



סיכום מפגש סיעור מוחות

Big Data



מנחי המפגש:

פיני כהן, עינת שמעוני



לקוחות נכבדים שלום רב,

אנו מודים למשתתפים במפגש סיעור המוחות בנושא Big Data.

מהן מפגש סיעור מוחות?

- מפגש של כ-20 CTOs, ארכיטקטים, יועצים, מובילי דעה, מנהלים טכנולוגיים בתחומים שונים בארגוני משתמשים, ואנליסטים של STKI. שלא כמו במפגשי שולחן עגול שעורכת STKI (המיועדים לארגוני משתמשים בלבד), במפגשי סיעור מוחות משתתפים הן ספקי IT והן ארגוני משתמשים בתחום ה-IT.
- מפגשים אלה סובבים סביב נושאים חדשים יחסית, כאלה אשר עדיין אינם נחשבים ל-Mainstream ולא ניתן למצוא אותם בקלות בכל ארגון. (מפגשים קודמים שערכנו במתכונת זו – מחשוב ענן, קוד פתוח, מובייל).
- במפגשים נבחנים תחומים/טכנולוגיות אלה, מידת ישימותן בארגונים בכלל ובארגונים ישראליים בפרט, ומועלות סוגיות המהוות חסם בפני תחום זה, הגורמים המניעים אותו, תועלות, חסרונות וכד'.

במסמך זה מובא סיכום עיקרי הדברים שעלו במפגש בנושא ביג דאטה ומידת ישימות תחום זה בארגוני Enterprise בישראל.

בברכה,

עינת שמעוני, פני כהן

תוכן עניינים

4.....	מבוא
4.....	האתגרים העיקריים בעולם ה-Big Data:
5.....	הפער בין חברות האינטרנט לארגוני Enterprise
6.....	דוגמאות לשימושים עסקיים:
7.....	ההיבט הטכנולוגי אל מול ההיבט העסקי:
8.....	מה המשמעות של ביג דאטה בהקשר של ארגוני Enterprise?
8.....	מה המשמעות מבחינת כוח אדם? כישורים נדרשים?
9.....	לסיכום:
10.....	נספח – תגובות ספקים:
10.....	חברת NESS
10.....	NETAPP
12.....	HP
14.....	אמת מחשוב
15.....	Software AG
15.....	EMC
16.....	יעל תוכנה
17.....	מיקרוסופט
17.....	מטריקס
18.....	טלדור

מבוא

Big Data הנה המגמה החמה ביותר בתחום מערכות המידע כיום. כל יום צצים מוצרים חדשים נוספים המכתירים עצמם כמוצרים בתחום זה. חשוב מכך - ברמה העניינית, ישנה אכן פריצת דרך טכנולוגית באפשרויות שלנו לעשות שימוש בנתונים – הן במידע בלתי מובנה והן במידע מחוץ לארגון. ההשלכות הפוטנציאליות לתופעת ה-Big Data הן רבות, הן מבחינה טכנולוגית והן מבחינה עסקית, החל מתובנות עסקיות חדשות שניתן להפיק מטכנולוגיות אלו, וכלה בהתייחסות שונה לגמרי מבחינת תשתיות (לדוגמה, מדוע לבנות cluster לפתרון אנליטי אם ב-Hadoop יש שילוש של המידע המובנה ברמת האפליקציה? מדוע להשתמש באחסון ארגוני? האם יש להחזיק "מחסן נתונים נוסף" שיטפל בעולם חדש זה, לצד ה-DW המסורתי? וכד').

במפגש סיעור המוחות הועלו הסוגיות הבאות:

- משמעויות עסקיות לשימוש בטכנולוגיות big data
- האתגרים העיקריים, רמת בשלות, חסמים
- לאילו ארגונים הנושא יותר מתאים? האם רק לחברות אינטרנט או גם לאנטרפרייז?
- כיצד מתבצעת בחירה טכנולוגית/עסקית בפתרון big data?
- כיצד על הארגון להתאים את עצמו בכדי שיתמוך בטכנולוגיות אלו?
- מהן המשמעויות התפעוליות/תשתיות של טכנולוגיות חדשות אלו?
- מה בין Big Data לעולם ה-DW המסורתי? האם יש מקום לשני תחומים אלה בארגון?

האתגרים העיקריים בעולם ה-Big Data:

- חלק מהותי משימושי הביג דאטה הנה היכולת לנתח טקסטים, אחד מהנושאים שהועלו כבעייתיים במיוחד בישראל הנו **NLP (Natural Language Processing) בעברית**. נושא עברית עדיין בעייתי ומסובך. אמנם יש כבר תמיכה בעברית וניתן למצוא פתרונות שכבר מיושמים בשטח, אבל השאלה היא – באיזו רמה התמיכה? הרמה היא עדיין לא הכי גבוהה. לדוגמה, Entity recognition - זיהוי/חילוץ ישויות מעל מילון בעברית הוא יותר קשה.
- התאמת השיטות האנליטיות לעולם זה הנו גם כן נושא שצריך להתייחס אליו. לדוגמה, צריך לטייב אלגוריתמים כדי להתאים אותם לעולם הביג דאטה – למשל ANOMALLY DETECTION לא רץ כ"כ טוב, היה צריך להגדיר אותו מחדש.
- חלק מהטכנולוגיות החדשות של Big Data מתאפשרות באמצעות ויתור על הקונסיסטנטיות המלאה של המערכת כלומר חוסר עקביות מוחלט בנתונים. במערכות מסוג זה מדובר על מונח חדש "eventually consistent". לכן נשאלת השאלה - איך אפשר "לחיות" עם חוסר עקביות

בנתונים בעולמות ביטוח/רפואה? ההסכמה הרווחת הייתה כי יש צורך לחלוקה בארגונים – לחלק מהצרכים נצטרך טכנולוגיות ביג דאטה (לדוגמה, טרנדים ושיווק) ובחלקים מסוימים יישארו עם ה-BI המסורתית.

- **חיבור עם עולם ה-DW הקלאסי** עלה כצורך חשוב. לדוגמה, היכולת לחבר נתונים של לקוחות מרשתות חברתיות למידע הקיים עליהם במערכות הארגוניות.
- **החיבור בין העולם המובנה ללא מובנה**. חייב להיות קשר וכיום יש כלים שמשלבים. הוזכר נושא של קולבורציה סביב הדאטה - ממשק חדש שהוכרז על ידי EMC ממשק CHORUS¹ קולבורציה שילוב אנשים בתוך הארגון לנתח ולדבר על ה-DATA שיוצא.

בכל מקרה, עולם הביג דאטה הנו עולם חדש לחלוטין, נדרשת חשיבה שונה מזו שהורגלנו לחשוב עליה בגלל מגבלות הטכנולוגיות הקיימות. אפשר לטפל ב RAW של המידע עצמו. איזה שימושים אפשר לעשות? איזה שימושים אפשר לעשות במידע הבלתי מובנה? איך מחברים את זה לעולם הקיים? אין מהפכות חדשות, תמיד מוסיפים לקודמות. השאלה הכי חשובה - איזה ערך אפשר להביא לארגון מבחינה עסקית?

הפער בין חברות האינטרנט לארגוני Enterprise

המאמצות הראשונות והעיקריות כרגע של עולם הביג דאטה הנו חברות האינטרנט. הבעיה של חברות אלה היא סקלביליות תוך מזעור עלויות. גם אם מדובר בחברות סטארט-אפ קטנות, הן מעוניינות לבנות מראש את המערכות בתשתית שהיא סקלבילית.

