



Moshav Bnei Zion P.O.Box 151, 60910 Israel Tel. 972-9-7907000 Fax. 972-97442444



סיכום מפגש שולחן-עגול

DBMS

מנחה
פיני כהן

לקוחות נכבדים שלום,

תודה על השתתפותכם במפגש שולחן עגול Round Table בנושא טכנולוגיות ומתודולוגיות מעולם ה-DBMS.

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

חלק גדול מהדיון הוקדש להרצאה החיצונית שהועברה. התייחסות לכך בהמשך. ומעבר לכך עלו מספר סוגיות משמעותיות (כמו Exadata).

באופן כללי לקוחות דיברו על הבעייתיות בקונסולידציה של מסדי נתונים הגורמת לכך שבזמן השבתה (מתוכננת או לא מתוכננת) אחוז גדול מהארגון נפגע. כמו כן לקוחות ציינו שלדעתם אחוז תקלות בתוכנה (או טעויות אנוש שקשורות לתוכנה) גדול הרבה יותר גדול מתקלות חומרה.

דבר ספציפי נוסף שעלה הוא השינוי במדיניות של מיקרוסופט והמעבר לתמחור לפי CORES (בדומה לשאר הספקים בתחום התוכנה).

בברכה,

פיני כהן

תוכן

3.....	Couchbase להרצאה שהעביר שרון בר מ-
5.....	Exadata מיישום רשמים
5.....	כללי
6.....	תפעול
6.....	Exadata בסביבת DBA תפקיד ה-
6.....	Exadata על עלויות
7.....	ביצוע sizing למערכות
8.....	VMWARE על אורקל
8.....	תגובות של ספקים
8.....	EMC

התייחסות STKI להרצאה שהעביר שרון בר מ- Couchbase

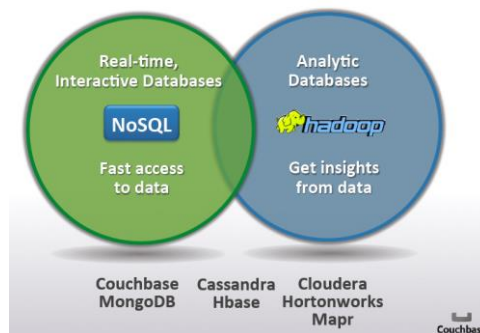
ברצוני להודות שוב לשרון על ההרצאה מאירת העיניים שהעביר שרון. החשיבות בסקירה היא להדגים עד כמה ישנן טכנולוגיות רבות ומפותחות בתחום מסדי הנתונים והאנליטיקה שעדיין לא מצאו את מקומם ברובם המכריע של ארגוני ה-IT, טכנולוגיות שזכות לעיתים להקרא Big Data ושהגיעו במקורן מחברות האינטרנט. חברות אלו מתנהגות באופן שונה מהותית מארגוני Enterprise IT. להלן מספרי מימדים בהם ישנה שונות בולטת:

1. חברות אינטרנט משתמשות באופן רחב בענן ציבורי
2. חברות אינטרנט משתמשות בקוד פתוח לא מסחרי
3. חברות אינטרנט מפתחות ב- AGILE ומכניסות לייצור ב- DEVOPS שמשמעותו גם מבנה ארגוני שוני בהקשר של היחס בין פיתוח לבין ייצור
4. חברות אינטרנט (במידה ולא בענן ציבורי) משתמשות בחומרה סטנדרטית – שרתים פשוטים ללא אחסון מרכזי
5. ברמת עובדי החברה משתמשים ב- BYOD באופן נרחב

אלו רק היו דוגמאות בולטות להבדלים. בה בשעה שהטכנולוגיות והתפיסות שהוצגו למעלה לא מתאימות באופן גורף לארגוני Enterprise IT הרי שארגוני ה-IT חייבים לעקוב אחר התפתחויות אלו תוך הבנה שבטווח הבינוני ארוך לפחות חלק מתפיסות אלו יחדרו גם אל ארגוני ה-IT השמרניים ביותר.

