Miami Dade College

Course Description

CAP4784 | Big Data | 4.00 credits

This course focuses on the processing of massive datasets, both structured and unstructured. Students will learn how to use Databricks and Spark to manage and analyze large datasets from a variety of sources. In addition, students will gain an understanding of how Databricks supports the end-to-end data science workflows that allow users to extract and share business insights. Prerequisites: CAP1788 and CAP2761C.

Course Competencies:

Competency 1: The student will demonstrate an understanding of Big Data concepts by:

1. Defining Big Data
2. Describing what Hadoop is and why it is important in managing Big Data
3. Describing Hadoop distributions
4. Explaining the components of the Hadoop Ecosystem and their functions
5. Discussing the future of Hadoop
6. Identifying elements of the Hadoop framework

Competency 2: The student will demonstrate an understanding of cloud computing by:

1. Defining the cloud
2. Comparing and contrasting Big Data cloud providers (i.e. Microsoft and Amazon)
3. Identifying cloud services
4. Discussing the key characteristics of cloud computing. 5. Describing service models. 6. Describing deployment models
5. Explaining cloud architecture. 8. Discussing issues of cloud security. 9. Discussing privacy issues in the cloud

Competency 3: The student will demonstrate how to set up Big Data in the cloud by:

1. Configuring a Big Data environment
2. Installing a Big Data environment
3. Loading files
4. Verifying two key components of the Hadoop ecosystem for processing large amounts of data, Hive and Pig

Competency 4: The student will demonstrate an understanding of how to store Big Data by:

1. Exploring the Hadoop Distributed File System (HDFS)
2. Explaining the HDFS architecture
3. Interacting with HDFS
4. Exploring the Big Data Warehouse
5. Designing, building, and loading tables in the cloud
6. Querying data.
7. Configuring the Hive Open Database Connectivity (ODCB) Driver

Competency 5: The student will demonstrate an understanding of how to manage Big Data by:

1. Providing structure for unstructured data
2. Enabling data access and transformation
3. Identifying Hive from traditional Relational Database Management Systems (RDBMS)
4. Creating and querying tables
5. Creating databases
6. Creating tables
7. Adding and deleting data
8. Querying a table
9. Using advanced data structures with Hive
10. Setting up partitioned tables

11. Loading partitioned tables
12. Using views to query data
13. Creating indexes for tables
14. Utilizing HDFS to store and manage Big Data

Competency 6: The student will demonstrate an understanding of how to work with Big Data by:

1. Moving data between Hadoop and relational databases
2. Integrating data
3. Importing and exporting data
4. Transforming data
5. Loading data into Hadoop

Competency 7: The student will demonstrate an understanding of how to work with Big Data by:

1. Organizing and formatting the data
2. Visualizing the data
3. Analyzing the data
4. Presenting conclusions and recommendations

Learning Outcomes:

- Communicate effectively using listening, speaking, reading, and writing skills
- Use quantitative analytical skills to evaluate and process numerical data
- Solve problems using critical and creative thinking and scientific reasoning
- Formulate strategies to locate, evaluate, and apply information
- Use computer and emerging technologies effectively
- Demonstrate an appreciation for aesthetics and creative activities