חלק מהלקוחות כבר מטפלות במיליארדי אירועים ביום (דוגמה לחברה ישראלית - מנתחת שלושה מיליארד אירועים ביום) או מטפלות רק בכמות קטנה של אירועים – מליונים עד עשרות מליונים ליום אך צופות לפרוץ בשנה-שנתיים הקרובות ולכן צריכות פלטפורמה שתדע לגדול פי עשר/ מאה/ אלף תוך זמן קצר, בעלות התחלתית נמוכה מאוד ועלות סבירה בהמשך. חלק מהלקוחות נוטות לכיוון מסד נתונים מקבילי, רלציוני, עבודה ב-SQL וחלק יותר לכיוון ההאדופ. הרבה פעמים ההעדפה להאדופ היא לא טכנולוגית אלא כי הוא נתפס זול יותר מכיוון שמדובר על פתרון שהוא ממשפחת הקוד הפתוח כמו גם פתרונות אחרים בתחום זה.

יש עדיין מעט פרוייקטים חדשים בישראל בקרב ארגוני Enterprise שהם באמת Big Data. כרגע, רוב הלקוחות מחפשים פשוט חיסכון בעלויות – יש להם מחסן נתונים קיים של מספר טרות, לרוב עד עשר, והם רוצים גם קצת לגדול, לשפר ביצועים וגם לחסוך בכסף (אחסון, רשיונות, שרתים). אחד ממשתתפי הדיון סיפר על פרויקט מתהווה של חיסכון בעלויות הגידול של ה-DW. עלתה שאלה באשר ל"האם זה פרויקט ביג דאטה או לא" ולא הייתה הסכמה ברורה בין המשתתפים. חלק טענו שזה לא ביג דאטה, כי אין כאן רצון להתחיל להזרים למחסן הנתונים כמות דרמטית של אירועים גולמיים, פנימיים וחיצוניים, כדי להגיע לתובנות חדשות, אלא רק ליעל את המצב הקיים. מהצד השני, יש להניח שהדברים ייכנסו באופן הדרגתי, ועם הזמן נתחיל לראות יותר הבנה של big data ותכנון פתרונות מתאימים. ייתכן שהצעד הראשון בכיוון יהיה מכיוון זה של טיפול בנפחים

¹ <http://www.greenplum.com/products/chorus>

והכנסת DW Appliances. ברגע שמכונות אלה יהיו כבר בתוך הארגון, הדרך לכיוון מימוש פתרונות אנליטיים חדשים בגישת הביג דאטה תהיה קצרה יותר.

נקודה שיש לשים לב אליה היא החזר השקעה. גם אם הכל "בחינם", לקנות שרתים, רשת, להתקין, להתנסות, לכתוב קוד, לבנות תהליכי העברת מידע, להחזיר את הפלט למערכות התפעוליות וכו' דורש הרבה עבודה והרבה כסף. ארגונים צריכים להבין איך השיפור היחסי הצפוי ביכולת קבלת ההחלטה ביחס לרמת התובנות הקיימות בארגון, מחזיר את עלות הפרוייקט.

פלטפורמות Enterprise Big Data יצמחו לרוב באופן אורגני, במקביל למחסן הנתונים הארגוני. כלומר, כפיילוט עבור מידע שגדול מדי על מחסן הנתונים או שדורש ניתוח אינטנסיבי עתיר משאבים. כך תיווצר בארגון סביבה אנליטית לניתוחי batch כבדים עבור בניית מודלים חדשים של התאמת מוצרים ללקוחות, ניבוי נטישה, זיהוי הונאות וכו'. הפלט של הסביבה ישתלב חזרה במחסן הנתונים כדי להזין חזרה תהליכים תפעוליים, וכך יחסוך השקעה יתרה במחסן הנתונים הארגוני היקר. זה נכון הן עבור האדופ והן עבור סביבות MPP Databases.

יש הבדל מהותי בין חברות ווב לחברות אינטרפרייז – לחברות ווב יש את הפריבילגיה להתחיל מאפס, כך שמצבן יותר קל. בצד ה ENTERPRISE הדבר מורכב – נדיר למצוא DW שבו יש פרויקט חדש שייצור מאסה אדירה של נתונים. אצל רוב הארגונים הגדולים יש כבר DW שקיים 10-15 שנה. הוא לא שומר את כל הנתונים אבל כן את המידע הרלוונטי. הדרך היחידה בה הוא יאמץ ביג דאטה – אם אני אשמור פי 10 יותר נתונים ב DW האם זה ישפר את החיזוי? כלומר, הדרך היחידה להכניס את הפתרונות – מהצד האנליטי. איזה צורך עסקי מצדיק כניסה לתחום?

עולם ה HADOOP עדיין לא בשל בכל המימדים. חברות שיש להן יכולת להשקיע במשאבים ואנשים (כמו חברות אינטרנט בארץ) – משתמשים אבל בשליפות של ה HADOOP ביצועי השליפות בעייתיים. מתמודדים עם נושא הרפליקציות (חברות האינטרנט פיתחו פתרונות). לחברות ETNERPRISES – יש וצריכים להיות פתרונות אחרים ומותאמים.

סיכום של אחד ממשתתפי הדיון על שכבות Hadoop השונות:

<http://it.themarker.com/tmit/article/19542>

דוגמאות לשימושים עסקיים:

כאמור, רוב ארגוני ה Enterprise שנכחו בדיון לא ידעו להגדיר מה הצורך העסקי שלהן, אבל אחד מהמשתתפים דווקא כן פירט באופן ברור מה היה מצפה לקבל, הוא רואה צורך בנושא בתחום רשתות חברתיות ואינטרנט - להביא עוד לקוחות שלא רשומים אצלם וגם מידע מודיעיני מסחרי כמו לדוגמה מחירים שהמתחרים מציעים. כמובן שישנן הרבה שאלות לגבי התארגנות שכזו - איזה פונקציות חדשות יהיו בארגון? איך תיבנה הסביבה? חלק ניכר מהנתונים הם זמניים (מישהו כתב בפייסבוק שהוא רוצה לעשות משהו שמעניין אותה החברה שלי) - לא צריך לשמור את המידע הזה אח"כ. זה נושא חשוב שלא חושבים עליו מספיק (מתעלמים מנושא הזבל – זריקה של נתונים שלא צריך). אותו משתתף תיאר שהיה רוצה לקבל שירות כזה מגוף חיצוני – חפשו לי את המידע הזה והזה... ומתוך זה הוא יבנה מאגר פנימי שלו. כלומר, יהיו אנשים אנליסטים שיגידו מה חשוב לחברה, סינון של מה שקיבלתי מהמסננת החיצונית הזו ויגדירו את המטהדטה על מנת לקשר אותה למערכות הקיימות היום.

תוארו שימושים נוספים (מתוך סיפורי לקוח של יישומים בעולם) - לדוגמה:

- שימוש של חברות ביטוח: לרוב המשאיות בארץ יש סנסור שנותן STREAM של דאטה - איפה הנהג היה וכמה נסע. לוקחים את כל הנתונים הללו ולפי זה מתמחרים את הפוליסות.
- ניהול סיכונים בבנקים – כיום פעם ברבעון מגלגלים את ה RISK DW וחישובים פיננסיים להקצאת ההון. בעולם ביג דאטה יוכלו לעבור גם לריצה יומית או אפילו בתוך יום העבודה בהתאם למצבי השוק. בהקשר של גילוי הונאות, צוין כי גילוי ההונאות כיום הוא רק 10% מהיקף ההונאות!
- כיום אצל הרבה חברות הנתונים הולכים לפח (לדוגמה בתחום הסלולר נתוני שיחות). ב HADOOP אפשר לאגור כמות גדולה של נתונים, אין את אותן המגבלות.
- באחד הארגונים הפיננסיים יש פרויקט עם רשת חברתית וחיבור המידע למידע הפנימי שיש בארגון. המטרה: לייצר תובנות חדשות.
- חיבור פייסבוק לעולם זיהוי הלקוח: סופר על שירות שאחד הספקים מציעים במודל של שירות חיצוני לניהול מועדוני לקוחות – משלבים אותו עם מידע מפייסבוק. אחת מהתובנות – לא כל האנשים מוכנים לתת מידע אך יש דרכים לעשות את זה בצורה חוקית. אפשר ללמוד העדפות, חברים, אירועים, בתי ספר בהם הלקוח למד. מנתחים רק את הנתונים שאותו אוהד נתן להם הרשאה וזה חוקי. חיבור הפייסבוק לעולם זיהוי הלקוח אפשרי.

