D-DS-OP-23 Dumps PDF 2024 Program Your Preparation EXAM SUCCESS [Q15-Q38
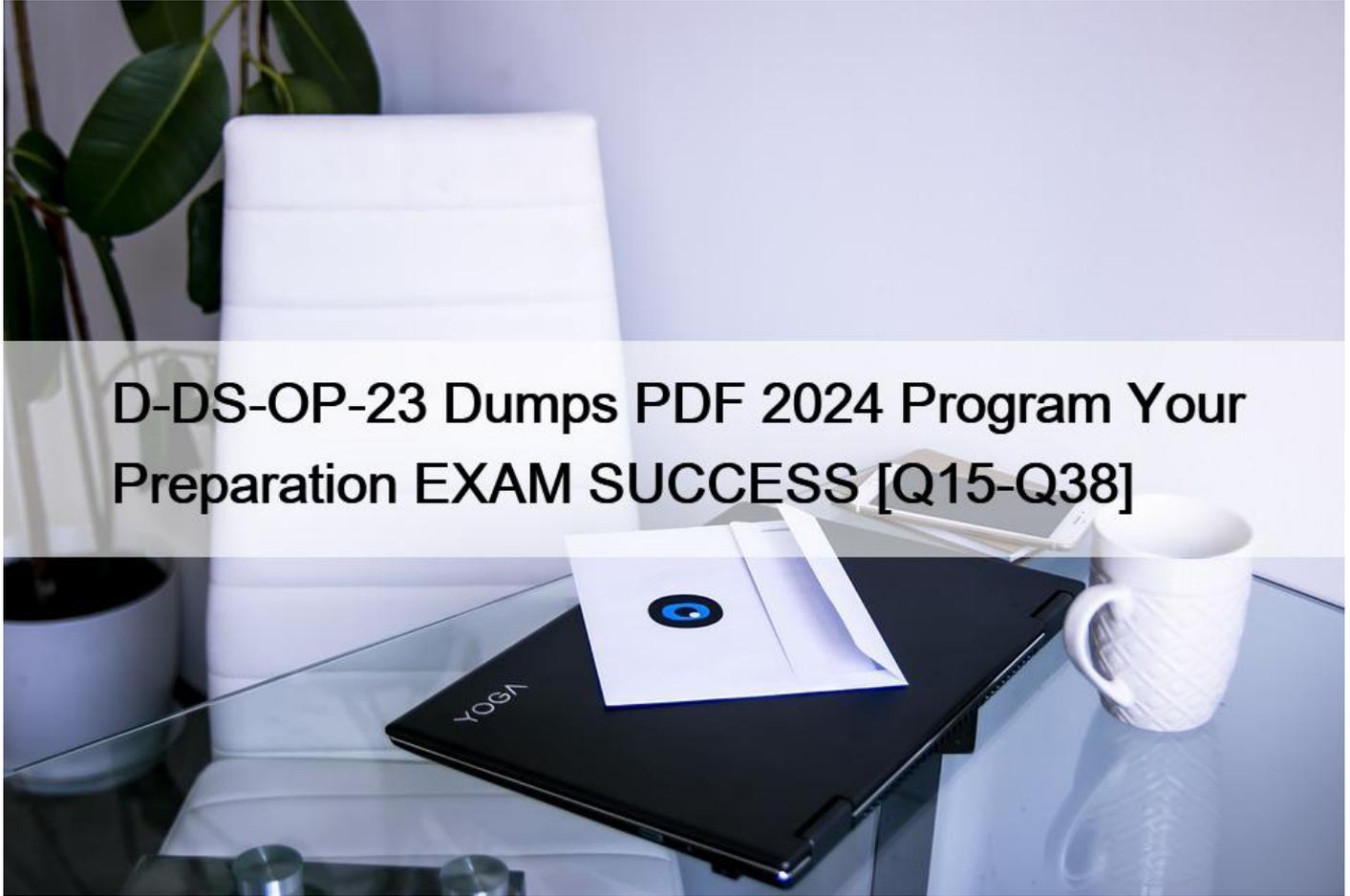

D-DS-OP-23 Dumps PDF 2024 Program Your Preparation EXAM SUCCESS Get Perfect Results with Premium D-DS-OP-23 Dumps Updated 103 Questions QUESTION 15

What is a characteristic of the Kerberos role in a Hadoop ecosystem?
* Bind-based authentication
* Secure user identification over insecure network
* Designed to access directory services
* Used for authentication and authorization

QUESTION 16

Which of the following are data ingestion tools commonly used in the Hadoop ecosystem? (Select all that apply)
* Apache Flume
* Apache Kafka
* Apache Sqoop
* Apache Flink
* Apache Beam

QUESTION 17

What ordered set of terms refers to the process in which data is sourced from a relational database, enriched and merged with other datasets, then stored in a data lake?

