

ניהול פרויקטים וניתוח מערכות בעולם ה **BIG DATA & Machine Learning & No Sql db**

קורס 24256 – 40 שעות

אודות הקורס

ארגונים רבים שוקלים כיום (או שכבר עברו) לעבור לעולם נתוני עתק (ה-Big-Data), במסגרת נסיון זה הם נתקלים במושגים חדשים רבים, מחזורי חיים חדשים, אינספור פלטפורמות חדשות, הגדרות חדשות של בעלי תפקיד, ספקי שירותים מסוגים חדשים וטכנולוגיות חדשות. פערי הידע של הארגונים אינם מתנקזים לכדי נושא אחד אלא מתפרשים על פני כמה תחומים שכל אחד מהם הוא עולם בפני עצמו. לשם כך על המנהלים, מנהלי הפרויקט/ מנתחי המערכות וראשי הצוותים ואנשי צוות רבים נוספים נדרשים לדעת ולהכיר עולם זה, בשלב הראשון יש צורך להבין את היתרונות העסקיים הטמונים בעולם זה, להכיר ברמה גבוהה את המרכיבים של עולם זה. מידע זה נדרש כיום לכל מי שעוסק בתחום המחשוב הארגוני

האתגרים העיקריים הניצבים כיום בפני ארגונים השוקלים או נמצאים בתהליך של מעבר והבנת והערך העסקי בעולם Big data

- בראש ובראשונה ישנם מספר נקודות המעסיקות את הארגון: האם המעבר מתאים לארגון שלנו ועוד יותר משמעותי – מה יצא לי מזה, ובנוסף כמה זמן יקח, השקעה, משאבים וכדומה
- בקורס זה נעסוק בשאלות אלו, ובנוסף:
 - ✓ הבנה מה הן מושגים והמונחים המרכיבים את עולם ה Big-Data
 - ✓ מהו מחזור חיים המתאים לעולמות תוכן אלו
 - ✓ מהי חלוקת התפקידים בין אנשי הצוות: מנהל הפרויקט, מנתח המערכות, המתכנת, איש הבדיקות ועוד.
 - ✓ אפיון הרכיבים המשתייכים לעולם זה לעומת אילו היכולים להיעזר בסביבות המסורתיות
 - ✓ איתור, איסוף וניקוי המידע
 - ✓ אחסון, המידע הרב והלא מובנה בבסיסי נתונים מהסוג החדש
 - ✓ ניתוח ועיבוד מידע לא-מובנה ורב (non-structured data)
 - ✓ הצגתו בכלים פשוטים ובעלי משמעות אסטרטגית לארגון
 - ✓ היכולת לבצע ניתוחים סטטיסטיים, כריית נתונים וביצוע למידה חישובית (Machine learning)

מבנה הקורס:

- מבוא הבנת הקשר וההבדל בין עולם ה BI לעולם ה BIG DATA
- שלביו של פרויקט BIG DATA מסוגיו השונים, כולל התייחסות ל SDLC הן בגישת מפל המים והן בגישה הספיראלית וכמובן בעולם ה AGILE
- הקורס מתבסס על פרויקט מעשי, עליו עובדים המשתתפים בקורס לאורך כל שלביו, משלב הייזום ועד להצגת המתווה לפרויקט (כמצגת) בשעור האחרון.
- הפרויקט מתבצע ומוצג בקבוצות של מספר תלמידים העובדים יחדיו, על בסיס תבנית המועברת לתלמידים בשעורים הראשונים.

בסיום הקורס החניכים יוכלו:

- לדעת מתי נכון להכנס לפרוייקט BIG DATA
- להגדיר דרישות עסקיות לפרוייקט big data
- להגדיר את הצוות הנדרש ותוכנית עבודה עקרונית לפרוייקט
- יכולת איתור מקורות המידע אפשריים, כולל מתודולוגיה והכלים לאיסוף הנתונים ממקורות מידע אלו
- חקירת איכות הנתונים וכלים לביצוע חקריה כזאת
- שיקולים לבחירת בסיס נתונים (SQL / NoSql) ומידול מודל הנתונים

