



סיכום מפגש שולחן-עגול

סנכרון נתונים תפעוליים

מנחים:

עינת שמעוני

פיני כהן

לקוחות נכבדים שלום,

תודה על השתתפותכם במפגש שולחן עגול Round Table בנושא סנכרון נתונים תפעוליים.

מצ"ב סיכום עקרי הדברים שעלו במהלך המפגש. במפגש עלו נושאים מהותיים שתומצתו בסיכום כפי שעלו. אין בסיכום זה המלצה גורפת ללקוחות אלא מתן פרספקטיבה והצגה של ההתלבטויות שעלו במפגש, כלומר - "מהשטח".

אחת התובנות המרכזיות שעלו בדיון היא שנכון להיום תחום "ניהול הנתונים" אינו מפותח בצורה מיטבית בקרב ארגונים רבים. סיבה מרכזית לכך היא שניהול נתונים מרכזי מחייב שיתוף פעולה של גורמים עסקיים שונים, דבר שהוא בעייתי מיסודו.

כדוגמה לחוסר הבשלות הכללי בתחום סיפר אחד הלקוחות שלפני שבנה סביבת ODS הוא פנה לאינטגרטורים מובילים בישראל. על פי הלקוח, כל אינטגרטור הגדיר בצורה אחרת את המשמעות של ODS MDM DW וכד'. בסיכום הדיון התייחסנו ספציפית להבדלים בין ODS לבין DW וכמו כן בקצרה למשמעות של MDM.

תובנה נוספת שעלתה היא השונות הרבה בין ארגונים בישראל בהתייחסותם לתחום. החל מארגונים אשר בהם קיימת כפילות נתונים ללא ODS וללא התייחסות כוללת, וכלה בארגונים אשר כבר הגדירו מדיניות של Data Governance וחילקו את האחריות בין הגורמים העסקיים לגורמי ה-IT.

יש ארגונים אשר אינהרנטית סובלים פחות מהבעיה של ניהול הנתונים. דוגמה לארגון כזה אשר החליט לפני זמן רב שכל הנתונים ישבו במערכת המרכזית: ארגון זה העביר לדוגמה נתונים שהוזנו בתוך ה-CRM לתוך המערכת המרכזית, למרות ה"בזבז" שבדבר. אולם כעת הארגון קוצר פירות מכיוון שברמה הבסיסית קיים מקור אחד של מידע בארגון.

כמו כן, יש לציין ש-DW הוא מונח קיים ובשל ברוב רובם של הארגונים. אולם לעומת זאת MDM היא תפיסה שנמצאת בתחילת דרכה.

בברכה,

עינת שמעוני ופיני כהן

תוכן

4	הצורך ב"ניהול נתונים" ארגוני
5	סוגיות אשר מקשות על טיפול ריכוזי בנתונים
6	מילון נתונים לסביבת DW
6	מוקד מידע ניהולי
6	הגדרות – מה בין ODS ל-DW?
7	Operational Data Store – ODS
8	Data Governance
9	התלבטויות הקשורות למדיניות נתונים ארגונית וליצירת MDM
9	בדיקות ובקרה ו-KPI לנתונים
11	נספח: תגובות ספקים ויועצים

הצורך ב"ניהול נתונים" ארגוני

הבעיה הקלאסית בתחום הנתונים היא שעם השנים פותחו הרבה מערכות תפעוליות, כאשר כל מערכת מנהלת עולם מושגים עסקיים משלה. דוגמה בולטת היא ישות "לקוח" אך ישנן הרבה דוגמאות אחרות – לדוגמה, במערכת "פיקדונות שקל" הטבה ללקוח הוגדרה תחת שדה "הטבות" כאשר במערכת "פיקדונות מטבע חוץ" הוגדרה ההטבה ללקוח תחת שדה "בונוס". מקרים מבלבלים עוד יותר כדוגמת מקרה בו שם השדה בשתי מערכות הנו זהה, אך התוכן שלו הנו שונה. במקרה זה, כאשר הארגון מעוניין לבנות תמונה המכילה את כל הסכמי הלקוח, לא ניתן לעשות זאת כי לא ניתן להביא למכנה משותף את תוכן ההסכמים החוזים בצורה פשוטה. ישנם גם הבדלים עסקיים בין מערכות באותו ארגון. לדוגמה, ישנן מערכות אשר מבצעות חישוב עד ליום העסקים האחרון של ההסכם, וישנן שמבצעות עד ליום קודם. ישנן מערכות הכוללות יום שבת, וישנן מערכות שאינן כוללות יום שבת. גם כאן, כאשר רוצים להעביר למכנה משותף הסכמים הקשורים לאותו לקוח במערכות שונות – הדבר בלתי אפשרי.