ההיבט הטכנולוגי אל מול ההיבט העסקי:

הדיון הלך בכמה כיוונים שונים והיה ברור כי עדיין אין הסכמה לגבי הגדרה ותיחום של השוק. רוב הארגונים עדיין חושבים במונחים של עולם ה RDBMS וה-DW המסורתיים וגם משתמשי הקצה רגילים למצב הקיים ו"לקבל תשובה מדויקת". הדבר העיקרי בו כן הייתה הסכמה בין המשתתפים הוא החשיבות הגבוהה שיש לדיון העסקי: איזה שימושים עסקיים ניתן לבצע בכלים אלו? איזה ערך עסקי חדש ניתן לקבל? האם ישנם תחומים שבהם מתבצע כיום ניתוח Batching ואם נצליח לספקו ב real time הדבר ייתן ערך משמעותי? האם יש מקור נתונים שכיום כלל לא מסתכלים עליו של הארגון, שכדאי לנתח אותו?

היה מעניין לראות כי לרוב המשתתפים מארגוני Enterprises שנכחו בדיון לא היה צורך עסקי ספציפי בארגונם שהם יכלו לחבר את הנושא אליו. כלומר, הצטייר מעין שוק חם עם המון הייפ ש"מחפש מטרה". מצד אחד ברור שהפוטנציאל הוא גדול וניתן לעשות דברים חדשים ושונים שעשויים להיות בעלי ערך עסקי רב לארגון, אך לא ברור ממה ואיך להתחיל.

חלק נכבד מהדיון נסב סביב תחומים טכנולוגיים ותשתיתיים. אחד הארגונים סיפר כי בעיית הגידול המתמיד בנתונים ועלויות האחסון הם אלה שמעלים את הצורך בארגונם מתוך ציפייה שמכונת Big Data (DBMS Appliance)

סוגיות ואתגרים טכנולוגיים נוספים כללו – צורך במתודולוגיות וכלים לגיבוי וארכוב חכם של נפחים כ"כ גדולים.



מה המשמעות של ביג דאטה בהקשר של ארגוני Enterprise?

ה DW לא יעלמו אבל יעבדו יותר יעיל. יהיה הרבה יותר שימוש ב HADOOP. יש חסמים רגילים – אבטחת מידע, רגולציה, רגישות מידע. חסמים רגילים שייפתרו. ארגונים משתנים לאט. בעולמות אינטרנט הם כבר שם.

ארגוני Enterprise כרגע שמים דגש על נושא הנפחים אבל לא מספיק על נושא הביצועים. בהחלט ייתכן מצב בו המשתמש מחכה 10 דקות לדוח. היום משתמשים צעירים – לא מחכים לשום דבר בפייסבוק, ומתרגלים לרמת שירות כזו גם בארגון.

ביג דאטה מתעסק באינטראקציות (עולם חדש של אינטראקציות בין אנשים, ש IT עד עכשיו לא ממש ניתח) לעומת טרנזאקציות (העולם שאנו מכירים ויודעים לנתח).

ארגוני Enterprise שנכחו בדיון לא היו בטוחים שבארגונם הדבר רלוונטי. כלומר, יש נושא מעניין, יש הרבה הייפ ועניין וברור שיש פה דבר חדש שצריך ללמוד. אבל ארגונים עדיין לא יודעים לחבר אותו לצורך ספציפי בארגונם. חלקם מתחברים לנושא דרך עניין הנפחים – אם יש DW גדול, אם יש בעיית נפחים, שוקלים את הנושא. אחד הארגונים ציין שגודל ה-DW בארגונו הנו TB8 ולכן זה לא נראה כ"כ רלוונטי. לעומת זאת, אחד הארגונים כיום חווה גידול של 40% בשנה בנפח הנתונים (בעתיד יגדלו ב100% בשנה בנתונים). גידול לינארי בנתונים – מביא לגידול לא לינארי בהוצאות.

היו כמה שאלות שארגונים העלו – איך בדיוק מנתחים מידע ונותנים תשובה? איך מנתחים מידע בלתי מובנה? איך מקשרים את עולם ה-Big data לעולם ה-DW הקלאסי אם בכלל? האם אפשר לתחקר ביג דאטה עם כלי BI קלאסיים? ללקוחות כיום מאוד חשובה הקונסיסטנטיות – לקבל תשובה עקבית, האם זה מתאפשר בביג דאטה? מתי בוחרים באיזו טכנולוגיה - DB רגיל, לעומת DB MACHINE, לעומת HADOOP? כלומר מתי בוחרים בטכנולוגיה גם בשביל לענות על דברים שהיום לא עונים, וגם בשביל לענות מהר יותר על אותן השאלות?

מה המשמעות מבחינת כוח אדם? כישורים נדרשים?

יש בעיה למצוא אנשים שבאמת יודעים Hadoop. מבחינת מיומנויות נדרשות מיומנויות חדשות של אנשי BI. HADOOP היא מערכת שלא נועדה לשליפה על ידי המשתמש. כתיבה של אפליקציה חדשה זו כתיבה של JOB. אפשר לתזמן את ה JOB שרוץ מתי שאני רוצה. צריך לגשר בין הידע של האנליסטים היום ללכתוב קוד ב JAVA – זה מה שחברות ה BI עובדות כדי לפתח – לגשר על הפער הזה.



לסיכום:

בחברות Enterprise רוב הדיון מסתובב עדיין סביב נושא הנפחים ופחות סביב נושאים של אנליטיקה מסוג חדש, ניתוח סוגים חדשים של תכנים וכד'. אחת המשמעויות של ביג דאטה - תובנות חדשות שמתבססות על פתרונות אנליטיים חדשים, אבל ארגונים היום לא יודעים לשאול או לבקש אותם הטכנולוגיה והפרקטיקה עדיין לא בשלים – גיבוי, אודיטים, עברית, ידע של אנליסטים שצריכים להכיר דברים חדשים כמו תיכנות ב-JAVA, ידע בארגונים שמכירים האדופ וכד'. הרבה שאלות לא פתורות – איך מחברים למידע הפנימי? נושא העברית וכדומה.

נספח – תגובות ספקים:

חברת NESS

נס טכנולוגיות הפעילה באמצעות ענן מעבדה לנושא Big Data לצורך הדגמות, השוואות וביצוע הוכחת יכולת (POC). המעבדה ממוקמת בענן של חברת Triple C ומותקנת בה טכנולוגיות Big Data של חברות מובילות בעולם, לצד טכנולוגיות קוד פתוח (Open Source) בתחום זה. במעבדה ניתן להדגים תרחישים שונים המראים את התועלת העסקית שארגונים יכולים להפיק מתחום ה-Big Data על מנת ליצור יתרון תחרותי, תוך שילוב רב-עוצמה של נתונים פנימיים וחיצוניים, מובנים ולא מובנים. לדוגמה: יצירת פרופיל לקוח המשלב נתונים "קשים" ומידע "רך", שנגזר מרשתות חברתיות ומעולם האינטרנט בכלל, כגון סנטימנט לקוח, סגמנטציה חברתית, דירוג מובילות דעה וכדומה – דבר המאפשר הגדלת ערך ללקוח ושיפור שביעות הרצון שלו. במעבדה יודגמו יתרונות נוספים בתחום ה-Big Data כגון יכולת לבצע שאילתות וחיפושים המשלבים מידע מובנה ולא מובנה באנגלית ובעברית, הדגמת יכולות של סקאלביליות-ליניארית, NO SQL הרצת תכניות Map Reduce, שילוב של Hadoop ובסיסי נתונים אנליטיים כגון Sybase IQ, והדגמת Text Analytics.