ספציפית לתחום מסדי הנתונים שרון הציג מיפוי שלו של תחום ה-Big Data. בצד אחד Hadoop לביצוע ניתוח של נתונים. בצד השני NoSQL DBMS לביצוע עדכון נתונים on line (כלומר עדכון טרנזאקציוני של מערכות תפעוליות).

Operational vs. Analytic Databases



שרון גם דיבר על תחום מעניין בשם NewSQL¹ שבו מנסים להרחיב את הטכנולוגיות הקיימות של מסדי נתונים רלציונים שעדיין עובדים ב-SQL ושומרים על טרנזאקציות (ACID) תוך עבודה ב-SCALEOUT.

בתוך עולם ה-NoSQL ישנם מספר ווריאנטים עקרוניים:

- Distributed key value – שומרים מסמך לפי מספר מזהה כאשר המערכת לא יודעת להסתכל על תוכן המסמך.
- Document store שבו האובייקטים נשמרים (מעבר למספר המזהה של המסמך) בפורמט הקריא למסד הנתונים – כמו XML או JSON. ואז מסד הנתונים יודע לעשות secondary indexed או לבצע מניפולציות ברמת המסמך בעצמו.

¹ <http://en.wikipedia.org/wiki/NewSQL>

- Graph dbms – שבהם היחסים בין האובייקטים מיוצגים על ידי גרף (דוגמה בולטת היא Neo4J).

הפתרונות בתחום זה עדיין לא בשלים במלואם. לדוגמה בחלק גדול ממסדי הנתונים NOSQL אין עדיין טרנזאקציות על כמה מסמכים (כמה קריאות כתיבות שיתבצעו או לא יתבצעו ביחד...) כלומר בדברים מסוג זה האפליקציה צריכה לטפל.

בדיון שהתפתח בהמשך עלה שלמרות שאין שימוש מסיבי בטכנולוגיות nosql בקרב ארגוני IT בישראל, DBA רבים כבר מודעים לעולם זה, מתחיל כבר שימוש ראשוני (לקוח דיבר על מערכת ראשונה ב-MONGO שנכנסת לייצור) ואף לקוחות מודעים לפתרונות רלוונטיים ל-IT כמו Hparser של אינפורמטיקה או על כך ש-Hortonworks יוצא בגרסה שתומכת ב-Snaps ובאופן זה מאפשרת גיבוי המקובל ב-IT.

לסיכום, מדובר על תחום מתפתח בצורה פנומנלית וארגוני ה-IT מתחילים להתכונן ליישום טכנולוגיות בתחום זה.

רשמים מיישום Exadata

כללי

ההתרשמות באופן כללי היא ש-Exadata "מספק את הסחורה" בהיבט של ביצועים וזמינות הן לסביבת DW והן לסביבות OLTP (בעיקר במצב של קונסולידציה של מסדי נתונים). עם זאת דובר על כך שהכנסה של Exadata מחייבת לבצע שינויים בתפעול התשתיות הרלוונטיות הן בהקשרים טכנולוגיים (גיבוי ב-RMAN לדיסקים פנימיים ורק לאחר מכן גיבוי בתהליכים הרגילים) והן בחלוקת המשימות בין צוותי ה-DBA, אחסון וסיסטם. חיבור Exadata לטכנולוגיות חיצוניות (כגון גיבוי ישירות לתוך DataDomain) אפשרי בדרך כלל אך מחייב מומחיות ותפירה.

תפעול

המעבר ל- Exadata היה די פשוט יחסית אבל בכל זאת מחייב התייחסות לדוגמה הורדת אינדקסים (כלומר הופכים אותם ל- invisible). וגם קינפוג מחדש של dataguard בייחוד למי שקינפג Dataguard ספציפית לפתרון אחסון כזה או אחר (SRDF, True Copy וכד'). בהקשר התפעולי ישנה עוד נקודה שונה משמעותית ב- EXADATA לעומת תפעול IT רגיל והיא חשיבות השמירה על עדכניות מבחינת patch level. חשוב במיוחד שהמערכת תהיה מעודכנת כל הזמן כולל firmware של רכיבי החומרה, גם בכדי שאפשר יהיה לטפל ביעילות בתקלה במידה ותתרחש, ולכן רצוי לשקול במידת האפשר את השירותים המתקדמים של אורקל בתחום זה.

