

Hadoop and Spark Developer - CCA 175

CCA 175 Prep Plan

The key to any certification preparation is to have a proper plan, as the old saying says 'failing to plan is planning to fail'. In this blog I am going to show you one possible way you can prepare and obtain CCA 175 certification. My goal is to accomplish two things on this blog

1. Identify the technologies to learn in order to accomplish the certification goals
2. Create a realistic schedule that encompasses learning and practicing before appearing for the certification

Note: video walkthrough of this plan is available at [\[Click here\]](#)

[Click here for the video version of this series. This takes you to the youtube playlist of videos.](#)

Below table maps the required skill to technologies one needs to learn in order to solve problems during the certification exam. Remember, CCA 175 is a hands on exam, it is an open book exam but the only content you can access during the exam is api and official framework documentation. Hence, it is very important to gain a good level of comfort in using a set of hadoop eco system technologies, generic or specific frameworks and programming/query languages.

Skill Category	Skill Description	Technology To Use
Data Ingest	Import data from a MySQL database into HDFS using Sqoop	Sqoop
	Export data to a MySQL database from HDFS using Sqoop	Sqoop
	Change the delimiter and file format of data during import using Sqoop	Sqoop
	Ingest real-time and near-real-time streaming data into HDFS	Flume or Spark Streaming
	Process streaming data as it is loaded onto the cluster	Flume or Spark Streaming
Transform, Stage and Store	Load RDD data from HDFS for use in Spark applications	Spark RDD and Spark DF
	Write the results from an RDD back into HDFS using Spark	Spark RDD and Spark DF
	Read and write files in a variety of file formats	Spark RDD and Spark DF
	Perform standard extract, transform, load (ETL) processes on data	Spark RDD, Spark DF and Hive
	Use metastore tables as an input source or an output sink for Spark applications	Spark RDD, Spark DF, Spark SQL, Hive, Impala
Data Analysis	Understand the fundamentals of querying datasets in Spark	Spark RDD, Spark DF, Spark SQL
	Filter data using Spark	Spark RDD, Spark DF, Spark SQL
	Write queries that calculate aggregate statistics	Spark DF, Spark SQL, Hive and Impala
	Join disparate datasets using Spark	Spark RDD, Spark DF and Spark SQL
	Produce ranked or sorted data	Spark RDD, Spark DF, Spark SQL, Hive and Impala
Configuration	Supply command-line options to change your application configuration, such as increasing available memory	Spark Submit and options that can be used along with Spark Submit

This essentially boils down to learning below tools, frameworks, libraries and technologies. Here are the pre-requisites before you start your learning journey and also for practicing these technologies.

- Basic knowledge of any programming language. If you have scala or python background then it makes it much more easier
- Good Understanding of what data and database means. Some knowledge of SQL querying also helps.
- Finally, the most import aspect of this practical learning is to have an environment. it may take hours or days for you to build a haoop environment with all these combination of technologies. Cloudera makes it easier for you by providing a quickstart VM that you can install on your machine. Please read the instructions carefully and watch some youtube videos on how to setup the quickstart VM for your practice. You can download the quick start VM here [\[Click HERE\]](#)

#	Technology	Languages	Description
1	HDFS	Unix like commands	The Hadoop Distributed File System (HDFS) offers a way to store large files across multiple machines.
2	Sqoop	Unix like commands with some SQL	Framework for bulk data transfer between HDFS and structured datastores as RDBMS.
3	Spark	Scala OR Python	It is a Data analytics cluster computing framework. Spark fits into the Hadoop open-source community, building on top of the Hadoop Distributed File System (HDFS). To its credit, Spark provides an easier to use alternative to Hadoop MapReduce and offers performance up to 10 times faster than previous generation systems like Hadoop MapReduce for certain applications. Spark is a framework for writing fast, distributed programs. Spark solves similar problems as Hadoop MapReduce does but with a fast in-memory approach and a clean functional style API. For the certification exam, the emphasis is on Spark and not on tradition map reduce.
4	Spark RDD	Scala OR Python	RDD stands for Resilient Distributed Datasets (RDD) is a fundamental data structure of Spark. It is an immutable distributed collection of objects. Each dataset in RDD is divided into logical partitions, which may be computed on different nodes of the cluster.
5	Spark DF	Scala OR Python	A DataFrame is a distributed collection of data, which is organized into named columns. A DataFrame can be constructed from an array of different sources such as Hive tables, Structured Data files, external databases, or existing RDDs.
6	Spark Streaming	Scala OR Python	Spark Streaming is an extension of the core Spark API that enables scalable, high-throughput, fault-tolerant stream processing of live data streams. Spark Streaming provides a high-level abstraction called discretized stream or DStream, which represents a continuous stream of data
7	Spark SQL	Scala, Python and SQL	Spark SQL is a Spark module for structured data processing. Unlike the basic Spark RDD API, the interfaces provided by Spark SQL provide Spark with more