מכיוון שמדובר בתחום חדש שארגונים חוששים להיכנס אליו, המעבדה מאפשרת לארגונים המתעניינים להתנסות במגוון רחב של כלים ללא העלויות הכרוכות בכניסה לטכנולוגיה חדשה. המעבדה מאפשרת להדגים באופן מוחשי את מה שתמיד רואים במצגות, אך לא תמיד מבינים.

אנחנו מציעים לארגון שרוצה לבדוק את התאמת הפרדיגמה החדשה עבורו, להתנסות בשלב ראשון בנתונים הקיימים שיש לנו במעבדה. בשלב השני ניתן להגדיר הוכחת יכולת על תרחיש עסקי שמעניין את הלקוח, כולל נתונים חיצוניים ופנימיים".

מעבר למעבדה, הפתרון של נס הוא חיבור של יכולות ה-BI ובעיקר בסיסי נתונים אנליטיים (נס גילון) יחד עם יכולות הניתוח הסמנטיות שפותחו בחטיבה הביטחונית (נס TSG) והותאמו למגזר העסקי תוך שילוב עם נציגות טכנולוגיות של חברות מובילות בתחום ה-BIG DATA בעולם (נס פרו). השילוב מייצר ערך ללקוחות ומאפשר להם לקבל פתרון מלא מגוף אחד בנושא שהוא חדש ומורכב. נס מהווה מורה נבוכים לארגון שמתחיל למצוא את דרכו בעולם המושגים החדש ויודעת להתאים את מרכיבי הפיתרון לצורך הייחודי של כל ארגון תוך הצעת ערכים עסקיים מוספים מיוחדים לתחום.

NETAPP

סיכום מנהלים

ארגונים נכנסים היום לעידן חדש של עבודה בקנה מידה חסר תקדים. עידן שבו כמות הנתונים שהחברה מעבדת ומאחסנת נמצא מעבר ליכולת של כל מבנה ארכיטקטוני בתעשיית האחסון. NetApp מספקת פתרונות אשר נותנים מענה לקנה המידה העצום של נפח נתונים גדול (Big Data) באמצעות ה-ABC של Big Data - כולל ניתוח, רוחב פס ותוכן (Content, Bandwidth, Analytics) - ומאפשרת ללקוחות לקבל תובנות בנושא מתוך ערכות נתונים עצומות, להעביר נתונים במהירות ולאחסן תוכן חשוב למשך תקופות ארוכות יותר ללא הגברת מורכבות התפעול.

פתרונות NetApp מבוססים על שילוב של טכנולוגיות FAS & ONTAP ויכולות מערכי E-Series, בשילוב

טכנולוגיות HPC, קוד פתוח HADOOP ו-StorageGrid המתאמים הן לדרישות הניתוח וניהול מידע בהקפים ובקצבים שלא הכרנו בעבר, והן לדרישות הניהול, התפעול והתמיכה הקיים בארגוני האנטרפרייז השונים.

בשנות ה-90, משימתם העיקרית של צוותי ה-IT היתה להפיק את הביצועים הטובים ביותר מיישומי המפתח ומהתשתית הארגון. אותן מערכות של רשומות (Systems of Record) אשר ישבו במאגרים מבודדים היו אמצעי נהדר לניהול מעקב אחר מידע חיוני, אך הן היו יקרות מאוד ולא הציעו לארגונים אפשרות לקבל תובנות עמוקות מספיק על הנתונים ולהשיג יתרון תחרותי בעסקים. בשנות ה-2000, מוקד מאמצי ה-IT עבר לנושא היעילות - כיצד להשקיע פחות ולהרוויח יותר. טכנולוגיות כגון וירטואליזציה, שיתוף ואיחוד של התשתיות הארגוניות הקיימות הפכו ללב הפעילות של מחלקות ה-IT.

גורם נוסף לעלייה בקנה המידה היא ההתרבות המהירה של נתונים שיוצרים מחשבים ומשתמשים. היום, טכנולוגיות דיגיטליות עוברות למדיה דחוסה יותר, כל התמונות דיגיטליות, סרטוני הווידאו מושתתים על רזולוציה גבוהה יותר ויכולות ניתוח מתקדמות דורשות נפח אחסון נוסף. יתרה מזאת, נתונים אוטומטיים המגיעים מרשתות חיישנים, המעקב אחר דפוסי קנייה וגורמים נוספים יוצרים ערכות נתונים גדולות בהרבה שצריכות להיות מובנות ומסחריות. לסיכום, כמות הנתונים גוברת ואובייקטי הנתונים עצמם הופכים גדולים יותר. כל אותם כוחות יוצרים לחץ אדיר על ארגונים להרחיב את התשתיות הקיימות, ובמיוחד את פלטפורמות האחסון. זהו אתגר ה-Big Data שעליו מדברת NetApp.

כעת אנו נכנסים לעידן חדש של עבודה בקנה מידה חסר תקדים. עידן שבו כמות הנתונים שהחברה מעבדת ומאחסנת נמצא מעבר ליכולת של כל מבנה ארכיטקטוני הקיים בתעשיית האחסון היום. כתוצאה מכך, מנסים צוותי ה-IT לקחת את המערכות הקיימות לניהול רשומות, שנבנו בשנות ה-90 וה-2000, ולהמיר אותן למערכות שיתוף פעולה (Systems of Engagement) - מערכות שיכולות לספק ביעילות את המידע הנחוץ לאנשים הנכונים בזמן אמת, כדי לסייע להם לבצע ניתוחים אנליטיים מתוחכמים יותר ולקבל החלטות עסקיות חכמות יותר.

NetApp פיצלה את ערכות הפתרונות לניהול נפח נתונים גדול וחילקה אותן לשלושה תחומים עיקריים, אשר נקראים: "ה-ABC של Big Data" – ניתוח – (Analytics), רוחב פס (Bandwidth) ותוכן (Content). לכל תחום יש אתגרים ספציפיים ודרישות תשתית ייחודיות משלו.

- ניתוח. תחום פתרונות זה מספק בעיקרו כלי ניתוח עבור ערכות נתונים בנפחים גדולים במיוחד. המטרה העיקרית של כלי הניתוח היא לספק תובנות, לנצל את יתרונות העולם הדיגיטלי ולהפוך נתונים למידע באיכות גבוהה, להקנות הבנה עמוקה יותר של הפעילות העסקית ולאפשר תהליך קבלת החלטות טוב יותר. פתרונות NetApp משלבים בין טכנולוגיות FAS ו-E-Series, יחד עם טכנולוגיות הקוד הפתוח HADOOP וכלים אחרים דוגמת בסיסי נתונים ממשפחת No SQL על מנת להשיג יכולות חישוביות לעיבוד וניתוח הנתונים מחד, ופתרון ניהול, תפעול ותמיכה מותאם לארגוני אנטרפרייז מאידך.
- רוחב פס. תחום פתרונות זה מתמקד בהשגת ביצועים טובים יותר עבור עומסי עבודה מהירים מאוד. יישומים עם רוחב פס גדול כוללים מחשוב עם ביצועים ברמה גבוהה: היכולת לבצע ניתוח מורכב במהירות גבוהה במיוחד, הזרמת וידיאו ברמת ביצועים גבוהה לצורך מעקב ותכנון משימות וכן עריכה והפעלה של קובצי וידיאו ביישומי מדיה ובידור. פתרונות NetApp משלבים טכנולוגיות HPC (דוגמת LUSTRE) ומערכות FMV, עם מערכי אחסון E-



