BIG DATA PROFESSIONAL CERTIFICATE

CertiProf®

Syllabus V092021

Introduction

CertiProf® offers professional certification in Big Data, to understand the importance of data analysis and how they can obtain insights that lead to better decisions and strategic business movements.

Big Data refers to a set of data whose size (volume), complexity (variability) and growth rate (speed) make it difficult to capture, manage, process or analyze it using conventional technologies and tools, such as relational databases and conventional statistics. . The purpose of this certification is to teach professionals and organizations to identify problems in an understandable way using Big Data, to provide useful solutions with the large amount of information and data that can be shaped or tested in any way that is considered appropriate.

Objectives

- Understand Big Data and its powerful business benefits
- Know the differences between Big Data and conventional data
- Know the 4 V's of Big Data: Volume, speed, variety and veracity
- Discover the three types of Big Data business applications
- Learn valuable tips on how to manage Big Data
- Conceptualize the Big Data ecosystem and its key components
- Appreciate the main technological challenges in Big Data management
- Relate to the main technological solutions to face these challenges

Exam Format and Duration

This study program has an exam in which the candidate must achieve a score to obtain the certification in Big Data Professional Certificate BDPC™.

- Format: Multiple choice
- Questions: 40
- Language: English / Spanish / Portuguese
- Pass Score: 24/40 o 60 %
- Duration: 60 minutes
- Open book: No
- Delivery: This examination is available online
- Supervised: it will be at the discretion of the Partner

Eligibility for Certification

Anyone interested in expanding their knowledge in Big Data and who wants to cover the needs not met by existing technologies in storage and management of large volumes of data.

Content

Big Data and its Technologies
Plan
Introduction

Big Data - A Definition
Big Data - Definition

Big Data - Why?
Big Data - Why?
Data Availability
Structured vs Unstructured Data
Analytical Capacity
Big Data - Why?
Applications

Techniques and Technologies
Distributed File Systems
Scale up vs. Scale Out
HDFS Distributed File Systems
Distributed File Systems - S3
CEPH Distributed File Systems

Techniques - Distributed Algorithms
Techniques - Distributed Algorithms
MapReduce
Spark
Flux vs. Batch
Apache Storm
Apache Samza
Spark Streaming
DStream Transformations
Flux vs. Batch

Technologies - BD Systems
Database Systems
NoSQL

Document

Key-value

HBase

Cassandra

Graph

Structured

Kudu

CockroachDB

Others

Database Systems

Technologies - Orchestration Systems

Orchestration Systems

Dark Data

Dark Data

1. Traditional Database Management Systems

2. Traditional Data Security

3. How valuable is Dark Data?

4. Dark Data Characteristics

Conclusion

Practice1 - Batch processing with Hadoop HDFS and Map Reduce

Objectives of the Practical Work

Hadoop

Getting Started with Hadoop

Web Interfaces for Hadoop

Map Reduce

Wordcount

Launch Map Reduce on the Cluster

Practice 2 - Batch Processing and Streaming with Spark

Objectives of the Practical Work

Spark

Testing Spark with Spark-Shell

Spark API

Spark Batch in Java

Local Code Testing

Launching the Code on the Cluster

Spark Streaming

Practice3 - Data Collection with the Kafka Bus

Objectives of the Practical Work

Kafka

Architecture of Kafka

Kafka and Zookeeper

First Use of Kafka

Creating a Custom Application

Integrating Kafka with Spark

Practice4 - Data Storage in a NOSQL Database with HBase

Objectives of the Practical Work

Apache HBase

Data Model

Architecture

First Handling of HBase

HBase API

Loading Files

Data Processing with Spark