Search This Blog

CCA 175 Hadoop and Spark Developer Preparation

- [Home](#)
- [CCA 175 Prep Plan](#)
- [Problem Scenario 1](#)
- [Problem Scenario 2](#)
- [Problem Scenario 3](#)
- [Problem Scenario 4](#)
- [Problem Scenario 5 \[SQOOP\]](#)
- [Problem Scenario 6 \[Data Analysis\]](#)
- [Problem Scenario 7 \[FLUME\]](#)
- [File Formats](#)
- [Youtube Playlist](#)

A leader with a unique blend of deep technology expertise and strong management skill



Arun Kumar Pasuparthi
[G+](#) [Follow](#) [212](#)

[View my complete profile](#)

Report Abuse

Blog Archive

April 2017 (1)

			information about the structure of both the data and the computation being performed. Internally, Spark SQL uses this extra information to perform extra optimizations.
8	Spark Submit	Unix like commands	it is a mechanism to run spark programs as applications by supplying configurable parameters to optimize spark code execution
9	Flume	Unix like commands	Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data. It has a simple and flexible architecture based on streaming data flows. It is robust and fault tolerant with tunable reliability mechanisms and many failover and recovery mechanisms. It uses a simple extensible data model that allows for online analytic application.
10	Hive	SQL	Hive provides a SQL-like interface to data stored in HDP.
11	Spark Streaming	Scala and Python	Spark Streaming is an extension of the core Spark API that enables scalable, high-throughput, fault-tolerant stream processing of live data streams. Spark Streaming provides a high-level abstraction called discretized stream or DStream, which represents a continuous stream of data
12	Impala	SQL	Cloudera's very own hive like interface but uses its own engines instead of relying on spark or map reduce engine for processing.

Now that we have the technology to skill mapping, lets translate this into a work schedule. I am assuming that you can spend 2 hours a day for 5 to 6 days a week. Given that Big Data is a mesmerizing world, i will not be surprised if you spend more than 2 hours a day purely out of interest and curiosity to learn. Hence a 6 week preparation should be good enough to crack the certification. I personally know people who did this in 2 weeks and hence nothing is impossible. So, there are just 6 weeks between the current "you" (possibly a no one in the big data context) to being someone i.e certified hadoop and spark developer. **Are you up to the challenge?** If the technologist and the curious learner inside you is urging you to shout **'YES I AM UP TO THE CHALLENGE'** then i recommend that you either spend the next few weeks in equipping yourself in these technologies and come back to this blog when trying to accomplish the 15th task in the schedule below OR use the videos in this blog to gain some understanding in a more real time learning environment where you learn concepts on the fly and in a hands-on fashion.

This series of blog posts will provide a series of mock problem exercises you can solve to test your skill and knowledge required for clearing the certification exam. Please explore the links in the menu (right side) and also go through the videos in the playlist.

#	Task	Hrs of Study	Hrs of Practice
1	Setup Cloudera quickstart VM	3 Hours	NA
2	Introduction to Hadoop and Basic understanding of what Big Data is in general	3 Hours	NA
3	HDFS	1 Hour	2 Hours
4	Sqoop	2 Hours	2 Hours
5	Scala	3 Hours	3 Hours
6	Python	3 Hours	3 Hours
7	Spark RDD	3 Hours	6 Hours
8	Spark DF	1 Hours	2 Hours
9	Spark SQL	1 Hours	2 Hours
10	Spark Submit	1 Hours	1 Hour
11	Flume	1 Hours	3 Hours
12	Spark Streaming	1 Hours	1 Hour
13	Hive	3 Hours	5 Hours
14	Impala	1 Hours	2 Hours
15	Scenarios	NA	4 Hours
16	Total	27 Hours	39 Hours
17	Grant Total	66 Hours	
18	Total Weeks of Prep at 2 Hours a day and 5 days a week	Around 6 Weeks	

GOOD LUCK.....