- יכולת הבנת הצורך בניתוחים סטטיסטיים, כריית נתונים וביצוע למידה ממוחשבת (Machine learning)
- סוג ה machine learning שכדאי להשתמש
- הבנה היכולת ניתוח ועיבוד מידע לא-מובנה ורב (non-structured data)
- היכרות וכלים פשוטים להצגת הנתונים ובעלי משמעות אסטרטגית לארגון

כלים:

- תבניות אפיון וניהול ככלי עזר למנהל הפרוייקט ומנתח המערכות
- תרגול והדגמת machine learning בעזרת ORANGE
- כלי ניהול משימות בעולם נתוני עתק כדוגמת TRELLO

קהל יעד

אנשי מערכות מידע כדוגמת מפתחים, מנתחי מערכות, ראשי צוותים, מנהלי פרויקטים, אנשי QA, מומחי יישום ורפרנטים, בעלי ניסיון מעשי בעולם מערכות המידע ומעוניינים להכיר, להבין להתמחות ולהתמקצע בתחום מערכות big data. הקורס מיועד הן לאנשי IT ולמנהלים המבקשים להיכנס לתחום ה- big data וללמוד כיצד ליישם פרויקטים מעולמות התוכן הרלוונטיים.

דרישות קדם

- ניסיון של מעל שנתיים בעולם מערכות המידע

תכני הקורס

שיעור	נושא המפגש	נושאים שיידונו במהלך המפגש
1	מבוא לעולם ה Big data	<ul style="list-style-type: none"> • מבוא לעולם ה Big Data, מדוע ארגונים רוצים לעשות פרויקטים מסוג זה • מחזור חיים פרוייקט BI / BIG DATA • בעלי תפקידים (מנהל הפרויקט, מנתח המערכת, מפתח, איש הבדיקות) • בעלי תפקידים חדשים: ארכיטקט נתונים, מדען נתונים • סימנים ל Big Data - 3V • ארכיטקטורה קלאסית בעולם ה BI בקצרה • ארכיטקטורה של BIG DATA • הגדרת הדרישות הפונקציונליות וכיצד מנהלים אותם. מה ההבדל בין איסוף דרישות רגיל להיבטים של ניהול דרישות לעולם ה BI, ולעולם ה Big Data
2	כדאיות עיסקית ומחזור חיים ניהול הפרוייקט BIG DATA	<ul style="list-style-type: none"> • הגדרת הצורך, למי זה מתאים • עלויות וכדאיות כלכלית • מחזור חיים (מפל מים, יחידות מסירה, אג'יל) • הרחבה של מפל והרחבה על אג'יל כולל טקסים • תפקידו של מנהל הפרוייקט • תפקידו של מנתח המערכות • פרוייקט BIG DATA • מה קורה כאשר יש "יותר מדי נתונים" - מדוע אנו עושים פרויקטים כאלו • פרויקטי BD - אתגרים עיקריים • אתגרים עיקריים בפרוייקט למידת מכונה • מחזור חיים של פרויקט CRISP - ML • אנשי צוות בפרוייקט וחלוקת תפקידים ביניהם
	אפיון הדרישות BIG DATA	<ul style="list-style-type: none"> • זיהוי הלקוחות והמשתמשים בעולם ה Big Data • ההבדל בין איסוף דרישות רגיל להיבטים של ניהול דרישות לעולם ה Big Data • הגדרות ראשוניות של אבני הבניין • הגדרת התועלות העיסקיות - מה ירוויח הארגון מהשימוש במערכת

