

Big Data with Spark and Hadoop

The term Big Data has been gaining popularity in recent years and has become a trend in the industry. The practical meaning of the term Big Data is technologies for analyzing and managing databases with unorganized data from many sources, in different quantities, and in varying qualities. Many solutions have been marketed to deal with this large amount of Data and many organizations find themselves preparing to enter this world. The leading solutions in the industry today is Spark and Hadoop, an efficient processing engine that on the one hand gives the organization flexibility in designing a solution tailored to the organization, and on the other hand is convenient and fast to use. In this course you will Learn the skills to use Apache Spark and Hadoop and practice the different components.

You Must Know!

Duration:

40 Hours

Who should attend?

DBA, BI Developers, Software\ Web Developers who wants to enter this world

Main Topics:

- Hadoop's Ecosystem
- Pig
- Hive
- Spark Programming Basics
- SQL with Spark
- Stream Processing Using Spark
- Data Science and Machine Learning Using Spark

המכללה שומרת לעצמה את הזכות לערוך מעת לעת, לפי שיקול דעתה, שינויים בתכנית הלימודים, היקף שעות הלימוד, סגל המדריכים וכד', ולא יראו בכל מידע המפורט בדפי מידע של המכללה כהתחייבות כלשהי מצד המכללה.

Course modules

Module 1 – Introducing Big Data, Hadoop, and Spark

- Introduction to Big Data, Distributed Computing
- Introduction to Hadoop
- Introduction to Apache Spark
- Deployment options

Module 2 – Hadoop's Ecosystem

- Understand the main Hadoop components
- Learn how HDFS works

Module 3 – Map Reduce and Pig

- Introduction to Map Reduce
- Map Reduce Programming
- Pig Basics
- Pig Relational Operators
- Pig Evaluation Functions and Other Commands

Module 4 – Hive

- Introduction to Hive
- Hive DDL Commands
- Hive DML Commands
- Hive Operators and Functions
- Hive Storage Formats
- Hive Configuration

Module 5 – Spark Programming Basics

- Input/Output Types for Spark Applications
- Introduction to Spark RDDs
- Loading Data into RDDs
- Methods for Creating RDDs
- RDD Transformations and Actions
- Pair RDDs
- Partitioning Data in Spark
- RDD Caching
- Persisting RDDs
- Optimizing Spark

המכללה שומרת לעצמה את הזכות לערוך מעת לעת, לפי שיקול דעתה, שינויים בתכנית הלימודים, היקף שעות הלימוד, סגל המדריכים וכד', ולא יראו בכל מידע המפורט בדפי מידע של המכללה כהתחייבות כלשהי מצד המכללה.

Module 6 – SQL with Spark

- Introduction to Spark SQL
- DataFrames
- Caching, Persisting, and Repartitioning DataFrames
- Saving DataFrame Output

Module 7 – Stream Processing Using Spark

- Spark Streaming
- DStreams
- State and Sliding Window Operations
- Using Spark with Messaging Platforms
- Apache Kafka



המרכז הבינלאומי
ללימודי הייטק וחדשנות

* 6377

מתקדמים
לקריירה בהייטק



Microsoft Partner
Gold Learning



GLOBAL
UNIVERSITY
SYSTEMS



ibat
COLLEGE
DUBLIN

HIGHQ
מכנים אותך לעולם האמיתי



ARDEN
UNIVERSITY

The University of
Law

London
School of Business
& Finance



HTK



TORONTO
SCHOOL OF MANAGEMENT

קמפוסים בפריסה ארצית:

באר שבע

רחוב האנגריה 77
פארק ההייטק

ירושלים

רחוב יפו 34

רחובות

רחוב אופנהיימר 5
פארק המדע

תל אביב

ראול ולנברג 36
קריית עתידים