73 comments:

- SV

May 8, 2017 at 1:04 PM

Thx for the helpful posts and the scenarios. Hopefully you will keep adding more content so users can clear the exam successfully!

Reply
- ▼ Replies

Arun Kumar Pasuparthi

May 8, 2017 at 2:05 PM

Sure more problem scenarios are coming shortly. Please go through what is available and provide your feedback. I am going to upload problem 5 by end of day today.

Anurag Edlabadkar

September 18, 2017 at 10:16 PM

Hello Arun, Excellent input. Activities are too good to add knowledge. Please add more problem scenarios if possible. Thanking you for your good inputs.

Reply
- Unknown

May 14, 2017 at 1:38 PM

thanks a lot the information is well organized and also very helpful...I will be waiting for your new problem scenarios....Keep up the good work!!!

Reply
- tympanine rhapsody

May 22, 2017 at 8:58 PM

Hi Arun,
Your roadmap is exemplary. Could you pls direct me to the right source to start preparation, ie, any study material or websites that will cover HDFS, Sqoop, Spark, Python, Impala, Hive and Flume, in line of your tasklist. Thanks much!

Reply
- Raphael L. Nascimento

May 30, 2017 at 1:26 PM

Hi Arun, are you planning post some problem related to the configuration topic of the exam ?
Best Regards Raphael

Reply

**amitrock** May 31, 2017 at 8:46 PM

Arun , I have one question. In CCA175 syllabus no where the term Data Frames mentioned. From which term we can say that Data Frame is included as part of exam ?

Reply

▼ Replies

**Arun Kumar Pasuparthi** June 1, 2017 at 6:30 AM

Cloudera doesn't necessarily list each individual feature of a tool. Data Frame is an important additional and you can expect questions on Data Frames in the Cloudera Exam. For exam, how would read an avro file without understand what a Data Frame is. `sqlContext.read.avro()` returns a Data Frame. You will need an understanding of Data Frames to deal with reading of Avro or Parquet or any data files that contain schema in it (like not text files). Also, it is quite normal for Cloudera to ask you questions where you need to complete a pre-existing incomplete code. That code can have Data Frame api calls. Hence, having a basic understanding of Data Frame Api helps tremendously.

**tympanine rhapsody** June 2, 2017 at 8:33 AM

Hi Arun, pls help with the study material in line with your 66 hours preparation plan.

**Arun Kumar Pasuparthi** June 4, 2017 at 5:09 PM

Given that the exam is hands on, I don't recommend a book as a study material. I recommend online videos and official documentation. There are several on youtube by topic. Based on the study plan presented, i recommend that you pick each topic and do individual research. These are very simple topics and mastering them for CCA 175 purpose should not be a difficult. All you need is an hour of two of dedication every day to write the code and test the result. Remember, 'writing the code' is most important part of the preparation, do you assume that you can tackle the exam questions without much practice. All the very best my friend.

Reply

**Sabyasachi Paul** June 8, 2017 at 4:25 AM

Thank you so much Arun for this useful detail. This would help a lot for the prep.

Reply

**Unknown** June 11, 2017 at 9:47 AM

great work

Reply

**Unknown** June 12, 2017 at 7:10 AM

Thaks a lot Arun.. Great work.. Your questions gave me confidence in attending the exam. i have cleared the exam today...

One quick question - i tried to do the Snappy compression for AVRO files by applying `sqlContext.setConf("spark.sql.avro.compression.codec","snappy")` .. but it not compressing the AVRO files. i am not able to see the extension as *.snappy.avro.

Reply

▼ Replies

**sakthi vinayagam** June 20, 2017 at 9:49 PM

Its by default compressed with snappy. Thats why you dont see any difference. Try uncompressing it then you will see the size increased.

Reply

**Kuldeep Gaur** June 15, 2017 at 11:21 AM

Great explanation Arun!!

I have been preparing for CCA175 but still something I was missing out and that was the solid plan. You broke down bigger problem in simpler ones, so it gave me sense of confidence to get ready for CCA175.

Thank you very much

Reply

**Tulasi Ram K** June 21, 2017 at 11:41 PM

Arun,

Could you please let me know which is best to practice hadoop. I am very good at theory part while coming to practice it is not possible.

Thank you very much in advance..

Reply

**SURJIT SINGH** July 11, 2017 at 11:26 PM

about api and documentation..how do we access these during exam?