<ul style="list-style-type: none"> ▪ הפלטים של המערכת: דוחות, גרפים, לוחות מחוונים ▪ זיהוי מקורות הנתונים הפנימיים והחיצוניים, הן המסודרים, החצי מסודרים והבלתי מסודרים ▪ הבנת עומק איכות הנתונים והתנהלות בהתאם. ▪ זיהוי הדרישות שאינן פונקציונאליות כדרישות למהירות העיבוד, תזמונים ועיבוד במקביל ▪ הגדרת דרישות טכניות (לא פונקציונאליות) לדוגמה: ארכיטקטורה, אבטחת מידע, גיבוי, שחזור, זמינות המידע, ביצועים. ▪ הגדרת הפלטים בהתאמה לסוגי המשתמשים השונים ובחירת כלי תצוגה. ▪ הגדרת המדדים (Matrix) KPI כולל סוג נתון ולאיזה דרישה עסקית הוא שייך, מידת הדיוק נדרשת, גראנולאריות, היסטוריה נדרשת ועומק היסטורי Visualization ▪ סוגי גרפים ▪ Dashboards ▪ KPI 		
<ul style="list-style-type: none"> ▪ מה זה Hadoop ▪ כיצד עובדת Distributed System - מה האתגרים ▪ Master/Slave Architectural ▪ HDFS : HADOOP DISTRIBUTED FILE SYSTEM ▪ Map Reduce ▪ Hadoop-EcoSystem 	<p>Hadoop : hdfs, map reduce sparx</p>	<p>3</p>
<ul style="list-style-type: none"> ▪ המשך נושאים מתקדמים ▪ התלמידים יתנסו בקליטת קובץ ויצפו כיצד מפרק ה Hadoop את הקובץ לבלוקים וכיצד משכפל את הבלוקים. ▪ התלמידים יתנסו בבניית שאילתות לשליפת נתונים 	<p>Hadoop נושאים מתקדמים + תרגול</p>	<p>4</p>
<ul style="list-style-type: none"> ▪ הכרות עם בסיסי נתונים בעולם ה NO SQL כולל דוגמאות על הפופלארים שביניהם גם אם ארגון לא יעבור ל big data באופן מלא, ארגונים רבים חלו לאמץ בסיס נתונים שאינם רלציונים כדוגמת mongo db או redis ▪ מנתח המערכות/ מנהל הפרוייקט נידרש להכיר את היכולות/ היתרונות והחסרונות בשימוש בבסיס נתונים אלו. ▪ רוב עולם מערכות המידע משתמש בסוג אחד של בסיס נתונים RDB ▪ העולם החדש משתמש כבר ב 4 סוגים עיקריים : <ul style="list-style-type: none"> ▪ Key value db ▪ Document db ▪ Column db ▪ Graph db 	<p>NoSql db מבוא והכרות</p>	<p>5</p>
<ul style="list-style-type: none"> ▪ NoSql db - Data modeling in Big data world ▪ Key value db ▪ Document db ▪ Column db ▪ Graph db 	<p>NoSql db תרגול מעשי - בניית ERD</p>	<p>6</p>
<ul style="list-style-type: none"> ▪ מחזור חיים בפרוייקט machine learning ▪ מבוא ל machine learning ▪ מהי "למידה" - סוגים שונים של למידת מכונה (מוכוונת, לא מוכוונת, חוזרת, עמוקה) ▪ מהו נתון, כולל סוגים שונים (רציף, בדיד, מדורג, בינארי) ▪ מהי מטרה ▪ מהו מידע "מתוייג" ▪ אלגוריתמים נפוצים ▪ כיצד מודדים הצלחה של למידה ▪ הגדרת מדדי הצלחה ללמידה ▪ כיצד מעריכים את היכולת שלנו לבצע אותה 	<p>Machine learning מבוא והכרות</p>	<p>7</p>

<ul style="list-style-type: none"> ▪ חקירה מקדמית של הנתונים <ol style="list-style-type: none"> (1) כמות עמודות (2) כמות נתונים בעמודות (3) כמה נתונים ריקים (4) עומק היסטורי שלהם (5) מושגי בסיס : ממוצע, שונות, תוחלת התפלגות ▪ מרחקים ואופן מדידתם 		
<ul style="list-style-type: none"> ▪ אסטרטגיות של השלמות נתונים ▪ אסטרטגיות של עיבודים של נתונים ▪ דוגמה למדידה מכוונת בעזרת רגרסיה לינראית שפוטה ▪ דוגמה של למידה בעזרת "עץ" ▪ דוגמה של למידה לא מכוונת ▪ דוגמה של למידה חוזרת ▪ דוגמה של למידה "עמוקה" 	<p>Machine learning</p> <p>העמקה ותרגול (orange)</p>	8
<ul style="list-style-type: none"> ▪ Why you should always start with questions ▪ How to choose the right chart type ▪ The ins and outs of creating effective views ▪ How to design useful and engaging dashboards ▪ The importance and tips for perfecting your work 	<p>Data visualization</p>	9
<p>כל קבוצה מציגה מצגת עם הפרוייקט שעבדה עליו כל הקורס</p>	<p>השלמות ומצגות סיום</p>	10