ארגונים שלא רוצים לרכוש Exadata גם לאתר הגיבוי צריכים לפרוש את הנתונים על שטחי אחסון גדולים (כי Exadata דוחס את הנתונים ביחסים מרשימים) או לרכוש בנפרד ZFS (מערכת קבצים של Oracle\SUN).

תפקיד ה- DBA בסביבת Exadata

תפקיד ה- DBA בסביבת Exadata מתרחב באופן משמעותי כי הוא מבצע 80% מהטיפול במכונה. לעיתים מגדירים תפקיד זה כ- dba machine. מצד שני חלק גדול מעבודת הכיוון של ה- DBA נחסכת ב- Exadata. לסיכום לא ברור עד כמה ה- Exadata מהווה תוספת עבודה ל- DBA.

על עלויות Exadata

פתרון Exadata אינו נחשב כפתרון זול אך לפי הנפוצות המרשימה נראה שלקוחות משוכנעים שמדובר על פתרון בעל ROI באופן מובהק. עם זאת בדיון עלה שלקוחות צריכים להתייחס ולהתכונן ספציפית לאפשרות שיבצעו הרחבה של Exadata בעתיד. זאת מן הטעם שבמקרים רבים הכניסה הראשונית ל- Exadata אינה מחייבת רכישה של רישוי התוכנה של מסד הנתונים עצמו, רישוי RAC ורישוי Partitioning (מכיוון שהארגון מבצע קונסולידציה של מסדי נתונים קיימים ומשתמש ברישוי שכבר קיים) אבל בזמן ההרחבה הארגון יצטרך לרכוש רישוי זה. יש להתייחס לנקודה זו כבר בזמן הרכש הראשוני.

ביצוע sizing למערכות

בדיון עלתה סוגייה של ביצוע sizing למערכות – בעיקר כמה שרתים יש להקצות מראש לאפליקציה מסויימת.

במידה מדובר על מערכת קיימת ורק רוצים לשדרג את השרתים ניתן לעיתים להשתמש בחישובים המתבססים על מבחני TPC ולפי תוצאות המבחן בשרת\מעבד הישן והחדש לחשב מה מספר השרתים\מעבדים שנדרשים.

באופן כללי – תמיד עדיף לבצע סימולציה (בדרך כלל Load Runner) אבל הבעיה כאן היא שבמידה ומדובר על הרבה אפליקציות המשימה של כתיבת הסקריפטים היא משימה קשה. לגבי שימוש ב- Oracle real application testing –RAT – עלו מספר סוגיות במימוש הפתרון.

אפשרות אחרת היא ביצוע סימולציה של מערכת בהיבט מסדי הנתונים. לפי קונספט זה בודקים מה השאילתות שמתבצעות באופן תדיר, רושמים את השאילתות האלה, בונים data set ייעודי לצורך ביצוע הניסוי וגם רושמים פרמטרים לשאילתות. את כל זה אורזים בקובץ הפעלה ואותו מבצעים כמה פעמים באמצעות כלי scheduling (הלקוח שתאר את הפתרון השתמש כאן באינפורמטיקה לביצוע ההרצה החוזרת). במובן מסויים יש כאן בנייה של הקונספט של load runner אבל רק ברמה של מסד הנתונים. מדובר על מאמץ של מספר ימי עבודה לאריזת הניסוי אבל אחר כך אפשר להריץ אותו על איזה קומביניציות של חומרה (שרתים אחסון תקשורת) שרוצים והדבר מאפשר קבלת החלטות טובה לגבי sizing של תשתיות.

לקוח גם הפנה אותנו ל- package של אורקל שמאפשר למדוד ביצועי מערכת. מדובר על http://docs.oracle.com/cd/B28359_01/appdev.111/b28419/d_resmgr.htm :

CALIBRATE_IO Procedure

This procedure calibrates the I/O capabilities of storage. Calibration status is available from the V\$IO_CALIBRATION_STATUS view and results for a successful calibration run are located in DBA_RSRC_IO_CALIBRATE table.