Reply

**Raj The Ultimate Analyser** July 17, 2017 at 8:07 PM

First of all I would like to say thank you for putting such effort and trying to explain which will make people life happy.

I know this is too much to ask, but it would be great help if you can help us getting into correct sources for each topic. At least some links would be of great help.

Reply

**Karthika** July 18, 2017 at 11:24 AM

sqoop export is not working without --direct option for any other folder structure except Hadoop default folder .

```
hdfs dfs -ls practice/problem2/products/data
```

lists the data (data imported from mysql from different table with same schema in the same db).

Permissions all good. productsexp table created. Same export statement works fine with --direct option included. For every export I check if the data exists and mysql table has no records.

```
sqoop export --connect jdbc:mysql://localhost:3006/retail_db --username root --password cloudera --export-dir practice/problem2/products/data --table productsexp
```

```
17/07/17 08:20:38 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.10.0
INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
```

```
INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `products1` AS t LIMIT 1
INFO mapreduce.Job: map 0% reduce 0%
INFO mapreduce.Job: map 100% reduce 0%
INFO mapreduce.Job: Job job_1500302907744_0005 failed with state FAILED due to: Task failed task_1500302907744_0005_m_000000
Job failed as tasks failed. failedMaps:1 failedReduces:0
INFO mapreduce.Job: Counters: 9
Job Counters
Failed map tasks=1
Killed map tasks=2
58 WARN mapreduce.Counters: Group FileSystemCounters is deprecated. Use org.apache.hadoop.mapreduce.FileSystemCounter instead
58 INFO mapreduce.ExportJobBase: Transferred 0 bytes in 17.2694 seconds (0 bytes/sec)
58 INFO mapreduce.ExportJobBase: Exported 0 records.
58 ERROR tool.ExportTool: Error during export:
Export job failed!
=====
```

What am I doing wrong? Please suggest.

[Reply](#)

▼ Replies



chandru G July 19, 2017 at 6:09 AM

Try this.. If you are using CDH, I think the port should be 3306 instead of 3006 and the username as retail_dba
 sqoop export --connect jdbc:mysql://localhost:3306/retail_db --username retail_db --password cloudera --export-dir
 practice/problem2/products/data --table productsexp

I tried with this and it works fine.. describe the table on mysql and see if the schema matches with the data
 sqoop export --connect jdbc:mysql://localhost:3306/retail_db \
 --username retail_db \
 --password cloudera \
 --export-dir /user/cloudera/sqooptest \
 --table products_test1



Karthika July 19, 2017 at 7:56 AM

Sorry , port number was a typo. I used port 3306 not 3006. But I didn't get why username should be retail_db . mysql username is root not retail_db. Authentication would fail in this case.
 Thanks for the reply .

[Reply](#)



Raj The Ultimate Analyser July 20, 2017 at 9:37 AM

Any body can suggest good resources to go through. I really appreciate the analysis of preparation topics.
 But finding/following the correct resource is also key for success for certification. So seeking some help here.

[Reply](#)

▼ Replies



Arun Kumar Pasuparthi July 20, 2017 at 9:40 AM

<https://www.youtube.com/playlist?list=PLf0swTFhTl8q0x0V1E6We5zBQ9UazHFY0>

Go through the playlist above. my good friend durga has done an excellent job in teaching the concepts. the course is designed for CCA 175.



Raj The Ultimate Analyser July 21, 2017 at 12:20 PM

Thank you, I have seen those. Planning to go through them as well.

[Reply](#)



Jayadas Chelur July 21, 2017 at 5:59 AM

Thank you for this wonderful blog and list of videos for CCA 175. This has been of immense help to me in getting certified (Yes, July 20th 2017). I had gone through the entire video playlist and did all the problems multiple times. Initially I had to look up the solutions and later I was able to do it and was keeping an eye on the timing aspect (which helped me a lot in the actual exam). These are some of the takeaways from my experience with the CCA 175 exam

Timing is extremely important - I had barely 1 minute left (9 Problems)
 Make sure the internet connection is good - the browser hung at one point and I lost some time.
 If you are planning to take the test from an office environment, make sure the network is not private - I had to reschedule mine.
 The test cluster is SLOW - do not expect the same response from the virtual machine in a laptop.
 The fonts are small and it is quite a strain (especially if you wear eye glasses, like me!)
 As Arun mentioned multiple times, always COPY and PAST file names, directories etc.
 Have multiple terminals open for working in parallel - you will need it.

