

## Big-Data Architects

### מספר קורס 9293 - 125 שעות

המונח **Big-Data** מתייחס בד"כ למאגרי מידע שעונים לאחת או יותר מן הדרישות הבאות:

- המידע מורכב מנתונים בנפחים גדולים
- מידע חדש נוצר ונכנס או יוצא מהמערכת במהירות רבה
- כמות סוגי המידע הנאספים ומספר המקורות הם רבים ומגוונים

#### רציונל הקורס

ארגונים השוקלים לעבור לעולם ה-*Big-Data* זקוקים לידע נרחב החולש על מספר פלטפורמות, ספקי שירותים, עננים וטכנולוגיות. פערי הידע אינם מתנקזים לכדי נושא אחד אלא מתפרשים על פני כמה תחומים שכל אחד מהם מהווה עולם בפני עצמו. לשם כך נדרשת פונקציה אשר תדע ותכיר את הנמצא בעולם ה-*Data* וגם תוכל לאפיין אילו רכיבים מתוך הפיתוח הקיים או העתידי שייכים לעולם זה – *Big-Data Architect*.

אלו האתגרים העיקריים הניצבים כיום בפני ארגונים השוקלים או נמצאים בתהליך של מעבר לעולם ה-*Big-Data*:

- אפיון החלקים הקיימים במערכת אשר צריכים להיות מוסבות לפלטפורמת *Big-Data*
- אפיון רכיבים חדשים המשתייכים לעולם זה לעומת אילו היכולים להיעזר בסביבות המסורתיות
- איתור טכנולוגיות המתאימות לניהול התשתית ואנליזת הנתונים בארגון

בארגונים גדולים וקטנים כאחד מתגלה צורך בהכרה והבנה יסודית של אתגרים אלו. ארכיטקט בעל יכולת לשלב ידע רחבי אודות פלטפורמות, טכנולוגיות, מוצרים ושירותים בעולם ה-*Big-Data* יוכל לתכנן וליישם ארכיטקטורת מעבר לתחום זה באופן היעיל ביותר והנכון ביותר עבור החברה.

#### אודות הקורס

קורס *Big-Data Architects* מכין את בוגריו לקראת השתלבות בתעשיית ה- *Big Data* ומקנה ידע נרחב בטכנולוגיות המובילות בשוק. כמו כן, קורס זה שם דגש רב על הכנה מעשית, ומשלב תרגולים רבים. במהלך השיעורים יחשפו התלמידים למאגרי מידע רבים ומגוונים ויתנסו במבני *Data* שונים.

קורס *Big-Data Architects* מקנה בין היתר:

- ידע מעמיק בארכיטקטורת **Hadoop** עם דגש על הפצות **Hortonworks** ו-**Cloudera**.  
כמחצית משיעורי ה- **Hadoop** יוקדשו לסביבת **Cloudera** המחצית השנייה ל-**Hortonworks**
- אבטחת מידע באמצעות כלים כגון **Kerberos** ו-**Sentry**
- ניתוח נתוני *Big Data* באמצעות **Hive**, **Pig**, **Impala** ו-**Hbase**
- תזמון משימות באמצעות **Oozie**
- שינוע נתונים באמצעות **Sqoop**, **Flume** ו-**Kafka**
- פיתוח אפליקציות **Spark** באמצעות **Python**
- חישובים סטטיסטיים באמצעות **R**

- יישומים מתקדמים לניתוח מידע באמצעות SQL
- היכרות עם טכנולוגיות NoSQL נוספות כגון MongoDB, ו-Neo4j

קורס זה מלמד את כל העקרונות החשובים בטכנולוגיות ה- Big Data ובנוסף מקנה הבנה בטכנולוגיות שונות, המאפשרות "לראות את התמונה הגדולה". כך, אדם העושה את צעדיו הראשונים בעולם ה- Big Data נשאר בקצב ההתקדמות של הטכנולוגיה, פותח לעצמו דלתות רבות יותר, ומבסס את עצמו בארגון בו הוא נמצא על הצד הטוב ביותר.

במהלך הקורס ינתנו מספר פרויקטים אשר יאפשרו הטמעה מעשית, ואינטגרציה בין טכנולוגיות שונות אשר נלמדו.

### ליווי מקצועי

בקורס מרצים מיטב אנשי המקצוע בתחום. כל המדריכים בקורס הינם בעלי רקע בעבודה בשטח ובהדרכה בתחום, כולל מרצים ממיטב חברות הייעוץ בארץ.

### קהל יעד

קורס זה נועד בעיקרו למנהלי מסדי נתונים, אנשי BI, כותבי קוד, ראשי צוותים, מנהלי פיתוח, ואנשי מקצוע נוספים אשר עוסקים בתחום ה- Big-Data.

