

ה-Big Data מאיים על פרדיגמות האחסון הקיימות

"מזה עשור אנחנו מדברים על אחסון מרכזי, אבל הטכנולוגיות החדשות, שמגיעות מהעולם האינטרנט, צפויות לגרום שינוי בתצורה זו ולהביא לאחסון מבוזר", אמר פיני כהן, סגן נשיא ואנליסט בכיר, STKI ♦ לדבריו, הטיפול ב-Big Data יעשה על ידי שימוש בבסיסי הנתונים בתחום ה-noSQL, "שהדור הקודם של בסיסי הנתונים אינו יכול לספק להם מענה"

יוסי הטוני < צילום: ניב קנטור

בטכנולוגיה זו יש מחיר: "הבעיה היא שבעזרתה לא נעשה עדכון מלא ומהיר לנתונים הנמצאים במקומות שונים, כי זה לוקח זמן. לכן, הטכנולוגיה אינה מתאימה, למשל, לטיפול ביתרות בחשבון עובר ושב".

לצד הטכנולוגיות הללו, ציין כהן, יש בעולם ה-Big Data מכונות בסיסי נתונים לנפחים גדולים - של יבם, אורקל, סאפ ו-EMC.

מגמות נוספות שציין כהן הן Scale Out Storage - "לקחת מכונות אחסון עובדות ולחבר ביניהן"; ההצטרפות של האחסון; ווירטואליזציה של האחסון, עליה הוא אמר ש-"אף על פי שזו טכנולוגיה מביחה, היא טרם נכנסה לשוק המקומי".

כהן סיים בהציגו מחקר שערכה החברה שממנו עולה, כי כל עובד ומנהל תשתיות ואחסון טיפל ב-2011 ביותר נפח אחסון לעומת 2010. "מצד אחד, עולם האחסון כמנהגו נוהג, ומצד שני, אנחנו מבחינים באופק במגמות שיטנו את האחסון באופן מהותי", סיכם.

המידע הלא מובנה תופס נפח גדול מאוד בעולם האחסון

"התעשייה, הלקוחות וספקי ה-IT הבינו כבר מזמן שהמידע הבלתי מובנה הולך ותופס נפח גדול מאוד בעולם האחסון, ונדרש לטפל בו", כך אמר **שגיב קרונבלט**, מהנדס פתרונות בהיטאצ'י.

בדבריו הציג קרונבלט את האתגרים הניצבים בפני מנהלי התשתיות בעולם האחסון כיום: הצורך בניהול מידע מרובה סביבות; עלויות האחסון הגדולות משנה לשנה; היותה של סביבת ה-IT מרובת איי מידע והעובדה שהחיפוש אחריו גוזל זמן יקר; הצורך בהגנה על נתוני המידע; קיומם של צווארי בקבוק וירידה בביצועים בהעברת נתונים; והעדרה של ראייה כלל מערכתית על המידע המאוחסן.

"אנחנו מציעים פתרון אחד, סביבה שלמה, שמנהלת את כל המידע הבלתי מובנה תחת שכבת אחסון ווירטואלית אחת, בעלת ביצועים גבוהים", אמר קרונבלט. "האסטרטגיה שלנו היא לספק פלטפורמה אחת אחודה, עם ניהול ממקום מרכזי אחד".

הוא הוסיף, כי "הפתרון מתבסס על איחוד הנתונים במבנה אחיד, ארכוב ושמירה של הנתונים לאורך זמן, הגנה על הנתונים בכל נדבך ונדבך ויכולת חיפוש של הנתונים לפי כללים ודרישות". לדבריו, "הפתרון שלנו מאפשר איחוד של כלל נכסי המידע, ניהול פשוט, הגדלת ניצולת המערכות, הגנה על הנתונים וכל אלה יחדיו, בעלויות מופחתות".



פיני כהן



שגיב קרונבלט

"אנחנו מבחינים בהמשך המגמות הרגילות בעולם האחסון - הגידול האקספוננציאלי בנפחי המידע, ההצטרפות של רכיבי האחסון, ואחרות, אבל המגמה המעניינת, על אף שטרם בשלה, היא כניסת ה-Big Data. מגמה זו מאיימת לשנות את הפרדיגמה של עולם האחסון כפי שהכרנו עד היום", אמר **פיני כהן**, סגן נשיא ואנליסט בכיר ב-STKI.

כהן דיבר בפתח כנס Storage2012, שהופק על ידי אנשים ומחשבים ונערך במלון דיוויד אינטרקונטיננטל בתל אביב בהשתתפות מאות מקצוענים מעולמות התשתיות והאחסון. הנחה את הכנס יהודה קונפורטס, העורך הראשי של הקבוצה. לדבריו כהן, "מזה עשור אנחנו מדברים על אחסון מרכזי, אבל הטכנולוגיות החדשות, שמגיעות מהעולם האינטרנט, צפויות לגרום שינוי בתצורה זו ולהביא לאחסון מבוזר". הוא הוסיף, כי הגדרת ה-Big Data נוגעת לכמה היבטים: היקף של עשרות טרה בייט; באיזו מהירות הנתונים מתעדכנים; מגוון סוגי המידע - מובנה ולא מובנה, כאשר "העלייה הגדולה היא במידע הלא מובנה"; עושר המידע; והאפשרות לעבוד עם סכמות משתנות.

טכנולוגיות לעולם ה-Big Data

הוא אמר, כי הטיפול ב-Big Data יעשה על ידי שימוש בבסיסי הנתונים בתחום ה-noSQL, "שהדור הקודם של בסיסי הנתונים אינו יכול לספק להם מענה". לדבריו, מימוש פתרון זה מיועד לעוצמות מיחשוב גדולות, עם מחשבים ונתונים רבים, שלא ניתן לעשות בבסיסי הנתונים הקנייניים. בסיסי הנתונים הללו, אמר, הם בעלי סכמה גמישה וניתנת לשינוי ולגידול. כהן ציין כמה מהפתרונות בעולם ה-noSQL, ביניהם קסנדרה ו-MongoDB. הוא הוסיף, כי "שימוש בבסיסי נתונים אלה, לצד טכנולוגיית Hadoop, יביא לכך שלא תהיה הסתמכות על אחסון מרכזי אלא על אחסון מבוזר בשרתים. זו פנייה של 180 מעלות ממה שהיינו רגילים לו בעשור האחרון. רוב מנהלי התשתיות לא מכירים את התחום החדש ועליהם להתוודע אליו, כי Big Data הוא דבר מהותי לעולם האחסון".

כך, אמר, חברה בשם אורביטז נעזרה בטכנולוגיית Hadoop לטיפול בנפחי האחסון שלה, בהיקף 750 טרה-בייט, תוך שאנשי התשתיות עובדים על מערכת אחסון אחת. "Hadoop היא טכנולוגיה אשר תומכת בעד 4,000 שרתים העובדים יחד ועד 20 פטה-בייט נפח על אשכול אחד", אמר כהן. הוא ציין, כי לשימוש