Series , המתאפיינים במתן ביכולת ספיקת נתונים גבוהה ביותר, ודחיסות פיזית גבוהה (כ- PB2 למד)

- תוכן. תחום פתרונות זה מתמקד בצורך לספק אחסון נתונים מאובטח ועם יכולת הרחבה ללא גבול. פתרונות התוכן נדרשים לאפשר למשתמשים לאחסן כמויות בלתי מוגבלות של נתונים, כדי שחברות יוכלו לאחסן נתונים רבים ככל שירצו, לאתר אותם בעת הצורך ולא לאבד אותם לעולם. NetApp מספקת פתרון מובנה המשלב בין מערכי אחסון מסוג E-Series או FAS בשילוב תוכנת StorageGrid

העידן החדש של עבודה בקנה מידה חסר תקדים מותח את ארכיטקטורות האחסון עד לקצה גבול היכולת. ארגונים נדרשים לשאול את השאלות הבאות: האם עומדות בפנינו הזדמנויות לנצל את הנתונים שלנו בצורה טובה יותר? אילו תובנות יכולות להיות שימושיות באמת לקידום הפעילות העסקית? כיצד נוכל להפיק מהנתונים יתרון תחרותי? כיצד נוכל להפיק תועלת מהיכולת לקשר בין המגמות בדפוסי הקנייה לבין המיקום הפיזי של המשתמשים בכל נקודה בזמן כדי לספק להם חוויה טובה יותר? כיצד תסייע לנו היכולת לזהות אירועי הונאה מבעוד מועד? האם יש לנו אפשרות לזהות כשלים פוטנציאליים בנקודות חמות ספציפיות לפני שהם מתרחשים?

אין סוף לשאלות. אך התשובה לכולן זהה. NetApp מציעה את פתרונות האחסון המאפשרים לארגונים להשתמש בפתרונות של Big Data לתועלתם ולהפיק מהם ערך עסקי רב יותר. עולם הנתונים יכול להפוך למכרה זהב. NetApp מסייעת לארגונים לזהות את הערך הגלום בעולם הנתונים ולהפוך אותו ליתרון עסקי אמיתי.

HP

מרבית המידע הקיים היום בעולם, נוצר בשנתיים האחרונות בלבד. וכאשר היקפי המידע עצומים כל כך, האתגר האמיתי הוא לנתח אותו בצורה מושכלת, להפיק ממנו תובנות, להסתייע בו בקבלת החלטות ולהיעזר בו להשגת המטרות השונות.

התובנה הזו הובילה את חברת HP לקבל את ההחלטה על רכישת Autonomy בשנה שעברה. בזכות חשיבה אחרת על הנראות העתידית של מערכות המידע ושימוש במה שקרוי על ידינו "טכנולוגיות בסיס המשמעות" (Meaning base technologies), אפשר כיום להבין ולקטלג גם מידע לא-מובנה, מידע שמוכן לכל אדם - תכנים ברשתות חברתיות, תכני וידאו ואודיו ותכנים שנוצרים במכשירים ניידים שונים.

ארגונים נדרשים ולהבין מהו המידע הרלוונטי מתוך מקורות תוכן רבים ומגוונים כגון מיילים, מסמכים, וידאו, רשתות חברתיות ועוד. מעבר לחיפוש של מילה או מונח מסוים, קיים צורך להגיע לתובנות מתוך מאסה של נתונים (Big Data) ובכך לאפשר לארגון להיות תחרותי ולהגיב לשינויים בזמן אמת.

ניתן להדגים זאת באופן הבא – חברת סלולר ולה שירות לקוחות היתה מעונינת לזאת בזמן אמת מתי לקוח מדבר על "נטישת החברה" בין אם המילה נאמרה במפורש או ע"י זיהוי טון השיחה. כאשר מדובר כל מס' אלפי שיחות המתקבלות ביום תהליך זה מצריך כלים אוטומטיים המאפשרים לדוג' לבצע ניתוח רגישות (sentiment analysis) ולזהות האם השיחה היא חיובית או שלילית או לחילופין לבצע אסקלציה בזמן אמת במידה ומיילים מסוימות עולות בשיחה.

דוגמא אחרת היא הקשר מול הלקוחות. חברות מקיימות כיום ערוצי תקשורת מול הלקוחות שלהם במס' ערוצים שונים: אינטרנט, מייל, רשת חברתית, מוקד שירות הלקוחות ועוד. גישה זו נקראת Multichannel, (הגישה הרב-ערוצית). בעידן ה Big Data, קיים צורך לכלי המסוגל להתחבר לכל אותן מקורות מידע, לנתח אותן ולאפשר לארגון להגיע לתבונות עסקיות, כגון מיהו קהל היעד של החברה, מהם הצרכים שלהם, באילו ערוצי הפצה כדאי להשקיע ועוד.

על בסיס הדוגמאות האלה ועוד יישמה HP Autonomy מס' פרויקטים רבים בעולם אשר אפשרו לארגונים לנתח ולהבין את המידע הקיים בארגון בצורה טוב יותר ולהפיק ממנו ערך עסקי אמיתי.

בנוסף לכך, בשנים האחרונות מתקיים שיח ער בנושא Big Data והיכולת של בסיסי הנתונים הרלציוניים, המבוססים מבנה שורה, לתמוך בכמויות ההולכות והגדלות של המידע. תהליך זה מואץ עם ההתפתחות של הטכנולוגיות ועם המעבר ממידע אשר מיוצר ע"י בן אנוש למידע אשר מיוצר ע"י מכונות (machine generated code), אשר מגדיל את כמויות המידע בסדרי גודל, ומייצר מידע בקצב מהיר.

האתגר הניצב בפני ארגונים היום הינו בכמה מישורים:

- צורך בכמויות חומרה גדולות ויקרות – שרתים, דיסקים, רשתות תקשורת וכו'
 - תוכנות יקרות המתומחרות לפי גודל החומרה
 - איסוף המידע באופן רציף ממערכות שונות וטעינתו למחסן הנתונים בזמן סביר
 - כלי תחקור אשר יהיו מסוגלים להריץ שאילתות ולקבל מענה בזמן סביר
 - ועוד
- גורמים אלו ועוד הביאו חוקרים לחשוב על כיוונים נוספים למימוש המבנה של בסיס הנתונים.
- בשנת 2005 פותח מהבסיס הקוד של ורטיקה ע"י פרופ' סטונברייקר, אשר בא לתת מענה לצרכים החדשים ולטיפול בכמויות המידע הגדולות.
- בסיס הנתונים ורטיקה משפר יכולות אל מול בסיסי הנתונים הקיימים ונותן ביצועים של בין פי 50 לפי 1000 יותר מהיר, בעלות תועלת נמוכה יותר מהסביבות הקיימות.

בין היתרונות המובהקים של ורטיקה:

- בסיס נתונים ניהול עמודה (Column store)
 - יכולות טעינה מסיביות ותחקור בו זמנית
 - יכולות כיווץ אגרסיביות מאוד
 - שימוש בתשתית סטנדרטית, גריד של שרתים ויכולת גידול ליניארית
 - זמינות גבוהה מובנית בפתרון
- כיום עושים שימוש במוצר כבר מעל 600 לקוחות, בין הלקוחות של ורטיקה – זינגה (Zynga), וריזון (Verizon), דאבל וריפי (DoubleVerify), מוזילה (Mozilla), קומקסט (Comcast) ועוד.

ורטיקה מאפשרת למשתמשים לייצר הכנסה ממידע.