קורס זה מעניק ידע רב, ומתנהל בקצב למידה גבוה. נדרשת יכולת הבנה גבוהה, למידה עצמאית, והתמצאות טכנית.

### תנאי קדם:

- ידע וניסיון בעבודה מול מערכות Data שונות
- רקע בשפת ה- SQL
- ראיון אישי / ייעוץ להכוונה מקצועית

### משך הקורס

- 125 שעות
- לימודי ערב מ- 17:30 עד 21:45, שני מפגשים שבועיים
- הלימודים אינם מתקיימים בחגים ומועדים.

### זכאות לתעודת גמר

- נוכחות ב- 80% מן המפגשים לפחות
- הגשת פרויקטים בהצלחה

Module Title	Module Description	Hours
Preparatory	<ul style="list-style-type: none"> <li>• Make yourself comfortable with the Linux OS</li> <li>• Understand basic concepts of virtual machines</li> <li>• Create and run virtual machines in the cloud</li> </ul>	10
Introduction to Big Data	<ul style="list-style-type: none"> <li>• Our roadmap</li> <li>• Big data basic concepts</li> <li>• Hadoop basics</li> </ul>	5
Hadoop Eco System Basics	<ul style="list-style-type: none"> <li>• HDFS, YARN and MapReduce (5) <ul style="list-style-type: none"> <li>○ Basic HDFS concepts</li> <li>○ Working with the HDFS</li> <li>○ Yarn components</li> <li>○ MapReduce workflow</li> </ul> </li> <li>• Securing the Cluster Using Kerberos and Sentry (5) <ul style="list-style-type: none"> <li>○ Hadoop's security system concepts</li> <li>○ What is kerberos and how it works</li> <li>○ Securing the cluster using kerberos</li> <li>○ Sentry basic concepts</li> </ul> </li> <li>• Analyzing Data using Apache Hive and Impala (10) <ul style="list-style-type: none"> <li>○ Hive and Impala - Introduction and use-cases</li> <li>○ Data types</li> <li>○ Creating databases, tables, and partitions</li> </ul> </li> <li>• Working with Apache Pig (5) <ul style="list-style-type: none"> <li>○ Pig - Introduction and use-cases</li> <li>○ Pig's data model</li> <li>○ Pig Latin</li> </ul> </li> <li>• Working with HBase (5) <ul style="list-style-type: none"> <li>○ Introduction to Hbase for processing huge tables</li> <li>○ Hbase data model</li> <li>○ Hbase vs. RDBMS</li> <li>○ Client API (CRUD, queries and batch operations)</li> <li>○ Interactive REST clients</li> </ul> </li> </ul>	30
Workflow and Automation	<ul style="list-style-type: none"> <li>• Introduction to Oozie</li> <li>• Workflow scheduler</li> <li>• Alerts and Notifications</li> </ul>	5
Transporting Data	<ul style="list-style-type: none"> <li>• Working with Flume</li> <li>• Kafka basic concepts</li> <li>• Using Sqoop</li> </ul>	10
Spark	<ul style="list-style-type: none"> <li>• Introduction to Python (10)</li> <li>• Working with PySpark (25) <ul style="list-style-type: none"> <li>○ Basics</li> <li>○ RDD Operations</li> </ul> </li> </ul>	35

Module Title	Module Description	Hours
	<ul style="list-style-type: none"> <li>○ Caching</li> <li>○ Spark Streaming</li> <li>○ Spark SQL and DataFrames</li> </ul>	
Statistical programming Using R	<ul style="list-style-type: none"> <li>● What is statistical programming</li> <li>● The need for Big-Data</li> <li>● R statistical language</li> <li>● introduction to R</li> <li>● installation</li> <li>● Basic usage &amp; general statistics with R</li> <li>● Data mining using R</li> <li>● Hadoop-R</li> </ul>	10
Advanced SQL	<ul style="list-style-type: none"> <li>● This module, unique of its kind in Israel, focuses on complex queries, and entirely composed of advanced “real-world” problems and solutions.</li> <li>● This training is based on a decade of experience working with various databases, challenging job interviews questions, and many complex SQL problems. In this course, students learn many different practical applications and “real-world” problems.</li> </ul>	10
Working with Neo4j	<ul style="list-style-type: none"> <li>● Graph database concepts</li> <li>● Basic querying - Nodes and Relationships</li> </ul>	5
Working with MongoDB	<ul style="list-style-type: none"> <li>● Introduction to MongoDB</li> <li>● Core concepts</li> <li>● Environments</li> <li>● CRUD and the MongoDB Shell</li> </ul>	5
<b>Total</b>		<b>125</b>