לגבי סוגיית ה"לקוח": בחלק מהארגונים הנושא טופל כבר לפני שנים. כך, לדוגמה, בבנקים קיימת "מערכת ניהולית" אשר רק בה מוגדרים לקוחות. עם זאת, גם בבנקים אשר טיפלו בסוגיה זו ישנן לעיתים קשיים נקודתיים – לדוגמה, במקרה של סניפים בחו"ל אשר פועלים עם מערכות מחשב נפרדות ומגדירים את הלקוחות בצורה נפרדת. ישנם בנקים אשר כבר מכילים בתוך נתוני ה-DW נתונים אשר הגיעו מסניפים בחו"ל, כאשר זיהוי הלקוח מתבצע למעשה ב-DW. אולם לעיתים ישנו סיכון בארכיטקטורה זו מכיוון שה-DW אינו מעודכן באופן מיידי וכן זיהוי לקוח הוא תהליך batch שלפעמים כולל התערבות אנושית – כלומר, אין עדכון מיידי. ולכן אפשרי שלקוח שחסום בישראל יקבל שירות בחו"ל, או לחילופין שהבנק אינו מודע לחשיפה שלו ללקוח מסוים. כמו כן, למרות שבאופן טבעי הרבה פעמים קורה שה-DW הופך להיות תפעולי, לא תמיד ה-DW בנוי ברמת השרידות המתאימה לתצורה תפעולית שכזו.

נקודה נוספת היא שבמקרים רבים אין חשיבה על מילון נתונים כללי בזמן בניה או שינויים במערכות התפעוליות. לדוגמה, אין שפה אחת לפרויקטים השונים. "ריבית משוערכת" במערכת אחת אינה "ריבית משוערכת" במערכת שנייה, וכן מערכת "פיקדונות" לא מתחשבת בצרכי ה-DATA של "מערכת עו"ש". מצב אותו ניתן למצוא בארגונים הוא שאנשי ה-CRM, אפליקציה "צרכנית נתונים" כבדה לכל הדעות, גזרו נתונים עסקיים מתאימים, אך המפתח לאחד מסוגי הרשומות לא נגזר, וכעת, כאשר המערכת כבר כמעט לפני ייצור -

מאוחר מדי להכניס את השינויים המתאימים. דוגמה נוספת שהוזכרה היא שסוג הנתונים שונה במערכות התפעוליות, ורק בזמן הגזירה ב- ETL הדבר התגלה. ברור שמבחינה תהליכית מחייבים את אנשי האפליקציה לשתף את אנשי ה- DW בשינויים מסוג זה, אולם עקב עומס וטעויות אנוש ישנם "פספוסים". אחד האתגרים הוא למנוע תקלות מסוג זה על ידי אמצעים טכנולוגיים ולא רק תהליכיים.

ישנם מקרים שבהם ישנם שינויים ארגוניים בחברה המחייבים איחוד בין חברות הבנות אשר מרכיבות את הקבוצה. במקרה זה ה- DW של הקבוצה, שכבר זמן רב מאחד נתונים מכל חברות הבת, שודרג והפך להיות "תפעולי" – מתעדכן ומספק מידע בזמן אמת, כאשר המערכות התפעוליות נשענות עליו.

דוגמה נוספת היא התפתחות במודל עסקי. לדוגמה, לקוח סיפר על כך שבזמנו, הישויות במערכות המידע (ובארגון!) היו "כרטיס אשראי" ו"בתי עסק". אולם מודל זה מפספס את הנקודה שלעיתים ללקוח יש מספר כרטיסי אשראי ומהצד השני ישנם מספר בתי עסק הקשורים לאותו ארגון (כמה סניפים) כך שהישות המרכזית צריכה להיות "לקוח" – אך לא ברור כיצד ניתן להגדירו לאורך כל הארגון, וקשה מאוד להסכים על כך ארגונית, פוליטית וכד'. התפתחות כזו מביאה גם לשינוי במודלי הנתונים ובהתייחסות לנתונים.

סוגיות אשר מקשות על טיפול ריכוזי בנתונים

אחת הסוגיות אשר מקשות על טיפול ריכוזי בנתונים בארגון הוא תקצוב ישיר של הגורמים העסקיים. כאשר התקצוב במערכות המידע מתבצע באופן ישיר על ידי הגורמים העסקיים ישנה פגיעה בכלל התחומים המהווים תשתית כללית למערכות המידע וספציפית בתחום של ניהול הנתונים המשותף, וזאת מכיוון שההשקעה בתחום זה אינה אינטרס ישיר של מחלקה עסקית ספציפית.

סוגיה אמיתית אחרת שמקשה מאוד היא שאם רוצים לגרום לכך ששתי מערכות ידברו באותו עולם מושגים, ברוב המקרים יש לכתוב את המערכת מחדש!!! לדוגמה, אחד הלקוחות אשר חויב לאחד שתי מערכות (פיקדונות שקליים ופיקדונות מטבע חוץ) ביצע את המשימה אך תוך איחוד טכני כאשר שדות 1 עד 100 משמשים את הפיקדונות השקליים ושדות 100 עד 200 משמשים את פיקדונות מטבע חוץ. המשימה בוצעה אך אינה מאפשרת לצור תמונת מצ"ב אמיתית של הלקוח ישירות מתוך המערכת. אם היה רוצה לאחד "באמת" את המערכות הוא היה חייב לכתוב מחדש את המערכות – השקעה של עשרות שנות אדם רבות – דבר שאינו ריאלי.