אמת מהשוב

חברת א.מ.ת מיחשוב משווקת, מטמיעה ומתחזקת מערכות Big Data מסוגים שונים. תפישת העולם של א.מ.ת גורסת שהאתגר ב Big Data אינו דווקא הגודל (Big) או ה Data-עצמו אלא איך להנגיש תובנות אפקטיביות מכמות גדולה של נתונים באופן פרקטי. כדי להבטיח שימוש נכון בנתונים א.מ.ת משווקת פתרונות Big Data במספר קטגוריות.

כידוע, נהוג לחלק את עולם המידע בהיבטים של Big Data לשלש קטגוריות :

1. נתונים מבניים לחלוטין – מרחב הפתרונות של בסיסי נתונים (יחסיים מבוססי SQL וגם מאלו ששיכים לקטגוריית, NoSQL)
2. נתוני אינטראקציות בין אנשים – הודעות דוא"ל, קבצי מלל, גיליונות אלקטרוניים, מצגות וכיו"ב.
3. נתוני מכונה – נתונים המיוצרים באופן אוטומאטי על-ידי מכונות (מחשבים, מערכות אחסון, ציודי תקשורת, מוצרי תוכנה) המתעדים פעולות, ביצועים, תצורה, שימוש ועוד.

בא.מ.ת ניתן למצוא פתרונות מובילים בתחומם בכל קטגוריה :

1. המוצר המוביל היום בתחום ה Big Data-הטבלאי הוא ללא ספק Oracle Exadata חברת א.מ.ת מיחשוב היא השותף הבכיר של חברת Oracle בהטמעת מערכות Exadata בישראל. א.מ.ת היא מרכז ההדגמה הישראלי למערכות Exadata ניתן לתאם תהליכי הוכחת יכולת ושימות של Exadata כפתרון Big Data או כתשתית קונסולידציה לבסיסי נתונים.
2. בתחום האינטרקציה הבין-אישית א.מ.ת מטמיעה מערכות מבית סימנטק וחברות אחרות להתמודדות עם ניהול, מפתוח, איחזור וכריית מידע ממידע בין אישי.
3. בתחום נתוני המכונה האתגר הוא מעט שונה. נתוני מכונה לא נמדדים בכמות שורות או במספר מסמכים אלא בקצב ייצור. אחת מהתופעות הנלוות לשימוש מסיבי בוורטואליזציה ושירותי ענן, כמו להשלמת מיחשוב כל תשתיות הארגון, היא הגידול המסיבי בנתונים שמערכות יודעות לספק על פעילותן. האתגר מרכזי הוא איך להפיק תובנות בזמן אמת מנתונים המיוצרים בקצב גבוה. א.מ.ת היא נציגת Splunk בישראל – המוצר הטוב בעולם לטיפול בכמויות ענק של נתוני מכונה Splunk. הוא מוצר Big Data ייעודי תחום נתוני המכונה ומאפשר לשמור, לנתח ולהפיק דוחות והתראות מנתוני מכונה בכל קצב. לקוחות Splunk מנתחים מידע בזמן אמת מנתונים המגיעים לקצבים של עשרות טרהבייט ביום ויותר.



בנוסף, א.מ.ת מקיימת פעילויות סביב מוצרי ניהול מידע נפוצים בעולם הקוד הפתוח. חברת אינטרביט מבית א.מ.ת מיחשוב היא שותפת ההדרכה של Cloudera בישראל ומקיימת בארץ הדרכות רשמיות של Cloudera לתשתיות מבוססות Hadoop. ניתן למצוא בא.מ.ת גם שירותי הדרכה, ליווי והטמעה עבור מוצרי ניהול מידע אחרים בתפיסות NoSQL שונות.

Software AG

חברת Software AG המתמחה בפיתוח פתרונות תוכנה לייעול והאצת תהליכים עסקיים, רואה בטכנולוגיות ה-Big Data נדבך נוסף וחשוב בפתרונות להאצת ביצועי הארגון. כחלק מתפישה רחבה, מציעה החברה ללקוחותיה את **Terracotta**, פלטפורמת לניהול המידע **In-Memory Terracotta**. הינה טכנולוגיה חדשנית מבוססת זיכרון Cache המאיצה עד פי 1000 יישומים ארגוניים ומאפשרת שיפור זמני התגובה של יישומים ובנוסף מאפשרת הקטנת התלות במסדי נתונים קיימים אשר מהווים צוואר בקבוק בכל הקשור לטיפול בכמויות מידע גדולות.

כפי שצוין במהלך מפגש סיעור המוחות, ארגוני Enterprise מתמקדים בעיקר בנושא נפחי המידע אך לא שמים דגש מספיק על נושא הביצועים וזאת למרות שטכנולוגיות האצת הביצועים בשלות יותר וכבר מוטמעות בארגונים רבים. ארגונים אשר מעוניינים להיכנס לעולם ה-Big Data וחוששים מבשלות טכנולוגיות האנליטיקה יכולים להתמקד בנושא האצת הביצועים ולהגדיל באופן משמעותי את כמויות המידע הנמצא בשימוש המערכות.

טכנולוגיות מבוססות זיכרון משמשות להאצת ביצועים במערכות זיהוי הונאות, שיפור מערכות Billing מקוונות, הגדלת המידע המוצג ללקוחות בפורטלים ארגוניים, קיצור זמני חישוב של מערכות ליבה כגון ניהול סיכונים ועוד.

כאמור, טכנולוגיות הענן וה-BigData הינן חלק מתפישה רחבה שמטרתה להטמיע טכנולוגיות אלו בפתרונות התוכנה אותן מציעה החברה. הטכנולוגיה של **Terracotta** תשולב בפלטפורמת ה-Business Process Excellence של Software AG ותאפשר האצה מהותית של הביצועים ויכולות גמישות אשר יבססו את מעמדה המוביל של החברה בתחום זה. באמצעות **Terracotta** תציע החברה ללקוחותיה הקיימים והחדשים יישומי BPM מואצים בסביבת ענן אשר יסייעו להם להגדיל ולייעל את הכנסותיהם ועסקיהם.

EMC

חטיבת **Greenplum** בחברת EMC מתמחה בתחום **Big Data Analytics**.

ל-Greenplum יש נוכחות גבוהה בישראל, כולל מרכז פיתוח בעל מעל לעשרים איש, ובנוסף, אנשי Professional Services, ארכיטקטים ואנשי מכירות.

לתפיסתנו, יש מקום בעולם ה-Big Data לשני סוגי התשתיות:

- תשתית למידע מובנה בעזרת מסד נתונים – **Greenplum Database** הוא פיתרון מוביל לתחום מחסני הנתונים. הוא מאפשר טעינה מקבילית וניתוח מקבילי של כמויות מידע בלתי

מוגבלות, וגידול פשוט ונוח אל מספר לא מוגבל של שרתים. הוא כולל יכולות מתקדמות כגון שמירת נתונים לפי עמודה או שורה (row or column orientation), Partitioning וכיווץ נתונים, חלוקת משאבים בין קבוצות משתמשים ועוד. בנוסף, הוא מממש סט גדול של אלגוריתמי **Data Mining** בצורה מובנית בתוך מסד הנתונים, כדי לאפשר ניתוח מלא של כלל הנתונים.

- **תשתית למידע לא מובנה** – (Hadoop) האדופ הוא הפיתרון המוביל בתחום זה. בתחום זה אנחנו מספקים זוג פתרונות המאפשרים לקבל יכולות תשתיות מתקדמות שלא קיימות בהפצות האדופ סטנדרטיות, כגון זמינות גבוהה, שיפורי ביצועים, יכולות Snapshotting ו- Disaster Recovery, גישה לקבצים גם ב-NFS ועוד. הפיתרון הראשון מבוסס על חומרה (אחסון Isilon) ומאפשר גם חיסכון משמעותי בנפחים ושיתוף פשוט של מידע ללא העתקת קבצים. הפיתרון השני מבוסס תוכנה בלבד (GPMR). כמו כן, אנו מספקים גם הפצת האדופ מבוססת על Apache בשם GBHD, כולל תמיכה ושירות.