Once again, a huge thank you to Arun for this valuable resource. This has been a great help in clearing the CCA 175 exam.

[Reply](#)

▼ Replies



Raj The Ultimate Analyser July 21, 2017 at 12:18 PM

Congrats for the certification. In terms of preparing the topics can you suggest any resources.



Unknown July 28, 2017 at 2:36 AM

Congrats and thanks a bunch for posting ur experience.
 Would you be kind enough to share your experience with job hunt after the certification? What package can we expect for certified resource without any on the job hand-on?
 Regards,

[Reply](#)



taoufik elk July 24, 2017 at 7:57 AM

Hi Arun,

Thanks for your huge effort! i just passed my exam and got the certification (8/9 correct answers).

Good luck for you guys!

[Reply](#)

**Murali Rachakonda** July 25, 2017 at 6:34 AM

Hi Arun, Thanks for great blog. Quick question related to exam day Can we launch spark shell for scala as spark-shell or do we need to mention --master yarn 2

[Reply](#)**Arun Kumar Pasuparthi** July 25, 2017 at 6:55 AM

spark-shell just like you do on the Cloudera CDH.

[Reply](#)**Unknown** July 25, 2017 at 6:58 PM

Thank you for putting this up Arun!

[Reply](#)**Ravi Ranjan** July 25, 2017 at 7:34 PM

Explanation is awesome!!

Before starting preparation for CCA175 certification exam. I see on Cloudera that java is essential for this certification. You being an expert, want to know from you..how much java is necessary for this?

And how much following simplilearn course can actually benefit me?

<https://www.simplilearn.com/big-data-and-analytics/big-data-and-hadoop-training>

[Reply](#)**Arun Kumar Pasuparthi** July 25, 2017 at 7:42 PM

you dont have to learn java to be CCA 175 certified. But you need to learn Scala or Python or both. Also, i dont know about the simplilearn training so cannot comment.

however, Below link takes you to a free training videos provided by my very good friend and more importantly a very good person and technocrat. They are free and i have heard from atleast 100 ppl that following those videos and solving the problems in my blog series helped ppl to get atleast 90% in the exam. All the very best

https://www.youtube.com/watch?v=IRSpBce_T5Y&list=PLf0swTFhTl8rJvGpOp-LujOcpk-Rlz-yE

[Reply](#)**Ravi Ranjan** July 26, 2017 at 8:18 PM

Hello Arun,

Thanks for the information and the link. I am about to start watching videos.

I do have few more questions:

1) Does the video link "https://www.youtube.com/watch?v=IRSpBce_T5Y&list=PLf0swTFhTl8rJvGpOp-LujOcpk-Rlz-yE" also teaches me scala and/or python?

2) Does signing up on "labs.itversity.com" for big data cluster access would be enough?

I mean, how do I get access/install to Sqoop, Spark, Flume, Impala, Hive etc?

OR

is there any other cloud cluster where I should signup on?

Just wanted you to know that, in case if it is required to install any tool on my system. My system is Windows 10, with 6GB RAM. Will that be sufficient?

My questions could sound a bit dumb. But, being completely new in this area, I prefer asking silly questions.

Thanks for patiently reading and replying.

[Reply](#)**Srini** July 29, 2017 at 3:19 AM

Hi Arun,

Thank you very much the blog, it is very useful.

Hi taoufik elk,

First of all Congratulations on clearing the exam. A quick question on exam questions after the recent syllabus change, did you get any spark streaming questions in the exam? in the new syllabus they have mentioned, "Process streaming data as it is loaded onto the cluster" which requires spark streaming at least from this exam prospective.

Thanks in advance.

[Reply](#)**Anil** August 2, 2017 at 7:41 AM

Hi Arun,

Could you please confirm me, I have gone thru your all videos related the CCA 175 exam and it is very much useful for me.but I just want to know one thing, Do we need to write the exam in particular institute suggested by Cloudera or can we write in our own laptop?

[Reply](#)

▼ Replies

**Arun Kumar Pasuparthi** August 2, 2017 at 9:59 AM

write on your own laptop. you should be in a clear room with no one else around you and the laptop should have webcam on all the time as it is remotely proctored exam.

**Anil** August 2, 2017 at 10:02 AM

Thanks a lot Arun for your quick response.:)

**Anil** August 2, 2017 at 10:25 AM

Hi Arun one more help, If we use the personnal laptop can we use the internet to browse for searching anything if we didn't get any answer? or will it be tracked?

