

## ההבדלים בין מערכות בינה עסקית (BI) ומחסני נתונים (Data Warehouse) בישראל ובארה"ב

לפני כשנתיים חזרתי לישראל לאחר שהות של 16 שנים בארצות הברית בהן עסקתי בארכיטקטורה של מחסני נתונים בארגונים בינוניים וגדולים בסביבות עסקיות שונות (מוסדות פיננסיים, ממשל, מלונאות ועוד).

במלאות שנתיים וקצת לשהותי המחודשת בארץ הקודש אני מרגיש נכון לסיכום ביניים על ההבדלים הלא קטנים בתחום בין שתי התרבויות.

### טרמינולוגיה - BI לעומת DWH

ההבדל המיידי הבולט לעין הוא בטרמינולוגיה. התחום בכללותו נקרא בארץ - BI (Business Intelligence) ואילו בארצות הברית - מחסן נתונים (Data Warehousing). לא מייד הבנתי שהבדלי הטרמינולוגיה נובעים מהבדלים בפילוסופיה. כולם משתמשים באותם כלי BI ו-ETL ואמונים לכאורה על אותן מתודולוגיות DWH (ראלף קימבל ובוב אינמן) אולם בצורת המחשבה הישראלית, יישום דוחות ה BI/OLAP עומד במרכז ומחסן הנתונים הינו פלטפורמה בלבד.

בארה"ב עיקר העשייה בזמן ומשאבים היא - בתכנון מחסן הנתונים (אינטגרציה, טיוב נתונים, סטנדרטיזציה וכדומה) ואילו יצור הדוחות הינו פועל יוצא. הנחת העבודה היא כי אם הבסיס (קרי - מחסן הנתונים) תוכנן נכון אזי בניית דוחות ה-BI הינה משימה קלה יחסית.

### מקורות המידע - כמות וגיוון

הבדלים בולטים נוספים הינם - מיעוט מקורות המידע (data sources) ושונות נתונים מוגבלת באותם מקורות מידע ישראלים ביחס לקיים באירופה או בארה"ב.

בישראל, מחסני המידע של החברות הגדולות כוללים בעיקר מקורות מידע תפעוליים פנימיים. השונות בין המקורות של המימדים (Dimensions) מועטה מכיוון והמידע נוצר רובו ככולו בישראל ולכן מפתחות (keys) רשומות המקור סטנדרטיים (ת.ז., ח.פ. וכדומה). במידה ויש לארגון הישראלי מקורות מידע זרים (כמו חברות בת) במקרים רבים המידע נטען למחסן הנתונים ללא אינטגרציה וללא סטנדרטיזציה. בבנקים ישראלים אחדים, המידע מחברות הבת עבר אינטגרציה רק עקב דרישות Basel II של בנק ישראל ועלה בדמים. ישום האינטגרציה בוצע בארץ בפיגור של שנים רבות אחרי הביצוע המקביל בבנקים בעולם. החוסן שהפגינו הבנקים הישראליים בצונאמי ככלכלי של 2008-2010 נבע ממיעוט פעילות בין לאומית של הבנקאות הישראלית ולא מניהול תקין של מערכות המידע.

חברות בחו"ל בסדר גודל של החברות הישראליות הגדולות נאלצות להתמודד עם מקורות מידע רבים לאותם המימדים (dimensions). זיהוי לקוח למשל אינו חד חד ערכי מכיוון והרשומה יכולה להופיע במסדי נתונים שמקורם בארצות שונות והמפתחות לזיהוי "מיהו הלקוח" עלולים להשתנות ממדינה למדינה. לכן תמיד קיים חשש של זיהוי שני לקוחות שונים בתכלית כאותו לקוח (positive false) או אי זיהוי שתי רשומות שונות של אותו הלקוח כשייכות (negative false). כמו כן חברות מערביות בעולם הגדול נוטות לקנות מקורות מידע חיצוניים המסייעים להבנה העסקית, דבר המוסיף לסיבוכיות אך מגדיל את התועלת שבמידע.

## המבנה הארגוני והתרבות הארגונית:

מחסן נתונים הינו מאמץ כלל ארגוני. ככל שמשימת ה BI של הארגון דורשת נתונים מחלקים נרחבים יותר של הארגון, כך מחסן הנתונים שתומך בדוחות מסובך יותר. יצירת מחסן נתונים ארגוני המאפשר משימות BI מורכבות דורש ארגון בו הרובד הניהולי הגבוה קובע נוהלי עבודה קבועים והחלטות ביצוע פרטניות. ככל שהחלטות הניהוליות נעשות ע"י הדרג הניהולי הבינוני או הנמוך כך קשה יותר (ולעיתים בלתי אפשרי) לכפות סטנדרטיזציה, ניהול ארוך טווח, בניית ארכיטקטורת המחסן (בטרם הפיתוח) כדרוש ושמירה על ה-"one single version of the truth" בארגון.