בנוסף, קיימת כבר היום **אינטרציה מקבילית ויעילה בין התשתיות**, כך שניתן למשל לבצע שליפה ממסד הנתונים ולהצליב בצורה מהירה בין נתונים טבלאיים לנתונים היושבים בהאדופ ללא צורך בהעתקת הנתונים.

מעבר לכך, ניתן לקבל כל מהמוצרים הנ"ל כמוצר תוכנה ולרכוש בנפרד את התשתית הפיזית (שרתים, רשת וכו') מכל ספק חומרה או לחלופין לרכוש Appliance מוכן (**DCA – Data Computing**) (**Appliance**) המאפשר קיצור זמן הפרוייקט, ביצועים ואמינות גבוהים ומוכחים ואחריות (ושירות) על כל הבטי התשתית.

תחום משלים נוסף לתשתית הנתונים הוא תחום סביבת העבודה. אנחנו מאמינים שעל מנת להתמודד עם כמויות ענק של נתונים מסוגים שונים, יש צורך בצוות יעודי של אנליסטים (**Data Scientists**). מנסינונו, לצוות כזה יש צרכים יחודים, בדגימת נתונים, תחקור נתונים, עבודה משותפת ועוד. לצורך כך פיתחנו סביבת עבודה יעודית לצוות של Data Scientists בשם **Chorus** שנועדה לשפר משמעותית את תהליכי העבודה של צוות ניתוח הנתונים.

יעל תוכנה

"גרסה חדשה של Spotfire (4.5) מאפשרת מהיום לארגונים להתמודד עם דרישות וצרכים בעולם Big Data בנוסף ליכולות ניתוח מידע באופן ויזואלי, שיתוף מידע, חיזוי תוצאות עסקיות, ניבוי מגמות שוק ועוד. יכולת זאת מתאפשרת באמצעות קישור ל-Hadoop ומנוע חיפוש Attivio.

Spotfire 4.5 מאפשר שילוב תוכן מובנה עם תוכן שאינו מובנה, איחוד נתונים ממקורות שונים ומקנה יכולות לניתוח נתונים באופן דינמי. יכולות אילו מיושמות תוך שימוש בטכנולוגיית In memory כמו גם באמצעות יכולות גישה לנתונים לפי דרישה (On Demand). Spotfire 4.5 הינו בעל יכולות לניתוח ושיתוף מידע מסוגים שונים והפצתו לגורמים הרלוונטיים בארגון. הפתרון מאפשר גישה למנהלים, היכולים מעתה לגשת אל המידע הרלוונטי, בזמן הנדרש ובאמצעותו לקבל החלטות מושכלות.

כעת, באפשרותנו להציע לארגונים נגישות פשוטה לכל סוגי המידע, ללא תלות בכמות המידע, סוגיו, מורכבותו או מהירות העברתו כולל מענה ל - Big Data. השימוש ב-Spotfire 4.5 מאפשר לארגונים לשלב בין מספר מקורות מידע וליצור תובנות עסקיות שלא היו אפשריות קודם. Spotfire מאפשר את

'יתרון שתי השניות' בו העברת מידע בזמן הנכון חיונית יותר מכל דבר אחר. נקודה זו, ממצבת את הפתרון כאחד מפלטפורמות ה-BI המובילות בעולם ועם זאת, הפשוטה ביותר לשימוש".

מיקרוסופט

1. בדיון הועלתה סוגיית הבחירה בין מסד הנתונים הארגוני לבין העבודה ב-Hadoop בחברות אינטרנט, כמו גם הצורך לחבר בין שתי הסביבות במידה וצמחו במקביל בתוך ארגון Enterprise.

פתרון ה-Big Data של מיקרוסופט מאפשר ליהנות משני העולמות על ידי שילוב מיטבי בין Hadoop ובין מסד הנתונים הארגוני. לשילוב שלושה אספקטים:

הרצת Hadoop על פלטפורמת מיקרוסופט. ניתן לבחור בין הרצת Hadoop על Windows Server, לבין הרצתו בענן Windows Azure המאפשר סקלאביליות לרמות ה-petabyte. הרצת Hadoop על פלטפורמת מיקרוסופט מקצרת את זמן ההטמעה בצורה משמעותית, עד כדי 3 שעות על Windows Server ו-10 דקות על Windows Azure. כמוכן מתאפשר ניטור וניהול של Hadoop באמצעות System Center וכן החלת מדיניות אבטחה ארגונית באמצעות Active Directory.

חיבור מובנה בין Hadoop למחסן הנתונים הארגוני. מיקרוסופט מציעה שני מחברים דו-כיווניים בין Hadoop ומחסי הנתונים של מיקרוסופט – SQL Server (פתרון SMP) ו-Parallel Data Warehouse (פתרון MPP המיועד למחסני נתונים עד TB600). בכך מתאפשר שילוב הפלט מסביבת ה-Hadoop חזרה במחסן הנתונים הארגוני כדי להזין תהליכים תפעוליים. החיבוריות מאפשרת גם אופטימיזציה של עלויות אחסון וניהול (עדיף ב-Hadoop) מול ביצועי שליפות (לעיתים עדיף במחסן הנתונים הארגוני).

כלי הבינה העסקית עובדים מעל Hadoop ומעל מחסן הנתונים הארגוני. ארגונים שמתמשים בפתרון ה-BI של מיקרוסופט (SSAS, אקסל, PowerPivot, Power View) מעל מחסן הנתונים הארגוני יכולים לתחקר באמצעותם גם מידע לא מובנה מ-Hadoop. זאת – בשל חיבוריות שמיקרוסופט בנתה כגון Hive add-in for Excel.

2. בדיון תואר שימוש עסקי בתחום ניתוח מידע מרשתות חברתיות, תוך הדגשת הצורך בקבלת המידע כשירות על מנת לייבא את הנתונים הרלוונטיים בלבד.

בנפחי הנתונים הקיימים היום אין אפשרות ואין תועלת בייבוא כל המידע לתוך הארגון ושמירתו באופן מקומי. לפיכך מיקרוסופט מאפשרת ללקוחות להתחבר למידע מכל העולם, לרבות נתונים מרשתות חברתיות, באמצעות שוק מידע – [Azure Marketplace](https://azuremarketplace.com).

מטריקס

אחד האתגרים המרכזיים בעולם ה-Big Data הינו אחזור המידע באופן שיאפשר הפעלת כלי BI ומנועי חיפוש לצרכי תחקור וניתוח. מנהלי IT בארגונים שונים בעולם המנסים להטמיע פתרונות בתחום מגלים לא אחת כי האתגרים בעולם ה-Big Data הם מעבר לנושא נפחי המידע הנדרשים לאחסון.



במבט מעבר לניהול הנפח, ניתן למצוא כי עדיין קיימים אתגרים תשתיתיים בכלים לניהול Big Data, בין אם על בסיס ADBMS, Hadoop, או שילוב של שניהם. ארגונים המבקשים למקסם את הערך העסקי של Big Data חייבים לתת מענה לאתגרים אלו.