מילון נתונים לסביבת DW

ארגון מתקדם מהתחום הפיננסי תאר מצב שבו ישנם 3 מאגרי מידע תפעולי מחוץ ל-DW. כמו כן, יש מילון נתונים ארגוני ל-DW וגם למערכות תפעוליות. במילון הנתונים ישנה התייחסות לדברים כמו פירוש המונח, מהי מערכת התשתית שממנה נגזר המונח, כמה זמן הנתון שמור ב-DW, מיהו האחראי, היכן הנתון נמצא וכד'. הארגון משתמש ב-Babylon Enterprise, מה שמאוד מקל על השימוש במילון הנתונים. בעקרון, לא מכניסים שום מאגר ל-DW אם אין הגדרות מתאימות במילון הנתונים! באופן כללי הטיפול בנושא זה אינו פשוט, במיוחד כאשר ישנם קשיים לגבי נתוני עבר, כי לא תמיד יש את המידע המלא לגבי נתונים אלו. למען הפרספקטיבה, בארגונים ותיקים וגדולים מדובר על סדר גודל של 5000 אובייקטים! תכונה מעניינת נוספת היא שאנשים מהשטח יכולים לבקש לעדכן מונח או לעדכן הסברים במילון נתונים.

מוקד מידע ניהולי

אחד הלקוחות תאר מצב שבו יש לארגון מוקד מידע ניהולי, הכולל 5 אנשים, שתפקידו לעזור למשתמשים בשאלות במידת הצורך. זאת מכיוון שבארגון זה ישנם מאות משתמשים במערכות ה-BI המתקדמות. המוקד מסייע בנושאים טכניים, כמו דוחות מורכבים, אבל גם בשאלות עסקיות של משמעות הנתונים והקשרם! ישנם ארגונים גם עם מאות משתמשים אשר לא מחזיקים מרכז מידע מסוג זה.

הגדרות – מה בין ODS ל-DW?

ארגונים אחרים התייחסו להבדל בין ODS¹ לבין DW² תוך התייחסות לכך ש- ODS עדיין נותן מידע זמין יותר, עדכני יותר, עם יותר לוגיקה עסקית זמינה (מחושבת מראש). כמו כן ODS גם שומר היסטוריה "נאמנת למקור", ללא המניפולציות שעושים ב-DW. לעומת זאת, ה-DW יכול, מעבר למידע ההיסטורי, גם מידע "אנליטי", ולא רק מידע תפעולי שאותו יכול ה-ODS.

שאלה כללית חשובה ביותר – היכן יש למקם לוגיקה עסקית? לדוגמה – "מיהו הבעלים של בית עסק?". מדובר במקרה זה על לוגיקה עסקית הקובעת מי הבעלים של עסק ומי הם אנשי המפתח בארגון. דוגמה רווחת נוספת היא "האם הלקוח הוא לקוח אסטרטגי?" האם

¹ ODS- Operational Data Store – מחסן נתונים תפעולי

² DW – Data Warehouse – מחסן נתונים ארגוני

הלוגיקה צריכה להתבצע ב- ODS, או ב- DW, או במערכת התפעולית? במידה ולוגיקה שייכת לתחום עסקי אחד ולמערכת תפעולית אחת, אזי אין ממש התלבטויות, כי הלוגיקה "תשב" במערכת תפעולית ותעדכן את ה- ODS ואת ה- DW. אולם במקרים רבים מדובר על מצב שבה לוגיקה עסקית קשורה לתחומים עסקיים ולמערכות רבות יש התלבטות מי יבצע את החישוב. האם אחת מהמערכות התפעוליות ש"תנדב", או שהלוגיקה תתבצע ב- ODS, או ב- DW. לפעמים מדובר על סוגיות טכנולוגיות: אם לדוגמה בכדי לחשב לוגיקה מסוימת צריך חישוב BATCH אבל ה- ODS אינו מבצע חישובי BATCH אבל ה- DW כן...?

