



E20-065

**Dell EMC Advanced Analytics Specialist for Data
Scientists**

Exam Summary – Syllabus – Questions



Table of Contents

Introduction to E20-065 Exam on Dell EMC Advanced Analytics Specialist for Data Scientists	2
Dell EMC E20-065 Certification Details:	2
Dell EMC E20-065 Exam Syllabus:	3
E20-065 Sample Questions:	3
Answers to E20-065 Exam Questions:	6

Introduction to E20-065 Exam on Dell EMC Advanced Analytics Specialist for Data Scientists

This page is a one-stop solution for any information you may require for Dell EMC Advanced Analytics Specialist for Data Scientists (E20-065) Certification exam. The Dell EMC E20-065 Exam Summary, Syllabus Topics and Sample Questions provide the base for the actual Dell EMC Certified Specialist - Data Scientist - Advanced Analytics (DECS-DS) exam preparation, we have designed these resources to help you get ready to take your dream exam.

The Dell EMC Advanced Analytics Specialist for Data Scientists credential is globally recognized for validating Dell EMC Advanced Analytics Specialist knowledge. With the Dell EMC Certified Specialist - Data Scientist - Advanced Analytics (DECS-DS) Certification credential, you stand out in a crowd and prove that you have the Dell EMC Advanced Analytics Specialist knowledge to make a difference within your organization. The Dell EMC Advanced Analytics Specialist for Data Scientists Certification (E20-065) exam will test the candidate's knowledge on following areas.

Dell EMC E20-065 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	\$200 (USD)
Training	Advanced Methods in Data Science and Big Data Analytics
Books	
Exam Registration	Pearson VUE
Sample Questions	Dell EMC Advanced Analytics Specialist Certification Sample Question
Practice Exam	Dell EMC Advanced Analytics Specialist Certification Practice Exam

Dell EMC E20-065 Exam Syllabus:

MapReduce (15%)	<ul style="list-style-type: none"> - MapReduce framework and its implementation in Hadoop - Hadoop Distributed File System (HDFS) - Yet Another Resource Negotiator (YARN)
Hadoop Ecosystem and NoSQL (15%)	<ul style="list-style-type: none"> - Pig - Hive - NoSQL - HBase - Spark
Natural Language Processing (NLP) (20%)	<ul style="list-style-type: none"> - NLP and the four main categories of ambiguity - Text Preprocessing - Language Modeling
Social Network Analysis (SNA) (23%)	<ul style="list-style-type: none"> - SNA and Graph Theory - Communities - Network Problems and SNA Tools
Data Science Theory and Methods (15%)	<ul style="list-style-type: none"> - Simulation - Random Forests - Multinomial Logistic Regression and Maximum Entropy
Data Visualization (12%)	<ul style="list-style-type: none"> - Perception and Visualization - Visualization of Multivariate Data

E20-065 Sample Questions:

Q 1: 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?

Options:

- A. Tuples ($y, f(y)$)
- B. 0
- C. Jython is not a supported language
- D. Value of $f(y)$ for all y

Q 2: 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?

Options:

- A. 4 based on the following reasoning: $TFIDF = TF \cdot IDF = 1.908$ You then know that IDF will equal $\log(3/1) = 0.477$ Therefore, $TFIDF = TF \cdot 0.477 = 1.908$ TF will then round to 4
- B. 6 based on the following reasoning: $TFIDF = TF \cdot IDF = 1.908$ You then know that IDF will equal $3/1 = 3$ Therefore, $TFIDF = TF/3 = 1.908$ TF will then round to 6

C. 11 based on the following reasoning: $TFIDF = TF \cdot IDF = 1908$ You then know that IDF will equal $\text{LOG}(3/2)=0.176$ Therefore, $TFIDF=TF \cdot 0.176 = 1.908$ TF will then round to 11

D. 2 based on the following reasoning: $TFIDF = TF \cdot IDF = 1908$ You then know that IDF will equal $\text{LOG}(32)=0.954$ Therefore, $TFIDF=TF \cdot 0.954 = 1.908$ TF will then round to 2

Q 3: You are analyzing written transcripts of focus groups conducted on product X. Your 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?

Options:

- A. Low TF score and low DF score
- B. High TF score and low IDF score
- C. High TF score and high DF score
- D. High TF score and high IDF score

Q 4: Why would a company decide to use HBase to replace an existing relational database?

Options:

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

Q 5: Which scenario would be ideal for processing Hadoop data with Hive?

Options:

- A. Structured data; batch processing
- B. Structured data, real-time processing
- C. Unstructured data; batch processing
- D. Unstructured data; real-time processing

Q 6: Which scenario is a proper use case for multinomial logistic regression?

Options:

- 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 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.

C. 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 its trucks

D. 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.

Q 7: According to Metcalfe's law, what is true about the value of a network?

Options:

- 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

Q 8: What are key characteristics of Random Graphs?

Options:

- 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

Q 9: What is a characteristic of stop words?

Options:

- A. Used in term frequency analysis
- B. Include words such as "a", "an", and "the"
- C. Meaningful words requiring a parser to stop and examine them
- D. Don't occur often in text

Q 10: Which problem type is best suited for simulation?

Options:

- 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"

Answers to E20-065 Exam Questions:

Question: 1	Answer: A	Question: 2	Answer: A
Question: 3	Answer: B	Question: 4	Answer: B
Question: 5	Answer: C	Question: 6	Answer: D
Question: 7	Answer: C	Question: 8	Answer: B
Question: 9	Answer: B	Question: 10	Answer: D

Note: If you find any typo or data entry error in these sample questions, we request you to update us by commenting on this page or write an email on feedback@analyticsexam.com