מטריקס, המייצגת את חברת Attivio מביאה פתרונות מקצה לקצה בתחום אחזור המידע המובנה והבלתי מובנה וכן יישומי תחקור וניתוח מתקדמים למגוון רחב של מגזרים עסקיים. פתרונות אלו משלימים את הערך העסקי של תשתיות ה-Big Data השונות ונותנים מענה ב-3 היבטים עקריים:

1. סגירת הפער הקיים בין המשתמש העסקי ל-Big Data :
היישומים הנוכחיים דורשים רכיבים רבים כגון DATA SCHEMA ושאלות SQL. רכיבים אלו בדרך כלל אינם נגישים לכלים כמו Hadoop העובדים ב-BATCH MODE. יחד עם latency נורמאלי הקשור במחסי נתונים, אנו מקבלים פגיעה ביכולת קבלת החלטות בזמן אמת. מנוע האחזור של Attivio מאפשר כלי BI מסורתיים לעבוד על הנתונים ללא צורך בהעברתם למחסן הנתונים
2. ניהול מספר רב של מקורות מידע בו זמנית:
מערכות מידע פועלות בדרך כלל על מקור מידע אחד בנקודת זמן. המנוע של חברת Attivio (UIA – Unified Information Access) בשילוב יכולות אנאליטיות מתקדמות של מטריקס מאפשרים קבלת החלטות מהירה יותר המבוססת על מידע רב יותר ומבוססת על בסיס נתונים אחוד.
3. פיתוח יכולת הבנה עסקית קונטקסטואלית
תוכן לא מובנה, כגון מסמכים, דואר אלקטרוני, תוכן אינטרנט, טקסט חופשי זורם, SharePoint, יומני קריאה, וסקרים, אינם מסווגים או מנותחים ולכן לא יכולים לספק תובנות עסקיות רלוונטיות. בעוד שרוב פתרונות ה-Big Data מספקים יכולת שמירה של מידע בלתי מובנה בלבד, בעוד פתרון המשולב של Attivio ו- מטריקס מספקים גם רכיבים כגון:
 - Sentiment Analysis
 - Text Analytics
 - Contextual Business Insight

מטריקס מספקת יכולת טכנולוגית בעלת ערך עסקי ישים המהווה פוטנציאל לבניית יתרון תחרותי. בנוסף, פיתחה החברה מתודולוגיה לביצוע הוכחת יכולת תוך מספר שבועות בודדים.

איתי גבאי, מנהל תחום פרויקטים, מטריקס BI, 054-7878950, itaygab@matrixbi.co.il

טלדור

בשנתיים האחרונות ביצעה טריבייס, מקבוצת טלדור, עבור לקוחותיה (חב' הייטק בינלאומיות) למעלה מ-10 פרויקטי Big Data. עם כניסת טריבייס לתחום לפני כשנתיים, הקימה החברה ע"ג ענן



של Amazon מעבדה אשר מכילה את מגוון הטכנולוגיות והפתרונות בשוק בתחומים הללו. יש לזכור שחלק רחב מהטכנולוגיות הינן טכנולוגיות Open Source אשר נעטפות היום תחת מעטפת שירות כוללת עבור ארגוני ה-Enterprise השונים בעולם.

טריבייס מקבוצת טלדור הינה מפיצה של Cloudera – מובילת שוק עולמית בתחום ה-Hadoop לארגוני אנטרפרייז בתחום ה-Big Data.

טריבייס רואה את תחום ה-Big Data מחולק ל-3 תחומים:

1. Hadoop

2. NoSQL

3. MPP

טריבייס מאמינה ששילוב של 3 הפתרונות הנ"ל מאפשר לארגונים לתת פתרונות שלא ניתן היה בעבר לנתח ולהניב מהם תובנות עסקיות, ולייצר בארגון במקביל חיסכון עלויות משמעותי בגין תשתיות קיימות. לטריבייס מתודולוגיה סדורה עבור הטמעת טכנולוגיות Big Data בארגונים הכוללת:

1. אפיון וניתוח צרכי הנתונים של הארגון - שדרוג תהליכים קיימים אשר אינם מקבלים מענה מספק מהתשתיות הקיימות, ו/או מציאת פתרונות אפליקטיביים חדשים אשר לא ניתן לבצע ע"ג התשתיות הקיימות

2. בניית ארכיטקטורת הפתרון ובחירה בטכנולוגיות המתאימות, תוך התייחסות לחשיבות הרבה לשילובם של מספר פתרונות. לדוגמא, שילוב MPP עם NoSQL/Hadoop

3. כתיבת אב טיפוס

4. בדיקות עומסים

5. כתיבת מסמך פתרון מלא עבור הלקוח

6. תחזוקה מלאה של הפתרון הן בצד התשתיות עם חברת GlassHouse, מקבוצת טלדור, בפתרונות מבוססי Cloudera והן בצד האפליקטיבי.

בשנה האחרונה העלתה טריבייס לאוויר מספר פרויקטי Big data המנתחים את המידע מרשתות חברתיות ובלוגים ברשת, ומתרגמים אותו לתובנות עסקיות והמלצות פרסונליות פר לקוח.

כמו כן, חברת טריבייס הקימה והסבה מספר של מערכות לסביבת Big Data, הן כפתרונות תחליפיים למטרות שיפור והתייעלות והן כפתרונות משלימים לסביבה המסורתית, ובפיתוח פתרונות פרסונליזציה ואנליטיקה חדשניים ע"ג סביבות אלו עבור חברות אינטרנט בינלאומיות.



לטריבייס ו-Cloudera ניסיון רב ביישום Hands On של פתרונות מבוססי Hadoop, עובדה
המאפשרת ללקוחותינו ליהנות מנתיב מהיר להצלחה בבחירת ובניית אפליקציות Big Data

למידע נוסף:

משה מורשת - סמנכ"ל מכירות, קבוצת מערכות, טלדור: 03-9298338, 054-7339997
moshem@taldor.co.il

Ness-Pro

חיפוש ארגוני בעולם ה-Big Data: נס החלה לאחרונה לייצג את חברת BMMSOFT שמשרדיה ממוקמים בסן פרנסיסקו ארה"ב. מוצר הדגל של החברה נקרא EDMT והוא מחבר את עולמות המידע המובנה והבלתי מובנה - טרנזקציות של בסיסי נתונים, אימיילים, מסמכים, לוגים של סנסורים, בלוגים ומדיה חברתית. כל אלה יכולים להישמר בתוך מאגר אחד המאפשר לחפש בתוכו מידע המאוחזר בשניות בודדות בלבד.

כבר בשנת 2007 חברת BMMSOFT פרצה דרך כאשר בנתה את מחסן נתונים הגדול ביותר בעולם (זכתה בשיא גינס) בגודל של 1 PB של מידע ממערכות ייצור (6 טריליון רשומות) בתוך מחסן נתונים אחד. בזמן הטעינה המערכת קלטה 285 מיליארד רשומות ביום (35 TB ליום). אבל החוכמה היא לא רק לאחסן כמובן, יש הרבה מאוד בתי קברות של מידע (Data Cemeteries). היתרון ב EDMT שהוא המאפשר לחפש רשומות בודדות בשניות.

חשוב מכך כיום EDMT מסוגל לחפש נתונים מובנים ולא מובנים באותו חיפוש ולעשות קורלציות בין הנתונים, דבר שהיה בעבר בלתי אפשרי. EDMT עושה זאת בשניות בודדות.

EDMT נמצא כיום בשימוש ע"י ממשלות, ארגוני טלקו, ארגוני ביון, חברות לעריכת דין, חוקרי הונאות, ארגוני ביטוח ועוד.

פתרון ה-EDTM בנוי על התשתיות המתקדמות בעולם מבית HP ו-SAP המאפשרים להזיל עלויות ולהגיע למחיר של 500 דולר ל-TB בלבד. הדבר הופך את החיפוש הארגוני בעולם ה-Big Data לאפשרי וזמין.

EDTM מגיע בגרסאות extra-small ועד ל-Extra-Extra Large. כיום המוצר נבדק במעבדות HP עם מאגרי מידע עצומים של 10 PB.

לפרטים יובל מלחי, מנהל פעילות Data Solutions נספרו 052-3655835