Operational Data Store – ODS

ארגון אשר בנה לאחרונה ODS תאר את מגוון השיקולים אשר הנחו אותו בתהליך זה. באותו ארגון נכנסת לשימוש מערכת CRM גדולה אשר מחייבת קבלת נתונים מעודכנים ממספר רב של מערכות, ביניהן סביבות legacy (MF ואחרות). הייתה אפשרות שה- CRM ייגש בעצמו למערכות התפעוליות, אולם היה חשש בארגון מ"סתימת" ה- MF וגידול משמעותי בעלויות ה- MF. עד לכניסת מערכת זו היה בארגון DW אשר סיפק לארגון הן נתונים אנליטיים והן נתונים תפעוליים. אפשר היה גם תיאורטית להפוך את ה- DW לזמין ו"שריד" עם ביצועים טובים מאוד אולם ה- DW בארגון זה מאוד כבד ולמערכת ה- CRM אין צורך בכל עושר הנתונים של ה- DW. כמו כן ה- CRM עובד 7*24 לעומת ה- DW אשר, עקב חישובים כבדים וטעינות מורכבות, אינו זמין לגמרי. לבסוף הוחלט על הכנסת סביבת ODS כפלטפורמה כללית לארגון, בעלת ביצועים ושרידות ברמה הגבוהה ביותר המעודכנת ברמה מתאימה – חלק מהמידע מעודכן ברמה יומית וחלק מעודכן כל 5 דקות. שמירת המידע ב- ODS תהיה לשנה (למעט מצב שבו יש עסקאות פתוחות מעל שנה, שגם הן ישמרו). זאת לעומת, ב- DW שמירת המידע תהיה לשנים רבות ברמת גרנוולריות שונה. בארגון גם הוחלט שה- ODS יהווה מקור המידע ל- DW, כלומר יהווה staging area ל- DW.

ברמה הארכיטקטונית נבנה ה- ODS משתי סביבות של שרתים העובדים כ- Active Pre-ODS. שרת אחד הוא ה- ODS המשרת את צרכני המידע. השרת השני הוא ה- Pre-ODS אשר מבצע עיבודים כגון טרנספורמציות של קודים (כל הקודים הפוכים להיות אחידים) ומבצע את הלוגיקות המוכנות שבהן ה- CRM ישתמש. ה- Pre ODS – משמש גם ל- BATCH, כלומר מערכות ה- BATCH ניגשות ל- preODS או ל- DW. ה- preODS עובד כאשר ה- ODS למטה.

חלק מהארכיטקטורה מדברת על החלפת השרתים בין ה- ODS ו- Pre-ODS. כלומר מבצעים filp-flop בין השרתים. ארכיטקטורה זו בודקת כל הזמן את נושא היתירות של המערכת.

Data Governance

בדיון השתתף ארגון שמבצע פרויקט מאוד מתקדם בנושא ניהול נתונים - data governance. בפרויקט זה מוגדרים אחראים על הנתונים ברמה ארגונית ברמות השונות. הסיבה העיקרית לכניסת הארגון לפרויקט לא קל זה הנה בקרות/רגולציה, במקרה זה BASEL II. אחד הנדבכים בתקינה זו מדברת על כך שנתונים צריכים להיות תואמי חשב, ותואמים בין המערכות השונות. ללא הגדרה של "אחראי" פורמאלי לנתונים לא ניתן לוודא שאכן הנתונים תואמים בין המערכות השונות.

לפיכך במסגרת הפרויקט הוגדרו התפקידים הבאים:

- Data Owners - אחראי על נתונים ברמה העסקית – בד"כ מנהל אגף. כל מה שקשור לעדכון הנתונים מבחינת המהות שלהם חייב להיות קשור אליו.
- Data custodian – אחראי לנתונים ברמה פרטנית יותר עדיין מתחום העסקי. בד"כ יש כמה Data Custodian הכפופים ל- data owner. תפקיד ה- Data custodian לטפל באופן פרטני בנתונים מסוימים. ה- Data Custodian משתמש בכלים לבדוק שהנתונים אכן נכונים – שאין חריגות, שיש התאמה לוגית (כמו - סכום הנתונים מהמערכת מתאים לסכום שמופיע במערכת אחרת). במידה ויש אי-התאמה נקודתית, אחראי ה- Data Custodian לטפל בחריגה ידנית (תאריך שהוקלד בטעות כ- 1.1.2100 הוא בפועל 1.1.2010). במידה ויש חריגה מהותית הנושא עולה ל- Data Owner בכדי למצוא פתרון מערכתי.
- Data Steward – סגן מנהל מחלקה ב- IT שאחראי לנתונים ברמת מערכות המידע. עובד מול ה- custodian.

הערה: לעיתים ישנה הגדרה שונה במקצת של בעלי התפקידים כאשר משתמשים במונח של Data Administrator³.

כמו כן, יש ועדה שאחראית על נושא המידע – אמורה להתכנס פעם בחודש. בארגון יש כבר מערכת התאמות ותיקה - מערכת שתפקידה הוא לגלות אי התאמות בנתונים.

אחד הרעיונות שהועלו בארגון במטרה לשפר את תחום הנתונים בארגון הוא שכל הדוחות הכספיים יצאו מה- DW ולא ממערכות התפעוליות (מערכות החשב)! כלומר, במידה ויגיעו למטרה זו המשמעות היא שהארגון צריך לסמוך על ה- DW בעיניים עצומות. ולכן איכות

³ ראה הרחבה ב- Master Data Management: Transcending Technology, Solving Business Problems - אשר ב- <http://www.cutter.com/content/itjournal/fulltext/2007/09/index.html>. ניתן לפנות אלינו בכדי לקבל העתק של מאמר זה.

