

DASCA - SDS - Authoritative Test Senior Data Scientist Engine



DumpsValid is an excellent source of information on IT Certifications. In the DumpsValid, you can find study skills and learning materials for your exam. DumpsValid's DASCA SDS training materials are studied by the experienced IT experts. It has a strong accuracy and logic. To encounter DumpsValid, you will encounter the best training materials. You can rest assured that using our DASCA SDS Exam Training materials. With it, you have done fully prepared to meet this exam.

SDS Guide Torrent compiled by our company is definitely will be the most sensible choice for you. In this website, you can find three different versions of our SDS guide torrent which are prepared in order to cater to the different tastes of different people from different countries in the world since we are selling our Senior Data Scientist test torrent in the international market. Most notably, the simulation test is available in our software version. With the simulation test, all of our customers will have an access to get accustomed to the Senior Data Scientist exam atmosphere and get over all of bad habits which may influence your performance in the real Senior Data Scientist exam.

>> Test SDS Engine <<

Free PDF Quiz 2025 DASCA SDS: Marvelous Test Senior Data Scientist Engine

You can enjoy free update for 365 days if you buying SDS study guide of us, that is to say, in the following year you can obtain the latest information for the exam timely. And the update version for SDS exam dumps will be sent to your email automatically. You just need to receive and exchange your learning ways in accordance with the new changes. In addition, SDS Study Materials are

compiled by experienced experts, and they are quite familiar with the exam center, therefore the quality can be guaranteed. We also have online and offline chat service, if you have any questions about SDS exam dumps, you can consult us.

DASCA Senior Data Scientist Sample Questions (Q59-Q64):

NEW QUESTION # 59

Spark is written in:

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

Answer: A

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 # 60

Example of amortized performance is:

- A. HDFS dictionaries
- B. Python dictionaries
- C. All of the above
- D. Hadoop dictionaries
- E. MapReduce dictionaries

Answer: B

Explanation:

Amortized performance refers to averaging the cost of operations over a sequence of actions, ensuring that while some operations may be costly, the overall average time per operation remains efficient.

Python Dictionaries (Option B): Implemented using hash tables. Insertions, deletions, and lookups typically run in $O(1)$ average time, but occasionally require rehashing (costly). The high cost of rehashing is spread over many operations, giving amortized constant-time performance.

Option A (Hadoop dictionaries): Not standard terminology.

Option C (HDFS dictionaries): HDFS doesn't use dictionary structures in this sense.

Option D (MapReduce dictionaries): MapReduce uses key-value pairs, but amortized dictionary performance is not its focus.

Thus, the correct answer is Option B (Python dictionaries).

Reference:

DASCA Data Scientist Knowledge Framework (DSKF) - Programming for Data Science: Hash Tables & Amortized Analysis.

NEW QUESTION # 61

Which of the following is NOT a correct situation to use Agile?

- A. None of the above
- B. When changes need to be implemented during the entire process
- C. When clients/stakeholders need to be able to change the scope
- D. When the final product isn't clearly defined

Answer: A

Explanation:

Agile methodology is widely adopted in data science projects because these projects often involve uncertain goals, exploratory analysis, and changing requirements. Agile thrives in environments where iteration, collaboration, and adaptability are necessary.

Option A: True for Agile. If the final product is unclear (common in data science), Agile works well because it allows incremental discovery and iterative prototyping.

Option B: True for Agile. Agile frameworks (Scrum, Kanban) emphasize flexibility, which means the scope can evolve as stakeholders learn more from data and models.

Option C: True for Agile. Agile welcomes continuous changes through iterative sprints and feedback loops.

This adaptability is crucial in machine learning model development where data insights often reshape project direction.

Since all three situations are valid for Agile, the correct answer to "Which is NOT correct?" is None of the above (Option D).

Reference:

DASCA Data Scientist Knowledge Framework (DSKF) - Business Applications of Data Science & Agile Methodologies in Data Projects.

NEW QUESTION # 62

Data wrangling is the process of getting the data from:

- **A. Its raw format into something suitable for more conventional analytics**
- B. None of the above
- C. Both A and B
- D. Its modified meaning format into something suitable for more conventional analytics

Answer: A

Explanation:

Data wrangling (also called data munging) refers to transforming raw, messy, or unstructured data into a clean and structured format suitable for analysis.

Option A: Correct. Raw data often contains missing values, duplicates, or irregular formats. Wrangling prepares it for conventional analytics and machine learning.

Option B: Incorrect. Wrangling does not involve "modified meaning"; it focuses on cleaning, structuring, and integrating.

Option C: Incorrect, since only A is correct.

Option D: Incorrect, because wrangling is explicitly described in A.

Thus, the correct answer is Option A.

Reference:

DASCA Data Scientist Knowledge Framework (DSKF) - Data Engineering Practices: Data Wrangling & Preprocessing.

NEW QUESTION # 63

Semi-structured data does NOT include:

- A. Scientific data
- **B. Schema-full data**
- C. File systems
- D. Database system

Answer: B

Explanation:

Semi-structured data falls between structured data (e.g., relational databases with fixed schema) and unstructured data (e.g., free text, audio, video). It typically includes irregular or flexible schema information, such as XML, JSON, email data, or log files.

Option A (Database systems): Correct, databases may hold semi-structured content (e.g., JSON or XML columns).

Option B (File systems): Correct, file-based storage (logs, JSON, Avro, CSV) often contains semi-structured data.

Option C (Scientific data): Correct, many scientific applications generate semi-structured data formats (sensor readings, genomic sequences, etc.).

Option D (Schema-full data): Correct Answer. Schema-full (strict schema-defined relational tables) represent structured data, not semi-structured.

Thus, semi-structured data does NOT include schema-full data.

Reference:

