Certified Big Data Foundation

GSDC's Big Data Foundation Certification is aimed towards sharing a deep understanding of all the basic fundamentals of Big Data.

Certified Big Data Foundation

ABOUT CERTIFICATION

Big data is a field that treats ways to analyze, systematically extract information from, or otherwise, deal with data sets that are too large or complex to be dealt with by traditional data-processing application software.

Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy, and data source. Big data was originally associated with three key concepts: volume, variety, and velocity.

OBJECTIVES

Certified Big Data Foundation certification shares a deep understanding of:

1. Explain Big Data, its origin, and its characteristics.
2. Discuss the tools applicable to Big Data processing.
3. Explain data mining.
4. Discuss the popular Big Data technologies - Hadoop and MongoDB.
5. Discuss the Big Data projects and the main players involved.
6. Identify and obtain relevant datasets when looking at a business problem.
7. Install and manage Big Data processing environments based on Hadoop or MongoDB at a departmental level.

Our Accreditation:

ANSI
American National Standards Institute

ABICB

COURSE SYLLABUS

1. Big Data Introduction:
- Big Data - History, Overview, and Characteristics
- Definition
- Benefits
- Characteristics

2. Big Data Technology - Overview:
- Hadoop - Introduction, Usage, Concepts
- MongoDB - Introduction, Features, Concepts

3. Big Data - Privacy & Ethics:
- Privacy - Compliance
- Privacy - Challenges
- Privacy - Approach
- Ethics

4. Sources for Big Data:
- Enterprise Data Sources
- Enterprise Systems
- Oracle
- SAP
- Microsoft
- Data Warehouses
- Unstructured Data
- Metadata

5. Social Media Data Sources:
- Introduction
- Facebook - Introduction, Public Feed API, Keyword Insights API, Graph API
- Twitter - Introduction, Streaming APIs, REST APIs
- Other Social Media Sources

6. Public Data Sources:
- Introduction
- Weather
- Economics
- Finance
- Regulatory Bodies

7. Data Mining - Concepts and Tools:
- Data Mining - Introduction
- Types of Data Mining - Overview
- Classification
- Association
- Clustering
- Weka
- Modules of Weka Applications
- KNIME
- R Language

8. Big Data Technologies - Hadoop:
- Introduction
- Main Components of Hadoop
- Additional Components of Hadoop
- How to Install and Configure
- Map Reduce

9. Data Processing with Hadoop:
- Introduction
- Twitter Sentiment Analysis - Overview & Algorithm
- Network Log Analysis - Overview & Algorithm

10. Big Data Technologies - MongoDB:
- MongoDB Fundamentals
- Install & Configure
- Introduction
- Replication
- Sharding
- Sharding and Replication
- MongoDB Ecosystem - Languages and Drivers
- MongoDB Ecosystem - Hadoop Integration
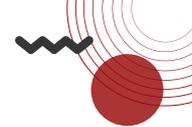- MongoDB Ecosystem - Tools

11. Document Databases:

- Introduction
- Documents
- Document Design Considerations
- Fields

12. Data Modelling with Document Databases:

- Introduction
- Twitter Sentiment Analysis with Algorithm
- Network Log Analysis with Algorithm

GSDC Technical Advisory Board :

The GSDC is the leading certification association which brings together innovative organizations and founding thought-leaders as Technical Advisors from over 40 countries to design curriculum on Blockchain, Devops, Six Sigma & Agile Certifications.

Our Future Information

Target Audience

- Software Engineers
- Application Developers
- Marketing Experts
- IT Architects
- Finance Specialists
- System Administrators
- Technical Recruiters
- Business Analysts
- Technical Support Specialists

Pre-requisites

There are no formal prerequisites for the exam, but knowledge of Hadoop and Mongo DB will help to understand the concepts

Find out more online at
www.gsdcouncil.org