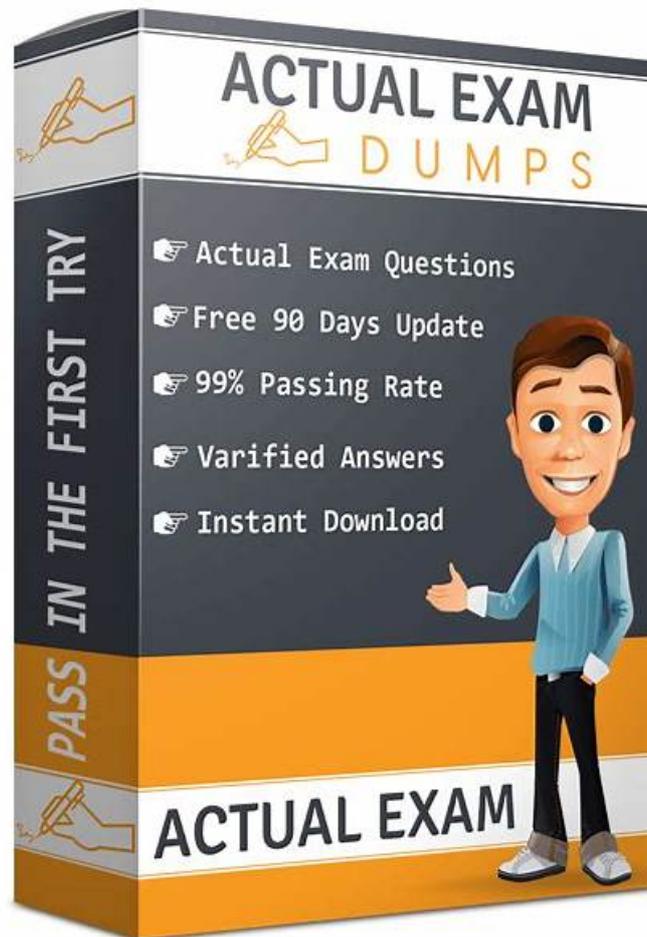


# SDS Valid Test Format - Real SDS Exam Dumps



BONUS!!! Download part of VCE4Dumps SDS dumps for free: <https://drive.google.com/open?id=1j96Ls91K04OSDccE8qaNXZarJIBsPB-o>

We are professional in this career to help all our worthy customers to obtain the SDS certification for years. You can get prepared with our SDS exam materials only for 20 to 30 hours before you go to attend your exam. we can claim that you will achieve guaranteed success with our SDS Study Guide for that our high pass rate is unmarshced 98% to 100%. And all the warm feedback from our clients proved our strength, you can totally rely on us with our SDS practice quiz!

We value every customer who purchases our SDS test material and we hope to continue our cooperation with you. Our SDS test questions are constantly being updated and improved so that you can get the information you need and get a better experience. The services provided by our SDS test questions are quite specific and comprehensive. First of all, our test material comes from many experts. The gold content of the materials is very high, and the updating speed is fast. By our SDS Exam Prep, you can find the most suitable information according to your own learning needs at any time, and make adjustments and perfect them at any time.

>> SDS Valid Test Format <<

## Real DASCA SDS Exam Dumps | SDS New Braindumps Ebook

Although it is not an easy thing for most people to pass the exam, therefore, they can provide you with efficient and convenience learning platform, so that you can obtain as many certificates as possible in the shortest time. We provide all candidates with SDS test torrent that is compiled by experts who have good knowledge of exam, and they are very experience in compile study materials. Not only that, our team checks the update every day, in order to keep the latest information of SDS latest question. Once we have latest version, we will send it to your mailbox as soon as possible.

## DASCA Senior Data Scientist Sample Questions (Q23-Q28):

### NEW QUESTION # 23

Which of the following statements is correct?

- A. Apache claimed that Spark is able to run parallel jobs 10 times faster in memory and 100 times faster on disk in comparison to the traditional Hadoop MapReduce
- **B. Apache claimed that Spark is able to run parallel jobs 100 times faster in memory and 10 times faster on disk in comparison to the traditional Hadoop MapReduce**
- C. Apache claimed that Spark is able to run parallel jobs 1000 times faster in memory and 100 times faster on disk in comparison to the traditional Hadoop MapReduce
- D. Apache claimed that Spark is able to run parallel jobs 50 times faster in memory and 5 times faster on disk in comparison to the traditional Hadoop MapReduce

**Answer: B**

Explanation:

Apache Spark is a distributed computing framework designed as an improvement over Hadoop's MapReduce.