**Arun Kumar Pasuparthi** August 3, 2017 at 6:36 AM

you cannot browse and search for anything on your browser. you will login into a VM and the VM gives you access only to the official documentation via the links.

[Reply](#)



Pavan Raj August 3, 2017 at 7:49 AM

Hi Arun,
Thanks for helping the ppl and have huge patience to you.I have one concerns from my side.
1.After write the CCA 175 exam, how they do the validation? we should get the output or simply can write the scenarios ?
2. Whatever you provided 7 problem scenarios is enough to prepare for certification? is all above 7 scenarios covered?

[Reply](#)



Pavan Raj August 3, 2017 at 7:54 AM

or if above are not covered, Could you please share us full dumps of CCA 175.Because I comes from Mainframes and done the Hadoop and Spark course, currently I am not able to get the job due to not having java background. So I need to get the certificate and then will start my job trails.Literally certificate is most important for me. Please suggest me some guidellines how to proceed on this.

[Reply](#)



Prateck Jain August 18, 2017 at 12:53 AM

Hi Arun,

Thank you so much for this amazing Blog. I was able to complete my certification with a score of 9/10.

[Reply](#)



Unknown August 23, 2017 at 6:22 PM

Hi Arun,
Thank you so much for putting up an amazing blog about CCA175. It helped me so much in completing my certification.

[Reply](#)



Kanan August 28, 2017 at 9:55 AM

Hi Arun,
Thank you so much. Your preparation plan and sample questions helped me a lot in clearing CCA175 certification. Keep up the good work. Thanks Again!

Kanan

[Reply](#)



sathya September 1, 2017 at 10:33 PM

Superb explanation & it's too clear to understand the concept as well, keep sharing admin with some updated information with right examples.Keep update more posts.

[Hadoop Training in Chennai](#)

[Base SAS Training in Chennai](#)

[Reply](#)



akshay joshi September 12, 2017 at 7:43 PM

Thanks Arun....
This is very helpful for me...

[Reply](#)



shazil saadat September 20, 2017 at 9:07 AM

Hi Arun. Outstanding Work!
I am a newbie in this field of Big Data Science and considering take CCA exam to kick start my career in this awesome field.
I am not from a Computer Science background I was a Business Student did my bachelors in Marketing however I spent a few months learning a bit of programming in python and R. Data Cleaning, EDA, intermediate Machine Learning that sort of stuff. How much time would it take for a guy like me to pass this exam. waiting for you response Thanks.

[Reply](#)



Carlo Vaccarini September 20, 2017 at 11:51 PM

Hi,
I have a question on the exam day. Performing the exercises in scala, is necessary to package a jar or is sufficient execute the instructions on the spark-shell?

Thanks in advance

[Reply](#)



Fakruddin Y September 22, 2017 at 5:57 AM

Hi Arun,
Great blog. quick question on spark-submit topics. do you happen to know when you can put these topics in the blog. thanks in advance

[Reply](#)



pavan P September 22, 2017 at 6:19 AM

Thanks Arun...it helped me in clearing CCA 175

[Reply](#)



Unknown September 22, 2017 at 8:11 AM

Hi Arun,
Thank you so much for sharing such an excellent information for us(Beginners).
I have one Question. Please reply..
How we can import Blob data from a particular column, existing in a particular table with the help of SQOOP.
It will be a great help from your side if you can explain with the help of an example.

[Reply](#)

Bala October 2, 2017 at 5:25 AM



Hello Arun !

I've cleared CCA175 today. Without your blog I doubt if I would have made it with this ease. And equally good was the course content from Durga that you had pointed out. Thanks a ton for your wonderful blog content. You have done a great job and thanks again !

[Reply](#)

▼ Replies



Ananthu December 8, 2017 at 4:36 AM

Congratulation on clearing the exam.

Is it mandatory to know Python? Is it sufficient to answer in scala rather than in Python?
Is it feasible to try out the solution in spark-shell before submitting the solution?
And how do we submit the solution?
Thanks

[Reply](#)



Unknown October 3, 2017 at 7:46 PM

Arun can you please add link or url where we can learn this technology faster and easy way from certification point of view

[Reply](#)



Lucy singh October 31, 2017 at 5:34 AM

your post is very good & creativity you just have shared
[Big Data and Hadoop Administrator Certification Training](#)

[Reply](#)



Unknown November 4, 2017 at 10:48 AM

Hello All

I am preparing for CCA 175. How can I access Apache documentation during the exam. I don't remember all the options of sqoop command. Please let me know.