הנתונים היא קריטית. הפרויקט מתוכנן בצורה כזו שבהתחלה יעבדו על שתי המערכות בו זמנית. מערכת מנהלית – נתוני לקוח.

התלבטויות הקשורות למדיניות נתונים ארגונית וליצירת MDM

באותם הארגונים אשר קיימת כפילות של נתונים כפי שהוצגה בפרק הראשון (שם הלקוח ופרטיו/או לחילופין - שם ופרטי המוצר, מופיעים בכמה מערכות) יש צורך לאחד נתונים אלו מהמערכות התפעוליות השונות. פרויקט זה נקרא פרויקט MDM⁴. אך ה-MDM אינו מטפל רק באתגר הטכנולוגי, מדובר הן בבעיה טכנולוגית המערבת כלים ותהליכים ב-IT והן בבעיה עסקית-ארגונית ותהליכית לא פשוטה כלל.

מחקרים נוספים על MDM ומצב השוק בארץ ובעולם ניתן למצוא מידע נוסף [בבלוג של עינת שמעוני](#), ו**[בבלוג של פיני כהן](#)**.

בדיון עלו מספר סוגיות הקשורות ל-MDM: מתי להשתמש ב-DW ומתי ב-ODS? כאשר סוגיה קשורה היא "מי בעל הבית של הנתון?" הסוגיה בעיקר רלוונטית כאשר ה-ODS וה-DW לא יושבים יחדיו מבחינת מבנה ארגוני. במקרה שבו ה-ODW וה-DW מנוהלים יחד סוגיה זו פחות בעייתית.

אחד הארגונים תאר תהליך שבו עבר ייעוץ/אבחון של IBM בתחום הנתונים העסקיים. אבחון אשר סייע לארגון להגיע למסקנה שיש צורך להשקיע בפרויקט MDM. בעיה בבניית MDM היא קושי עסקי ליצור את ישות הלקוח. הכוונה לכך היא שכאשר רוצים ליצור את ישות הלקוח חשוב להכניס את הפרטים המהותיים ללקוח, אולם, לדוגמה, בתחום הביטוח ישנם המון פרטים כאלה ולא קל לקבוע מה מהותי ומה לא. לדוגמה, בתחום הביטוח גיל ההורים חשוב לביטוח חיים ולביטוח בריאות (שתי מערכות שונות) ולכן נראה שגיל ההורים חשוב ומהותי ולכן ראוי שישב ברמה הראשית של ה-MDM – רמת הלקוח. אולם לביטוח חו"ל נתון זה אינו רלוונטי. במידה ובכל זאת יבקשו פרטים של גיל ההורים בזמן יצירת לקוח עקב הזנת פוליסת חו"ל, יש סיכוי גדול שימלאו פרטים לא נכונים.

בדיקות ובקרה ו-KPI לנתונים

אחד הארגונים סיפר שלאחר כל שינוי (שאינו טריוויאלי) המשתמש העסקי מבצע בדיקת תקינות.

באחד הארגונים בנו "מערכת בדיקות" ל-BI. מדובר על כמה טבלאות גנריות עם "תשובות" ידועות מראש (סיכומים, מינימום, מקסימום, count וכד'), במקרה זה משימת הבדיקות

⁴ MDM – Master Data Management

מוטלת על ה-QA. בדיון הוזכר גם מוצר DQA של גילון המבצע בדיקות בנפרד של החוקים. כמו כן, הוזכר Trillium בייצוגה של חברת אמן ככלי מתקדם בתחומו. אחד הארגונים דיבר על הרצון לקבוע KPI לאיכות הנתונים. אם יש יותר מדי "בעיות נתונים" הדבר יוצף ל- Data Owner ולעיתים אף למנכ"ל. בארגון שומרים היסטוריה של שגויים והטיפול בהם. "חיפוש" השגויים מתבצע ברמת ה-DW אולם אם מגלים שיש תקלה, ה-Data Stewart אמור לתקן גם את הנתונים בפועל וגם את המסכים ואת הלוגיקה כך שהתקלה לא תחזור (בקשה מהפיתוח לשיפור בדיקות ה-data entry אם צריך). באחד הארגונים דובר על בדיקה פעם בשבועיים אם יש נתונים חסרים בכל הטבלאות.

נספח: תגובות ספקים ויועצים

חברת SAP ישראל

הדברים האמורים בסיכום הינם נכונים וברורים ואף מסכמים את המצב האמיתי בשוק בפועל. יש אירגונים רבים שמבינים את הצורך בנוכחות יישום MDM באירגון.