Syntax

```
DBMS_RESOURCE_MANAGER.CALIBRATE_IO (  
  num_physical_disks IN PLS_INTEGER DEFAULT 1,  
  max_latency IN PLS_INTEGER DEFAULT 20,  
  max_loops OUT PLS_INTEGER,  
  max_mbps OUT PLS_INTEGER,  
  actual_latency OUT PLS_INTEGER);
```

Parameters

Table 105-3 CALIBRATE_IO Procedure Parameters

Parameter	Description
num_physical_disks	Approximate number of physical disks in the database storage
max_latency	Maximum tolerable latency in milliseconds for database-block-sized IO requests
max_iops	Maximum number of I/O requests per second that can be sustained. The I/O requests are randomly-distributed, database-block-sized reads.
max_mbps	Maximum throughput of I/O that can be sustained, expressed in megabytes per second. The I/O requests are randomly-distributed, 1 megabyte reads.
actual_latency	Average latency of database-block-sized I/O requests at max_iops rate, expressed in milliseconds

אורקל על VMWARE

מספר לקוחות דיברו על כך שהם מיישמים אורקל על VMWARE בצורה סבירה. ואפילו דובר על מימוש RAC מעל VMWARE (רק באורקל גרסא 11.3 ומעלה). במידה ואפשר לבצע קונסולידציה של מערכות אורקל על חוות VMWARE ייעודית מדובר על חסכון גדול ברישוי.

הלקוחות מודעים לתמיכה הלא מלאה של אורקל על VMWARE אך בפועל לא נתקלו במצב שבו נפגעה התמיכה בפועל בגלל עבודה על VMWARE (לא היו צריכים לעביר מסד נתונים מ- VMWARE למכונה פיזית לצורך טיפול בתקלה).

עם זאת במקרים שהמשאבים שהאורקל דורש משאבים הקרובים לגודל שרת פיזי מלא, או במקרים שבהם ישנן דרישות מיוחדות משרת האורקל (כמו cluster בתצורות שונות), יכולות DR או כל יכולת שגורמת לשרת הוירטואלי שלא יהיה סטנדרטי – הדבר אומנם אפשרי אבל מקשה תפעול חוות ה-VMWARE.

תגובות של ספקים

EMC

חברות EMC ו-VMware מאמינות בקונספט של Software Defined Data Center כרעיון המורכב משלוש תכונות עיקריות: סטנדרטיזציה, וירטואליזציה ואוטומציה של משאבים. על מנת להגיע למצב אידיאלי של ניהול ה-Data Center בתוכנה, יש צורך בוירטואליזציה של כלל הרכיבים ומכאן שבשלב כזה או אחר המטרה היא שכלל מערכות ה-Data Center ירוצו על פלטפורמה וירטואלית.

באופן ספציפי, כפי שרואה EMC - המטרות העיקריות בביצוע איחוד מסדי נתונים (Consolidation) ומעבר של Oracle לסביבה ווירטואלית הן:

- הפחתת עלויות (OPEX & CAPEX) באופן כללי, ורישוי תוכנה בפרט.
- שיפור רמת הביצועים (Performance).
- שיפור הזמינות (Availability).
- מניעת זמן השבתה (Downtime).
- הפחתת תלות בחומרה והמגבלות הנגזרות מכך.

הוכחה לכך, ניתן למצוא בתוצאות מחקר שנערך ע"י Wikibon: [Virtualization of Oracle Evolves to Best Practice for Production Systems](#) שהשווה בין שימוש בתשתיות מסורתיות פיזיות של Oracle לבין תשתיות מבוססות EMC ו-VMware ששמו דגש על וירטואליזציה ושימוש בטכנולוגיה ה-Flash (XtremSF) וה-FAST VP Storage Tiering של EMC.

