Question-28: Is there really time-pressure in the exam?

Answer: Simple rule, if you take pressure then certainly it is. If you don't take pressure and calmly go through each question it is fine. Even you don't know the API search in the doc (use CTRL+F for browser search and find specific keyword etc.)., always keep document opened in another tab of the browser, so you can immediately check the doc

To Get access to entire book : Visit this page