* Atomicity, Consistency, and Isolation
* Isolation, Atomicity, and Consistency
* Extract, Transform, and Load
* Extract, Load, and Transform

QUESTION 18

What ethical principle emphasizes that collected data should be used for its original intended purpose?

* Data minimization
* Purpose limitation
* Transparency
* Accountability

QUESTION 19

Which skill is crucial for a data engineer to effectively manage and optimize large-scale data processing systems?

* Data analysis
* Front-end development
* Cloud computing
* Graphic design

QUESTION 20

What is a data pipeline best practice related to Extract-Transform-Load (ETL) jobs?

* ETL jobs should be flexible to source data changes such as new variables or attributes.
* A failed ETL job should always be marked as completed so that any other scheduled jobs can execute.
* ETL jobs should be manually executed and monitored to ensure the highest quality.
* A failed ETL job should be re-run immediately without taking any corrective actions.

QUESTION 21

How is replication implemented in Redis?

* Client-server communication protocol is used, whereby every client replicates data from the server
* Data is shared among many computers to enable fault tolerance and data accessibility
* Meta information is generated based on the client-server communication protocol
* The developer copies and pastes the data in multiple databases

QUESTION 22

Which IoT data processing tool enables businesses to make real-time decisions to capitalize on the Analytics of Things (AoT)?

* Pravega
* Apache NiFi
* Nautilus
* Apache Kafka

QUESTION 23

Which of the following describes Apache Flink?
* A distributed file storage system
* A graph processing framework
* A stream processing framework
* A batch processing engine

QUESTION 24

What is the purpose of memory management in Apache Flink?
* Convert all of the data into Java objects
* Control how much memory the runtime operations use
* Eliminate the need for serialization of the data
* Ensure no disk space is ever required

QUESTION 25

What is a characteristic of Pig?
* Performs real-time reads and writes in HDFS
* Uses HiveQL to translate SQL queries into MapReduce jobs
* Data warehouse infrastructure that manages jobs in the cluster
* Alternative language to Java programming for MapReduce

QUESTION 26

What is metadata in the context of data governance?
* Data encryption keys
* Data quality metrics
* Data access logs
* Data about data

QUESTION 27

What is the data governance role that leverages the data for business purposes, such as increasing sales or reducing manufacturing costs?
* Data steward
* Data custodian
* Security specialist
* End user

QUESTION 28

Which of the following are components of the ETL process? (Select all that apply)
* Extraction
* Transmission
* Loading
* Transformation
* Visualization

QUESTION 29

In Python, which primitive data type covers real rational and irrational numbers?
* String
* Integer
* Float
* Boolean

QUESTION 30

An organization has implemented an IoT solution in their environment to gain competitive advantage.

This solution generates huge volumes of streaming data from different sources.

The organization wants to leverage this data to provide predictive, actionable results. They need a real- time data processing tool that can process the data using a fault-tolerant, horizontally scalable method.

Which tool should the organization use?
* Apache Ranger
* Apache Oozie
* Apache Storm
* Apache Ambari

QUESTION 31

Which Cassandra data type represents a collection of key-value pairs, and is represented in curly brackets ({, }) and colon (:)?
* List
* Map
* Set
* UUID

QUESTION 32

You are designing a database for an e-commerce data warehouse. The data is normalized in 1NF format and will be used to create data marts for different departments inside the organization.

Which schema type should be used for this requirement?
* Star
* Columnar
* Highly Normalized
* Graph

QUESTION 33

What is a primary task of a data engineer during a data analytics project?
* Installing, configuring, testing, and maintaining operating systems, application software, and system management tools
* Ensuring the successful initiation, planning, design, execution, monitoring, controlling, and closure of the project
* Designing and integrating infrastructure, analyzing algorithms, modifying source code, and writing system instructions
* Preparing the Big Data environment to be utilized by data scientists and other business stakeholders

QUESTION 34

In Python, which data type is one dimensional and immutable?
* List
* Array
* Tuple
* Set

QUESTION 35

What is the purpose of sensor operators in Apache Airflow?
* Perform validation checks in parallel
* Move data sequentially from one system to another
* Use triggers to report each successive retry
* Use a poke method to monitor external processes

QUESTION 36

What is the primary purpose of Apache Kafka in a data processing architecture?
* Storing historical data
* Running machine learning algorithms
* Processing real-time data streams
* Running complex SQL queries

QUESTION 37

Which of the following data structures in Python is immutable?
* List
* Dictionary
* Tuple
* Set

QUESTION 38

Why is Python commonly used for building data pipelines?
* It only supports batch processing
* It has no libraries for data manipulation
* Its readability and extensive libraries
* It requires compilation for execution

D-DS-OP-23 PDF Dumps Extremely Quick Way Of Preparation:
https://www.dumpsmaterials.com/D-DS-OP-23-real-torrent.html]