מחקר זה הוביל לשתי מסקנות חד משמעיות:

1. הפחתת עלויות בצורה משמעותית ע"י שילוב של –
 - i. קונסולידציה שרתים
 - ii. שימוש בשרתים המתקדמים ביותר מבחינת מעבדים וזיכרון
 - iii. חיזוק מערכי האחסון באמצעות שימוש בטכנולוגיות ה-Flash ו-Auto Tieringשילוב העקרונות הללו מוביל לאופטימיזציה תשתיות מסדי הנתונים של Oracle כך שפחות מעבדים נדרשים וכך נחסכות עלויות בצורה משמעותית.
2. שיפור ביצועים בסביבות בסיסי הנתונים ע"י –
 - i. הגדלת וייעול הזיכרון בשרתים
 - ii. שימוש ב-Flash במערכי האחסון
 - iii. שימוש בטכנולוגיה נידוד אוטומטי של הנתונים בין שכבות הדיסקים השונות

בהתבסס על מחקר זה, בנינו כלי [Oracle TCO Calculator](#) שיכול לאמוד את החיסכון בעלויות במימוש פתרונות משולבי EMC ו-VMware.

לסיכום – פתרון משולב של וירטואליזציה בסיסי הנתונים של Oracle, המבוסס על פלטפורמת EMC זה הפתרון המוביל מהסיבות הבאות:

- שיפור וייעול רישוי Oracle: מקסום כוח העיבוד על בסיס מספר מעבדים קטן יותר.
- שיפור הביצועים: יכולת בשימוש המעבדים העדכניים ביותר (נובע גם מהיכולת להפחית את התלות בחומרה), אפשרות של פעולות בסביבה הווירטואלית שמטרתן ניווד עומסי העבודה בין מכונות ושרתים שונים גם ע"י מימוש VMware DRS, שימוש בפתרונות מבית EMC כדוגמת XtremSF/SW ו-FAST VP בשכבת האחסון.
- זמינות: העלאת הזמינות תוך שימוש בכלי VMware HA, כדוגמת VMware HA (למתקדמים - יש גם אפשרות לממש Stretch Active cluster עם VPLEX)
- מניעת השבתה בזמן תחזוקה: ע"י שימוש בטכנולוגיית vMotion.
- הפחתת תלות בחומרה: שימוש בטכנולוגיית vMotion מאפשר להעביר בצורה פשוטה את מערכת ההפעלה ממערכת HW אחת למערכת HW אחרת.

EMC מציעה כיום מגוון פתרונות מתוך הסתכלות על עולם ה-Data Center כיחידה אחת המספקת שירותים לכלל האפליקציות הארגוניות והמטרה היא לצאת מפתרונות הבנויים כאיים וכפתרונות סגורים.

שני הפתרונות הבאים מציעים מעטפת מקיפה או לצורך העניין פתרון כולל לסביבות בארגון:

VBlock הוא מוצר של מיזם משותף של החברות: EMC, CISCO, VMware. מוצר זה נמכר דרך "חברת בת" הנקראת VCE שמטרתה לייצר, למכור ולתמוך במוצר זה. למוצר זה יש אחריות ממקום אחד על סך כלל המרכיבים כך שללקוח אין צורך לפנות לכל חברה בנפרד במקרה של תקלה ולא פחות חשוב - ה VBlock מגיע מותקן מהמפעל. כלומר, מדובר בפתרון מקיף וכולל ולכן כל הרכיבים שלו הם אינטגרטיביים החל מהארון המיוחד עם זרימת האוויר עד לרכיבים הלוגיים כגון תוכנת הניהול המיוחדת שלו הנקראת UIM ויכולה להקצות Storage, להתקין שרתי ESXi ולהקצות רשתות תקשורת - הכל מתוכנה אחת. בנוסף, התוכנה מנטרת את כלל המרכיבים ב VBlock, מספקת Root Cause Analysis ולמעשה מקצרת את הזמן הנדרש היום בניטור מקור התקלה.

פתרון נוסף הוא ה-VSPEX שלמעשה מהווה מגשר בין העולם המלא של ה VBlock לבין Reference Architecture (RA) שבו הלקוח צריך להתקין ולתמוך בכל הרכיבים בעצמו. הפלטפורמה גמישה (Open Platform) כך שנותנת אפשרות לשילובים שונים של רכיבי תקשורת, שרתים, מערכי אחסון ורכיבי וירטואליזציה ובכך בעצם מאפשרת ללקוח את האופציה לבחור בין מגוון ספקים וחומרות.