Dell EMC E20-065

DELL EMC ADVANCED ANALYTICS SPECIALIST CERTIFICATION QUESTIONS & ANSWERS

Exam Summary – Syllabus – Questions

E20-065

Dell EMC Certified Specialist - Data Scientist - Advanced Analytics (DECS-DS)
60 Questions Exam – 63 % Cut Score – Duration of 90 minutes

www.AnalyticsExam.Com

Table of Contents

Know Your E20-065 Certification Well:

The E20-065 is best suitable for candidates who want to gain knowledge in the Dell EMC Data Science. Before you start your E20-065 preparation you may struggle to get all the crucial Dell EMC Advanced Analytics Specialist materials like E20-065 syllabus, sample questions, study guide.

But don't worry the E20-065 PDF is here to help you prepare in a stress free manner.

The PDF is a combination of all your queries like-

- What is in the E20-065 syllabus?
- How many questions are there in the E20-065 exam?
- Which Practice test would help me to pass the E20-065 exam at the first attempt?

Passing the E20-065 exam makes you Dell EMC Certified Specialist - Data Scientist - Advanced Analytics (DECS-DS). Having the Dell EMC Advanced Analytics Specialist certification opens multiple opportunities for you. You can grab a new job, get a higher salary or simply get recognition within your current organization.

E20-065 Dell EMC Advanced Analytics Specialist Certification Details:

Exam Name	Dell EMC Advanced Analytics Specialist for Data Scientists
Exam Code	E20-065
Exam Duration	90 minutes
Exam Questions	60 Questions
Passing Score	63
Exam Price	$230 (USD)
Training	**Advanced Methods in Data Science and Big Data Analytics (MR-1CP-ETAAMUSD)**
Exam Registration	**Pearson VUE**
Sample Questions	**Dell EMC Advanced Analytics Specialist Certification Sample Question**
Practice Exam	**Dell EMC Advanced Analytics Specialist Certification Practice Exam**

E20-065 Syllabus:

Objective	Details	Weight
MapReduce	- MapReduce framework and its implementation in Hadoop - Hadoop Distributed File System (HDFS) - Yet Another Resource Negotiator (YARN)	**15%**
Hadoop Ecosystem and NoSQL	- Pig - Hive - NoSQL - HBase - Spark	**15%**
Natural Language Processing (NLP)	- NLP and the four main categories of ambiguity - Text Preprocessing - Language Modeling	**20%**
Social Network Analysis (SNA)	- SNA and Graph Theory - Communities - Network Problems and SNA Tools	**23%**
Data Science Theory and Methods	- Simulation - Random Forests - Multinomial Logistic Regression and Maximum Entropy	**15%**
Data Visualization	- Perception and Visualization - Visualization of Multivariate Data	**12%**

Dell EMC E20-065 Sample Questions:

Question: 1

Which scenario would be ideal for processing Hadoop data with Hive?

a) Unstructured data; batch processing
b) Structured data, real-time processing
c) Structured data; batch processing
d) Unstructured data; real-time processing

Answer: a

Question: 2

What is a characteristic of stop words?

a) Meaningful words requiring a parser to stop and examine them
b) Don't occur often in text
c) Used in term frequency analysis
d) Include words such as "a", "an", and "the"

Answer: d

Question: 3

You develop a Python script "logisticpy" to evaluate the logistic function denoted as f(y) for a given value y that includes the following Pig code:

Register 'logistic.py' using jython as udf; z = FOREACH y GENERATE $0, udf.logistic ($0); DUMP z;

What is the expected output when the Pig code is executed?

a) 0
b) Jython is not a supported language
c) Value of f(y) for ally
d) Tuples (y, f(y))

Answer: d

Question: 4

Which problem type is best suited for simulation?

a) One with a few. non-random input variables
b) One that has a closed-form solution
c) One with numerous, non-random Input-variables
d) One that compares "what-if scenarios

Answer: d

Question: 5

You conduct a TFIDF analysis on 3 documents containing raw text and derive TFIDF ("data", document y) = 1.908. You know that the term "data" only appears in document 2.

What is the TF of "data" in document 2?

a) 2 based on the following reasoning: TFIDF = TF1DF = 1 908 You then know that IDF will equal LOG (32)=0.954 Therefore, TFIDF=TF*0.954 = 1.908 TF will then round to 2
b) 4 based on the following reasoning: TFIDF = TF1DF = 1.908 You then know that IDF will equal LOG (3/1)=0.477 Therefore, TFIDF=TF'0 477 = 1.908 TF will then round to 4
c) 6 based on the following reasoning: TFIDF = TF1DF = 1.908 You then know that IDF will equal 3/1=3 Therefore, TFIDF=TF/3 = 1.908 TF will then round to 6
d) 11 based on the following reasoning: TFIDF = TF1DF = 1908 You then know that IDF will equal LOG(3/2)=0.176 Therefore, TFIDF=TF"0.176 = 1.908 TF will then round to 11

Answer: b

Question: 6

You are analyzing written transcripts of focus groups conducted on product X. You approach is to use TF-IDF for your analysis.

What combination of TF-IDF scores should you examine to ensure you only report on the most important terms?

a) High TF score and high DF score
b) High TF score and high IDF score
c) High TF score and low IDF score
d) Low TF score and low DF score

Answer: c

Question: 7

What are key characteristics of Random Graphs?

a) Low clustering coefficients; high network diameters
b) Low clustering coefficients; small network diameters
c) High clustering coefficients; high network diameters
d) High clustering coefficients; small network diameters

Answer: b

Question: 8

According to Metcalfe's law, what is true about the value of a network?

a) Proportional to the number of edges
b) Proportional to the logarithm of the number of edges
c) Unrelated to the number of edges
d) Proportional to the square of the number of edges

Answer: c

Question: 9

Which scenario is a proper use case for multinomial logistic regression?

a) A marketing firm wants to estimate the personal income of a group of potential customers. Using inputs such as age, education, marital status, and credit card expenditures, a data scientist is building a model that will estimate a person's income
b) A logistic distribution company wants to minimize the distance traveled by its delivery trucks. A data scientist is building a model to determine the optimal route for each of tis trucks
c) To improve the initial routing of a loan application, a financial institution plans to classify a loan application as Approve, Reject, or Possibly_Approve. Based on the company's historical loan application data, a data scientist is building a model to assign one of these three outcomes to each submitted application.
d) A manufacturer plans to determine the optimal number of workers to employ in an assembly line process. Utilizing the observed distributions of the task durations of each process step, a data scientist is building a model to mimic the interactions and dependencies between each stage in the manufacturing process.

Answer: c

Question: 10

Why would a company decide to use HBase to replace an existing relational database?

a) It is required for performing ad-hoc queries.
b) Varying formats of input data requires columns to be added in real time.
c) The company's employees are already fluent in SQL.
d) Existing SQL code will run unchanged on HBase.

Answer: a

Study Guide to Crack Dell EMC Advanced Analytics Specialist E20-065 Exam:

- Getting details of the E20-065 syllabus, is the first step of a study plan. This pdf is going to be of ultimate help. Completion of the syllabus is must to pass the E20-065 exam.
- Making a schedule is vital. A structured method of preparation leads to success. A candidate must plan his schedule and follow it rigorously to attain success.
- Joining the Dell EMC provided training for E20-065 exam could be of much help. If there is specific training for the exam, you can discover it from the link above.
- Read from the E20-065 sample questions to gain your idea about the actual exam questions. In this PDF useful sample questions are provided to make your exam preparation easy.
- Practicing on E20-065 practice tests is must. Continuous practice will make you an expert in all syllabus areas.