SAP כספק פתרון MDM מנסה בשנים האחרונות לעזור לאירגונים להטמיע מתודולוגיית MDM, אבל שמדובר בפריסה אירגונית רחבה יש לעיתים בעיה בהכנסת המתודולוגייה, לדוגמא עם ניקח חברת ביטוח שיש לה ריבוי מערכות תפעוליות לסוגי המוצרים השונים או ניקח גם בנק שרוכש בנק אחר וכמות המערכות התפעוליות רק גדלה אזי לקוח מופיע במספר מערכות רב וקשה לנהל את המידע עליו, MDM תפקידו לעזור לאירגון כזה לבצע איחוד, טיוב וסינכרון המידע מהמערכות השונות ולהציג שורה אמיתית אחת לגבי הלקוח ביחד עם זאת לקיים את ה- Data Governance שהיא תמיד באחריות הגורמים העיסקיים באירגון. MDM חייב לתמוך גם ב- Data Entry שמאפשר לאירגון לפתוח לקוח חדש ו/או לעדכן פרטי לקוח קיים בכל המערכות התפעוליות מתוך ה- MDM.

לדעתי ובדרך ש- SAP מתייחסת ל- DW או ODS, המענה לכל קיים ב- DWH, מבחינת SAP לא מערבבים בין הנושאים של MDM ו- ODS. במקרה שקיים DW באירגון ו- MDM ה- DW מחובר לנתוני האב שמנהלים ב- MDM והוא מהווה מקור מרכזי לנתוני האב באירגון.

נושא נוסף שלא הוזכר בסיכום מההיבט הטכני, הוא ש- MDM נמדד גם ביכולת להגיע ולשלוף מידע ממקורות מידע שונים, תכונה זו חשובה לאירגון שמנהל כמה מערכות תפעוליות שונות בטכנולוגיות שונות.

נכתב ע"י: שמוליק סיטון, SAP ישראל

חברת אורקל ישראל

מטרה מרכזית בארגונים היא להביא למקום אחד לאמת ("Single point of truth"). מדובר על אתגר לא פשוט מהסיבה שהנתונים התפעוליים נמצאים במספר אפליקציות בתוך ולפעמים גם מחוץ לארגון. כאן נכנס הצורך במחסן נתונים תפעולי שתפקידו לשמור את העותק העדכני של הנתונים ולשרת את המערכות התפעוליות. החצנת שירותים של תחקור ממחסן המידע עדיפה מבחינה תפעולית אך לא תמיד אפשרית כיון שרוב הארגונים לא יכולים לשנות את המערכות התפעוליים שיצרכו את השירותים האלו – אם בגלל קיום של מערכות Legacy ואם בגלל שימוש באפליקציות מדף עם מודל נתונים משלהן. במקרים כאלו יש צורך בפרויקט MDM שתפקידו לסנכרן את המידע המעודכן חזרה למערכות התפעוליות. בשתי הצורות על מחסן הנתונים התפעולי להתעדכן ולעדכן ככל הניתן בזמן אמת. ברמה התפעולית יש לאפשר קבלת מידע בזמן אמת ללא הפרעה לפעילות השוטפת של האפליקציות. במקרים בהם לא ניתן לקבל את המידע בצורה רציפה ניתן לבנות שכבה שתאפשר שליפה מבזרת ישירות מהמערכות התפעוליות (MDM וירטואלי). בכל מקרה, דרישות הזמינות והעדכון ממחסני הנתונים התפעוליים והאנליטיים יעלו רמה בזמן הקרוב ואמצעים תפעוליים הקיימים היום כמו חלון לילה יהפכו לפחות רלוונטיים עם הדרישות הנוספות. הטכנולוגיות החדשות בתחום ה SOA ו BI משתלבות בכדי לתת מענה למערכות בינה עסקית תפעוליות.

UCM – Universal Customer Master הנו MDM המתמחה באיחוד רשומת הלקוח ממקורת מידע שונים בארגון. הפתרון כולל מנוע חוקה, מנוע תהליכים מובנה תוך ניקוי, מיזוג ואימות הנתונים.

ארגונים מפיקים בעזרת הכלי מידע נוסף על הלקוחות, משפרים את איכות הנתונים, מפיקים תובנות מפעולות הלקוח ומפחיתים עלויות ניהול.

PIM – Product Information Management הנו MDM המהווה קטלוג פריטים ארגוני המאפשר ניהול של כל נתוני המוצרים והישויות הנלוות אליהם במקום מרכזי אחד. גם בעבודה מול מערכות חיצוניות תוך טיוב והגדרת מאפיינים מיזוג ואימות נתונים ומנגנוני workflow לאישורים והרשאות לפי בעלי תפקידים.

חברת אמן מחשבים

MDM הינה תפיסה חדשה ומורכבת, אנו רואים מגמה בשוק העולמי חלוקה ל-3 שלבים של הפתרון על מנת להוריד סיכון ולתת מענה מידי לחלק מהבעיות.

