D-DS-OP-23

Dell Data Engineering Optimize 2023

Certification Description

Proven Professional Website

Engage with your peers in our Proven Professional Community

Certification Overview

This certification will benefit practicing or aspiring data engineers, data scientists, data stewards, or anyone who is responsible for managing or processing data sets.

Certification Requirements

To successfully complete this certification, a candidate must:

1. Have a sufficient knowledgebase/skill set through hands-on product experience and/or by consuming the recommended training.
2. Pass the Dell Data Engineering Optimize 2023 exam.

Note: These details reflect certification requirements as of February 3, 2024.

The Proven Professional Program periodically updates Certifications to reflect technical currency and relevance. Please check the Proven Professional website regularly for the latest information.

Dell Technologies Partners: Achieving a certification validates capability; however, it does not imply authorization to deliver services. Services Competencies provide partners with the ability to deliver services under their own brand or co-deliver with Dell Technologies. Tiered partners are eligible to obtain Services Competencies upon completing the specific requirements outlined in the Services Competencies Matrix. Only partners that have met these requirements should be delivering their own services in lieu of Dell Technologies Services.

Dell Technologies
1 Dell Way
Round Rock Texas 78682

DELL Technologies D-DS-OP-23

Exam Overview

This exam focuses on the role of a data engineer in successful analytic projects and the various tools and techniques including SQL, NoSQL, the Hadoop ecosystem, Apache Spark, data governance, streaming and IoT data processing, Python, and building data pipelines.

Exam Topics

Topics likely to be covered on this exam include:

The Role of the Data Engineer (5%)

- Describe the skills of a data engineer
- Describe the role of a data engineer in a data analytics project

Data Warehousing with SQL and NoSQL (17%)

- Describe characteristics and performance considerations of a relational database
- Describe relational database schemas and normalization techniques
- Describe use cases and features of various NoSQL tools

Extract-Transform-Load (ETL) Offload with Hadoop and Spark (18%)

- Describe ETL, ELT, and related schedulers
- Describe the Hadoop ecosystem, HDFS, and data ingestion tools
- Describe Apache Spark and its architecture

Data Governance, Security and Privacy for Big Data (20%)

- Describe data governance, key roles, and related models
- Describe metadata and Master Data Management
- Describe security considerations with Hadoop and the Cloud
- Describe the uses of Apache Atlas, Ranger, and Knox
- Describe privacy regulations and ethics

Processing Streaming and IoT Data (20%)

- Describe uses and application of IoT tools
- Describe the Apache Storm system and topology
- Describe the Apache Kafka queueing system and architecture
- Describe Apache Spark - Streaming processing and architecture
- Describe Apache Flink and its architecture
- Describe Pravega and its storage architecture
- Describe EdgeX Foundry and its architecture

Building Data Pipelines with Python (20%)

- Describe Python, reasons to use, and its libraries
- Describe the use of lists, dictionaries, tuples, sets, and strings
- Describe the use of Apache Airflow
- Describe data pipeline best practices

The percentages after each topic above reflects the approximate distribution of the total question set across the exam.

Duration

90 minutes

Recommended Training

The following training is recommended for candidates preparing to take this exam.

Please choose one of the options below.

Option 1: **Please complete the following courses**

Course Title	Course Number	Mode	Available
Introduction to Data Engineering	ES731OCMIDENG	On-demand	3/25/20
Data Warehousing with SQL and NoSQL - On-Demand Course	ES732OCMDATAW	On-demand	3/25/20
ETL Offload with Hadoop and Spark - On-Demand Course	ES732OCMETLHS	On-demand	3/25/20
Data Governance, Security and Privacy for Big Data - On-Demand Course	ES732OCMDGSPR	On-demand	3/25/20
Processing Streaming and IoT Data - On-Demand Course	ES732OCMPSIOT	On-demand	3/25/20
Building Data Pipelines with Python - On-Demand Course	ES732OCMBDPPY	On-demand	3/25/20

Option 2: **Please complete the following courses**

Course Title	Course Number	Mode	Available
Introduction to Data Engineering	ES731OCMIDENG	On-demand	3/25/20

Note: These exam description details reflect contents as of February 3, 2024

Copyright © 2024 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.

Dell Technologies believes the information in this document is accurate as of its publication date. The information is subject to change without notice.