עבדתי בעונותיי גם עם Freddie Mac וגם עם Fannie Mae שקרסו והולאמו ע"י הממשל בארה"ב, עקב משבר הבנקאות העולמי. שתי חברות הנ"ל הן שהובילו את ניפוח בועת הנדל"ן בארה"ב. שתי החברות היו זהות לכל עניין ודבר. כמות המשכנתאות בשתייהן היתה זהה, מקומן הגיאוגרפי היה זהה (ווישינגטון, די סי) שתייהן העסיקו את אותם מומחים, שתייהן הוקמו תחת אותו צ'רטר פדראלי, שתייהן היו תחת חקירות פליליות ארוכות וקשות של ה SEC ועוד. ולמרות כל זאת בפרדי בנו מחסן נתונים טוב מאוד ושימושי (אחיד ויחיד) ואילו בפאני יצרו מספר מחסני נתונים בעלי סטנדרטים שונים עם שטחי חפיפה עסקית. העדר מחסן נתונים אחיד ויחיד בפאני גרם לכך שפאני לא יכלה לספק דוחות ספציפיים בכל תחום עסקי אליו נדרשה או חמור מכך - עקב ריבוי המחסנים הופקו דוחות סותרים.

ההבדל היחיד בין החברות שהוביל להבדלים בישום המחסנים הוא במבנה הארגוני. בפרדי, ההנהלה הבכירה החליטה לבנות מחסן נתונים ולשם כך השקיעה סכומי עתק וזמן ניהול רב ב- תכנון המחסן, בקביעת מטרות מאקרו ברורות וביצירת סטנדרטים אחידים. ההחלטות (כשבקבלתן ניתנה עדיפות ארגונית למחסן הנתונים) נאכפו על הארגון כולו וכולם פחות או יותר ישרו קו. בפרדי המידע היה שייך לארגון ולא למייצרי המידע או למנהליו. ההחלטה לגבי איזה מידע ייטען למחסן ומתי, נקבעה בצורה מרכזית ולא הייתה בבעלות מתחזקי הנתונים. אני עצמי עמלתי במשך כשנה וחצי להגדיר את המימדים המתואמים (conformed dimensions) של החברה כולה, שכללו - קביעת סטנדרטים, ניתוח, ו-modeling של ה- star schemas העיקריות. הדרישה העיקרית של ההנהלה היתה לבצע את המשימה בצורה מדויקת תוך שמירה על סטנדרטיזציה וחשיבה על צרכים עסקיים עתידיים. ביצועי השאלות העתידיים נבדקו בשלב התכנון, עקב המימדים הענקיים. כמו כן בוצע-benchmarking של פתרונות אלטרנטיביים. בתהליך תכנון סכמות הכוכב נלקחו בחשבון כל תוצאות הבדיקות הראשוניות. נבנה לוח הזמנים על מנת לעמוד בתנאים הנ"ל בלי מגבלת תאריכים אקראיים כסוף רבעון.

בפאני המבודרת לעומת זאת, למנהלי ה work streams (המקבילה למנהל מחלקה פחות או יותר) הייתה רשימת מטלות קצרות טווח לביצוע. צורת הבצוע, התזמון והסגנון היו נתונים להחלטתם. כל חריגה מהרשימה הנ"ל וכל קביעת תזמון בצוע המשימה היה נתון בצורה בלעדית למנהלי ה work streams. בנוסף מנהל ה work streams לא חייב היה לאפשר הרשאות select על שום מידע שבבעלותו גם לא לנתוני דמה. התוצאה הייתה שיותר קל היה להקים מחסן יעודי פרטני חדש לקבוצת מטלות עסקיות ולא להתמקד בעיבוי הקיים הארגונים הגדולים בישראל קרובים יותר בטבעם לפאני מיי מאשר לפרדי מק, עקב אופיו ה"ישראלי" של הארגון.

## תכנון וארכיטקטורה

במחסני הנתונים הבודדים שראיתי בחברות ישראליות גדולות, בלטו בהעדרם - סטנדרטים מחייבים, ניתוח מקיף של המצב הקיים וחסרונם של ה ERDs (שרטוט טבלאות בודדות בכלים סטטיים כגון Word או ב Visio אינו (!ERD).

כל אותם מאפיינים המוזכרים למעלה עולים כסף וזמן אך הם אילו הקושרים את הטבלאות הבודדות ל data marts ואת ה data marts למחסן נתונים אחיד ומאפשרים ניהול ותחזוק תקינים של המחסן לאורך זמן והגדלתו לפי הצורך.

## תקצוב

בעוד שעלות כוח האדם ב IT הישראלי אינה רחוקה מהעלות באמריקה, תקצוב מחסני נתונים בישראל נמוך בכסדר גודל שלם מהמקובל בארה"ב לפרויקטים מקבילים. חלק מהפער בתקציב נובע מכך שהפרויקטים בארץ "קלים" יותר (מקורות מידע מועטים ואחידים) והחלק הנוסף נובע מקיצוץ בשלבי התכנון המאוד נדרשים, כמו למשל – ארכיטקטורה, דוקומנטציה ו-modeling. קיצוצים אלה הופכים את הפרויקט לרזה יותר (זול יותר), אבל בפועל מסכנים את פוטנציאל המימוש בהצלחה של הפרויקט בהווה ובטווח הרחוק.

סיבה נוספת לחוסר התקצוב בישראל נובעת מהקצאת כוח אדם מצומצם לביצוע משימות שהוכרו לכאורה כחשובות כגון – אפיון, פיתוח ו DBA, התוצאה היא פגיעה נוספת בשלבי התכנון וקשירת הקצוות הנובעת שוב מהתמקדות קיצרת טווח.

הקמת מחסן נתונים/בינה עסקית אינה מותרות ומהותה היא תכנון וייעול לטווח הרחוק.

ערן ציפריס

eran@dwlogic.com