[Reply](#)



Anurag Edlabadkar November 15, 2017 at 8:58 PM

Pasuparthi, sir please add more videos of SPARK for CCA175

[Reply](#)



Indranil Roy November 17, 2017 at 8:12 AM

Hi Arun Sir -

I have a simple question if you can help

I have very basic idea and some hands on practice with the entire syllabus including scala RDD , DF etc . My concern is Python , I have already spent money to acquire the rest of the knowledge expect Python . Is Python really required for this exam ? If yes , can you share any video that will show me how to use Python for this test on a Cloudera VM . I already have Cloudera VM installed and I use it for Scala programming. My target is very precisely the CCA certification and I do not want to spend any more money or time on something that is not required. I guess I just need to know a few basic uses of Python working with Big data

[Reply](#)



Ganesh November 29, 2017 at 10:05 AM

Hi Arun, I have gone thru few of your responses. Its really good and usefull. I am preparing for CCA175 & Installed VM cloudera for practice (initial stage). - From exam perspective - Do I need to practice in cloud labs before to write exam. Please suggest.

[Reply](#)



Ramya Reddy November 30, 2017 at 12:24 AM

nice thank you for sharing
[Hadoop Admin Online Training Hyderabad](#)

[Reply](#)



Unknown December 2, 2017 at 7:22 PM

hi i am preparing for cca 175 exam can you share the sample questions to prepare for the certification

[Reply](#)



Thanuja Sri December 3, 2017 at 10:12 AM

Nice post about hadoop.are you looking for [hadoop online training](#) .

[Reply](#)



Shravan B M December 14, 2017 at 12:21 AM

Thank you so much for your blog Arun. This helped me a lot to clear the CCA examination. Your expertise has helped us the most.

[Reply](#)



christopher December 18, 2017 at 10:44 PM

If you face some problems about the preparation of CCA175 exam then don't be worry. Download [CCA175 Question Answers](#) from Dumps4Download.com and prepare your exam in a easy way. Dumps4Download.com is the best organization which provide updated and actual exam dumps. You can download any type of exam from this leading organization without any tension. They also gives you 3 months free updating of your exam. So don't waste your time and download your exam now.

[Reply](#)



Naveen Naidu December 19, 2017 at 6:09 AM

Can someone provide the valuable docs or links for CCA Spark Developer Certification. So that it will be more helpful. Thanks in advance.

[Reply](#)



GStreak December 19, 2017 at 11:13 AM

Let us say I get the CCA hadoop and spark certification , but please tell me what happens after 2 years? does it expire? what do I do then? All so I'm a java developer who has done a course on hadoop does the certification help me land a hadoop developer job?

Reply



Sirisha Rao December 20, 2017 at 1:47 AM

• I very much enjoyed this article. Nice article thanks for given this information. I hope it useful to many People**Hadoop admin Online Training Bangalore**

Reply



Michael Jhon December 20, 2017 at 11:35 PM

Realbraindumps is the main site that providing **CCA175 braindumps**. If you facing problem in cca175 exam preparation. Then Realbraindumps will help you to your exam preparation when will your preparation then you can pass the exam on the first attempt surely.

Reply



Thanuja Sri December 28, 2017 at 12:46 AM

Are you looking for **data science online training** institute ?

Reply



sanvi Janvi January 8, 2018 at 3:34 AM

It is nice blog Thank you porovide importent information and i am searching for same information to save my time.
Big Data Hadoop Online Training Bangalore

Reply



Vamshi Singh January 9, 2018 at 3:02 AM

Thanks for sharing this valuable information to our vision.
Hadoop Training in Gurgaon

Reply



sanvi Janvi January 9, 2018 at 4:05 AM

It is nice blog Thank you porovide importent information and i am searching for same information to save my time.**Hadoop Admin Online Training Bangalore**

Reply

[Home](#)

Subscribe to: [Posts \(Atom\)](#)

If you have landed on this page then you are most likely aspiring to learn Hadoop ecosystem of technologies and tools. Why not make you...

(no title)

If you have landed on this page then you are most likely aspiring to learn Hadoop ecosystem of technologies and tools. Why not make you...