According to the official Apache Spark documentation:

Spark can run workloads up to 100x faster in memory.

Spark can run workloads up to 10x faster on disk.

This performance gain comes from Spark's use of in-memory computation, DAG execution engine, and optimized query execution, compared to the slower, disk-heavy Hadoop MapReduce framework.

Thus, the correct statement is Option A.

Reference:

DASCA Data Scientist Knowledge Framework (DSKF) - Big Data Ecosystem: Spark vs Hadoop Performance Comparisons.

### NEW QUESTION # 24

Machine learning can be used in:

- A. Fraud detection
- B. Web search results
- C. Real-time ads on web pages and mobile devices
- **D. All of the above**
- E. Pattern and image recognition

**Answer: D**

Explanation:

Machine Learning has broad applications across industries and technologies:

Fraud Detection (Option A): Detecting anomalies in financial transactions, credit card usage, and cybersecurity threats.

Web Search Results (Option B): Ranking algorithms (e.g., Google's PageRank enhanced by ML techniques) improve relevance of search queries.

Real-time Ads (Option C): Online ad systems use reinforcement learning and recommendation models to target ads dynamically.

Pattern & Image Recognition (Option D): ML (especially deep learning) powers facial recognition, handwriting recognition, medical imaging, etc.

Since ML is used in all these applications, the correct answer is Option E (All of the above).

Reference:

DASCA Data Scientist Knowledge Framework (DSKF) - Applications of Machine Learning Across Domains.

### NEW QUESTION # 25

Maximum Likelihood Estimation (MLE) is a way to frame:

- A. Small class of problems in HDFS
- **B. Large class of problems in Data Science**
- C. Both A and C
- D. Large class of problems in HDFS
- E. Small class of problems in Data Science

**Answer: B**

Explanation:

Maximum Likelihood Estimation (MLE) is a statistical method used to estimate the parameters of a model by maximizing the likelihood function - i.e., finding the parameters that make the observed data most probable.

Option A: Correct. MLE provides a framework for a large class of problems in data science, including regression, classification, generative models, and probabilistic inference.

Option B: Incorrect - it applies to many problems, not just a small subset.

Option C & D: Incorrect. HDFS (Hadoop Distributed File System) is a storage technology, unrelated to MLE.

Option E: Incorrect because C is invalid.

Thus, the correct answer is Option A (Large class of problems in Data Science).

Reference:

DASCA Data Scientist Knowledge Framework (DSKF) - Statistical Foundations: Maximum Likelihood Estimation and Inference in Data Science.

### NEW QUESTION # 26

Spark is written in:

- A. C++
- B. Python
- C. Scala
- D. C
- E. Java

**Answer: C**

Explanation:

Apache Spark is an open-source distributed computing framework widely used for big data processing and machine learning pipelines.

The core implementation of Spark is written in Scala (Option A), which runs on the JVM (Java Virtual Machine).

Spark also provides APIs for Java, Python (PySpark), R, and SQL, but its native language is Scala.

Options C (C) and D (C++) are incorrect; Spark is not written in these languages.

Python (Option E) is a supported API, but Spark itself is not written in Python.

Thus, the correct answer is Scala (Option A).

Reference:

DASCA Data Scientist Knowledge Framework (DSKF) - Programming Tools for Big Data & Distributed Computing.

### NEW QUESTION # 27

Which of the following is an SLAs specification in case of Internet Service Provider (ISP)?

- A. Mean Time Between Failures (MTBF)
- B. Turnaround Time (TAT)
- C. Mean Time To Recovery (MTTR)
- D. All of the above

**Answer: D**

Explanation:

Service Level Agreements (SLAs) define performance commitments between service providers (like ISPs) and customers. Key SLA metrics include:

MTBF (Option A): Measures reliability by defining the expected average time between service failures.

MTTR (Option B): Measures availability by defining how quickly service can be restored after a failure.

TAT (Option C): Measures responsiveness in resolving customer requests or incidents.

All three are standard SLA performance specifications. Hence, the correct answer is Option D (All of the above).

Reference:

DASCA Data Scientist Knowledge Framework (DSKF) - Data Engineering and IT Practices: SLAs and Performance Metrics.

