



**Data
science**

BIG DATA

תיאור הקורס

המונח Big Data הולך וצובר תאוצה בשנים האחרונות, הוא הופך למגמה רווחת בתעשייה. המשמעות הפרקטית של המונח Big Data הינה טכנולוגיות ניתוח וניהול מאגרי מידע בעלי נתונים שאינם מאורגנים, המגיעים ממקורות רבים, בכמויות שונות, ובאיכויות משתנות. כדי לטפל בכמות גדולה זו שווקו פתרונות רבים וארגונים רבים מוצאים עצמם נערכים לכניסה לתוך עולם זה.

הקורס מהווה כמבוא לעולם ה - Big Data. נלמד להכיר Hadoop על רכיביו, ונבצע סקירה של הכלים הקיימים בשוק על מנת להבין מה מסתתר מאחורי 2 המילים הכל כך מדוברות: Big Data.

מטרת הקורס

- הקורס חושף את משתתפיו לתחום החדש הנקרא Big Data
- הקורס יעשה סדר בעולם המושגים הקיים בשוק בתחום חדש זה
- הקורס כולל תרגול מעשי של חלק מהחומר הנלמד

היקף הקורס

40 שעות אקדמיות

אוכלוסיית יעד

אנשי BI, DBA, אנשי פיתוח המעוניינים להיכנס לתחום.

מתודולוגית הלמידה

הרצאה פרונטאלית, הדגמות וסימולציות בסביבת הלמידה ותרגול אישי בליווי ספרות מקצועית



סגל המרצים

למכללת iNT טכנולוגיות סגל מרצים ומומחי הדרכה, מהמובילים בתחומם, בעלי ניסיון מעשי רב ביישום והדרכת נושאי הלימוד בתעשיית ההי-טק הישראלית והעולמית.

זכאות לתעודת גמר מטעם מכללת iNT

תעודת גמר מטעם מכללת iNT תוענק לבוגרים העומדים בכל דרישות התוכנית כמפורט:

1. נוכחות ב 85% מהשיעורים לפחות.
2. הגשת משימות/תרגילים בהתאם לדרישות הקורס
3. עמידה בתקנון הלימודים



תוכנית לימודים מפורטת

☑ Introduction to Big Data

- Understand what Big Data is (and we will review and clarify the Big Data buzzwords)
- Review the existent technologies (MPP, Hadoop, Columnar Database)
- Understand how Big Data solutions can work on the Cloud
- Learn what NoSQL is all about

☑ Introduction to Hadoop

- Understand what Hadoop is
- Learn about other open source software related to Hadoop

☑ Hadoop architecture

- Understand the main Hadoop components
- Learn how HDFS works
- List data access patterns for which HDFS is designed
- Describe how data is stored in an HDFS cluster

☑ Hadoop Administration

- Add and remove nodes from a cluster
- Verify the health of a clusterStart and stop a clusters components
- Modify Hadoop configuration parameters
- Setup a rack topology



☑ **Hadoop Components**

- Describe the MapReduce philosophy
 - Understand the concepts of map and reduce operations
 - Describe how Hadoop executes a MapReduce job
 - List MapReduce fault tolerance and scheduling features
 - List MapReduce fundamental data types
 - Describe a MapReduce data flow
- Explain how Pig, Hive, and Jaql can be used in a Hadoop environment to query data
- Describe how Flume and Sqoop can be used to move data into Hadoop
- Describe how Oozie is used to schedule and control Hadoop job execution

☑ **Map Reduce Programing**

- Describe the term map in regard to Hadoop
- Explain the term reduce in regard to Hadoop
- Describe how the JobTracker and TaskTrackers work with MapReduce
- Explain the fault tolerance capability of MapReduce
- Describe the basic Java code required to code
 - The mapper class
 - The reducer class
 - The driver
- Code a simple MapReduce application

☑ **Pig**

- Pig Basics
- Pig Relational Operators
- Pig Evaluation Functions and Other Commands



✓ SQL Access for Hadoop

- Introduction to Hive, Big SQL and Impala
- Working with SQL using Hive
- Working with SQL using Big SQL
- Accessing Hbase with Hive and Big SQL
- System tables and troubleshooting

✓ Hive

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

✓ Moving Data to Hadoop

- Load Scenarios
- Using Sqoop
- Using Flume

✓ Controlling Hadoop Jobs Using Oozie

- Introduction to Oozie Workflows
- Oozie Coordinator

לתשומת לב ולמען הסר ספק, מובהר כי 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.