הרכיבים העיקריים ביישום הינם לפי סדר היישום בפרויקט במקרה של חלוקה לשלבים:

- a. ממשק יעיל עם כל המערכות
- b. תחקור נתונים וטיוב (Cleansing , Matching, Enrichment)
- c. מודל נתונים.

לוגיקה עסקית - הלוגיקה העסקית הינה אחת מהדרכים בהן יכול הארגון להשיג יתרון עסקי מובהק על פני המתחרים, על הארגון לגבש אסטרטגיה היכן היא תשב. כמו כן כלי GUI יכול ליעל את הטיפול בלוגיקה, לעזור בניהול, ניטור, והכי חשוב - בהטמעת שינויים עסקיים במהירות.

ראשית נציין כי נושא הממשקים והלוגיקה העסקית מטופלים באמן על ידי שימוש יעיל בכלי אינפורמטיקה, וזאת על מנת להקל במידת הניתן על החיבוריות של מערכות, הן רבות כצרכני מידע והן כספקי מידע. בנוסף, השימוש באינפורמטיקה מבוצע מנת להגיע למידול יעיל של החיבוריות והלוגיקה מאחוריה.

נדברך נוסף בנושא ה MDM הנו שהנושא ברובו על היכולת לזהות ולהבין בעולם הלקוחות מיהו הלקוח ואיזה נתונים מיצגים אותו או משויכים אליו על פני מערכות שונות. יחד עם זאת הדבר תקף גם כאשר המיקוד של פרויקט ה - MDM הוא מעולם המוצרים (PIM).

בין אם מדובר בלקוחות או במוצרים, ישנן מערכות דוגמת כלי **Trillium** המסוגלים להיענות לדרישה. לשם חידוד הבעיה, נתאר תרחיש בו רישום לקוח יופיע הן במערכת הבילינג הן במערכת ה - CRM והן בכל אחת מן המערכות התפעוליות.

מאחר ומדובר במערכות המרובות ב Data Entry נמצא בחלק מהמקרים כי פרטי הלקוח מופיעים בצורה שונה מאוד בכל אחת מהמערכות, כך לדוגמא: במערכת אחת יכולה להיות שגיאה במספר ת.ז. של הלקוח אל מול מערכת אחרת, או מקרים בהם ישנה טעות בשם או בשימוש בכנויי חיבה דוגמת יצחק ואיזיק (או כל וריאציה אחרת), אשר במערכות רבות יגרמו לכפילות בנתונים בשל חוסר היכולת לצפות מראש את סוג השגיאה ולתקנה (די אם נחשוב על מקרים בהם הרשומות מכילות מספר שגיאות שונות בו זמנית).

ניתן לרכז את אתגרי מערכות תחקור וטיוב הנתונים ל MDM על פי הצרכים הבאים:

1. נדרשת מערכת המסוגלת לטפל בנתונים בכל שפה שהיא.
2. נדרשת מערכת בעלת יכולת parsing חכמה ועוצמתית (יוסבר בהמשך).
3. נדרשת מערכת אשר יכולה לעבוד אל מול מילונים חיצוניים ו reference data כשלב מקדים לסטנדרטיזציה.
4. נדרשת מערכת אשר מסוגלת לאחזר מידע אודות ממצאי כפילויות הנתונים על ידי הקניית "ציון" למידת הדמיון בין הרשומות.
5. נדרשת מערכת המסוגלת לבצע linking (קישור) ואיגודן לרשומות אב אל מול רשומות פרטניות (יוסבר בהמשך).
6. נדרשת מערכת המסוגלת לעבוד באופן מקוון.

הכלים המתקדמים ביותר, מבין אלו הנמצאים בשימוש, כגון כלי **Trillium** עובדים על בסיס אמידת הדמיון בין הרשומות השונות ומסוגלים לצור תשתית לאיחוד הרשומות וליצירת "רשומת הזהב" (Golden Record). זאת בניגוד לכלים מדורות קודמים אשר מבססים

את האיתור על מציאת זהות מוחלטת ברמת השדה (למשל איתור לקוחות על ידי השוואת מספר הזהות בלבד) מה שגורם ל"פספוסים" רבים באיתור הכפילות.

יכולת נוספת שיש לכלים אלו (מלבד parser חכם ודינאמי, המסוגל להתאים את חוקיו בהתאם למגוון מקורות הנתונים), היא היכולת לבצע Matching (חיפוש כפילויות על בסיס דמיון) ו linking – כלומר היכולת להבחין במידת השונות בין רשומות שונות, ולהחליט על קישורם ואיגודם לרשומת אב ורשומות בנות. מדובר ביכולת מהותית מאד עבור פרויקטי MDM, שכן לקוחות מאוגדים לעיתים ב"בתי אב" וכך גם מוצרים. מדובר ביכולת ייחודית לכלי **Trillium** שכן רמת המורכבות והיכולת להבחין בין טעויות סטטיסטיות שונות הנה גבוהה.

בנוסף, מציעים כלי **Trillium** את היכולת לעבוד אל מול מילונים (כגון מילוני שמות או קטלוגים) וביצוע השוואה חכמה של data entry מולם. באופן טבעי, עובדים הכלים גם ב on line וזאת על מנת לסייע ככל הניתן בפרויקטי MDM המושתתים על יכולת השוואה של כל נתון בזמן אמת, אל מול ה golden record.

כיום, מרבית פרויקטי ה-MDM המבוצעים בעולם מקצים זמן ומשאבים לטיפול חכם בנושא ריבוי מופעי הנתונים כפי שתואר לעיל" מתוך הבנה שפרויקט MDM יכשל ללא מענה אמתי לבעיה זו.

בארץ, בעוד שפרויקטי ה-MDM נמצאים בראשית דרכם, מומלץ לשלב את הטיפול באיכות הנתונים כשלב ראשי בפרויקט תוך מחשבה על יצירת Data Quality Fire Wall והיכולת לחפש זהות בין הנתונים על בסיס דמיון בלבד, גם כאשר אין מזהה חד - חד ערכי של הרשומה.

עוד נוסף כי אחת התפיסות המוטעות בקשר לנתוני לקוחות היא שנתונים כגון מס' ת.ז. הן בבחינת "כזה ראה וקדש" אולם, מספרי ת.ז. הינם לרוב נתון "מזובל" ביותר. במקרים רבים ישנם בעיות של רישום חלקי של המספר במערות אחדות אל מול רישום מלא במערכות אחרות, בעיות של סיכול ספרות ושיבוש נתונים אחרים כך שאירגון המנסה לאתר את לקוחותיו רק על בסיס נתון זה יטעה בחלק רב של המקרים. הפתרונות הקיימים היום נשענים על עוגנים דינמיים ומידת הדמיון בין הרשומות בין אם מדובר במהלך ראשוני אצוטי או כחלק מפתרון בזמן אמת.

חברת גילון

להערכת חברת גילון, אשר מימשה פרויקטים בתחום, בעיקר בקרב לקוחות ה-enterprise הגדולים:

1. קיימת התעניינות גוברת בתחום אינטגרציית הנתונים בקרב לקוחות ה-enterprise הגדולים
2. כל אפליקציה אשר מוטמעת בארגון סובלת משיעור הולך וגובר של ממשקים – כך שחלק מהותי מהשקעות הפיתוח בארגונים מופנה לתחום זה
3. ה-DWH ביעודו מוגבל לתחום של תמיכה בקבלת החלטות בדרגים השונים (מטה ושטח) ולאפליקציות קרובות (פיננסי, שיווקי וכיוב')
4. למרות הפיכת ה-BI/DWH לאופרטיבי יותר, ארכיטקטורת הפתרון מחייבת גישה שונה ליישום MDM או ODS – אשר תייצר BUS נתונים שיחליף את הספגטי הקיים

5. ההשקעות שיידרשו מארגונים ליישום פתרון אינטגרציית נתונים גבוהות מאוד, עם זאת להערכתנו ה-ROI יהיה גבוה ביותר ויתרום להקטנת משמעות ה-Data explosion שארגונים גדולים חווים
6. הדרך המומלצת היא גיבוש אסטרטגייה וארכיטקטורה ארוכת טווח ויישום מדורג, לדוגמא : רשומת לקוח מאוחדת, רשומת מוצר מאוחדת וכיוב'
7. הצורך בהבטחת איכות נתונים גובר
8. מיישום אחרון בחברת כרטיסי אשראי – מומש פתרון Tailor made, עם זאת לאורך הזמן פתרונות הספקים הגדולים הם אלו שיובילו. עדיין אין ספק מוביל
9. ארגונים הצופים רכישות / מיזוגים וכן ארגונים הכפופים לרגולציה דינאמית יהיו חייבים להכנס לתחום במהרה, אחרת יעמדו בפני סכנה מהותית בזמן שינויים אלו

גל רימון

מנכ"ל גילון ישראל

חברת ליבי טכנולוגיות

אפשרות נוספת להתמודד עם נושאים כמו שפה ארגונית, זמינות מרבית (Near Real Time), והטרוגניות של מקורות המידע היא בעזרת **Data Virtualization** ו- EII - Enterprise Information Integration. אפשרות זו משלבת את השימוש ב DWH המרכזי של הארגון, אך מאפשרת שימוש במאגרי מידע תפעוליים נוספים בשליפת Online של שאליות. בעזרת כלים ייעודיים (דוגמת Composite שמיצגת ליבי טכנולוגיות, WII של IBM ו- Denodo) יש אפשרות ליצור שכבה וירטואלית המאפשרת גישה מהירה (בעזרת אופטימיזציה ו Caching) לנתונים ממגוון מקורות מידע תפעוליים ואחרים ולאפשר למנהלים העסקיים שליפת נתונים לישות ספציפית אחידה – לדוגמה תמונת לקוח כוללת וכדומה.