

דקה אחת באינטרנט

מאת: יאיר גני, TNN - SGI

אינו נדרש בזמינות מהירה, אך הוא חייב להישאר זמין. ישנו סיכוי סביר שבכל זאת נצטרך להעזר בצילום. בנוסף, רגולציות, בחלקן חדשות, מכתיבות לא פעם כמה זמן יש לשמור מידע ולפעמים גם כמה זמן הוא צריך להיות. דוגמאות נוספות למידע ש"מתקרר" עם הזמן: תצלומי אוויר ולווין, צילומי המחאות בבנק, הסכמים וחוזים סרוקים וכו'.



יאיר גני

SGI הינה מחלוצות הטיפול במידע קר. לפני יותר מ-20 שנה פיתחה החברה את מערכת DMF - Data Migration Facility שמטרתה שינוע אוטומטי של אינפורמציה בין פלטפורמות אחסון שונות. זו הייתה מערכת HSM - Hierarchical Storage Management הראשונה מסוגה. זוהי מערכת Multi Tier התומכת במגוון גדול של אמצעי אחסון כגון פלאש, דיסקים מסוגים שונים, מערכות MAID, קלטות בתצורות שונות ואמצעי אחסון ברשת. המערכת דואגת באופן אוטומטי לחלוטין להעביר מידע מאמצעי לאמצעי - תוך יישום מדיניות האתר. כלפי המשתמש, או היישום שמשמש בנתונים, המידע תמיד זמין וקיים. אין צורך בהתערבות חיצונית כדי לשחרר מידע מגיבוי או כל אמצעי חיצוני אחר. רבים ממאגרי המידע הגדולים בעולם משתמשים

זה שנים רבות במערכת DMF של SGI כדי לאחסן בצורה נוחה כמויות גדולות של מידע הפרושים על אמצעי אחסון בעלי דרגות זמינות שונות. לדוגמה, סוכנות החלל האמריקאית (NASA) מחזיקה מעל 60 פטהבייט, מערך חיזוי האקלים האמריקאי (NOAA) מחזיקה מעל 35 פטהבייט, אולפני ווטה דיגיטל (זוכרים את שר הטבעות?) מאחסנים כ-24 פטהבייט. לאחרונה, הכריזה SGI על מוצר חדש - InfiniteStorage Gateway שהוא שילוב של תוכנת DMF יחד עם מערכת אחסון חדשה מבין SGI. השילוב בין השניים יוצר Appliance אטרקטיבי המאפשר הטמעה מהירה וזולה של מערכת אחסון Multi-Tier. המערכת תוכננה בישראל.

Storage 2013 בכנס

חשבתם פעם מה קורה בדקה אחת ברחבי האינטרנט? כמה מידע עובר? כמה מידע נאגר ומאוחסן?

לפי הערכות שנעשו ע"י חברת אינטל, בכל דקה מתבצעות הפעולות הבאות על גבי האינטרנט:

- 320 חשבונות מתווספים לטוויטר
- 100,000 ציורים עוברים בטוויטר
- 3,000 העלאות תמונה לפליקר
- 20,000,000 צפיות בתמונות בפליקר
- 30 שעות של וידאו חדש נוספים ליוטיוב
- 1,300,000 צפיות וידאו ביוטיוב
- מעל 2 מיליון חיפושיים בגוגל
- 6,000,000 צפיות בפייסבוק
- 100 חשבונות חדשים נוצרים בלינקדין
- 47,000 אפליקציות מורדות מה- AppStore של Apple
- מעל 200 מיליון אימיילים נשלחים
- 6 ערכים חדשים מתווספים לוויקיפדיה
- 1,300 משתמשים חדשים מצטרפים לעולם המובייל
- 20 גניבות זהות מתבצעות

כמובן שאלו לא הפעולות היחידות שמתבצעות על גבי האינטרנט - אך די באלה כדי להמחיש את כמות האינפורמציה שזורמת סביבנו בכל רגע נתון ויותר מכך - כמות האינפורמציה שנאגרת בכל רגע.

אנחנו לא אוהבים למחוק. הטבע האנושי גורם לנו לאגור. למשל, יש לנו לא מעט תמונות כפולות על המחשב האישי, אולי גם על עוד לפטופ ובטאבלט, העלנו אותן אולי לפיקאסה וחלקן גם לפייסבוק, גיבינו אותן בדראפובוקס ואולי גם בענן החדש של בזק. גם סרטי הווידאו והמוסיקה שלנו אינם טומנים ידם בצלחת.

מבחינת זמינות - לא כל האינפורמציה היא בעלת אותו ערך. יש מידע שחשוב שיאוחסן על אמצעי אחסון מהירים מאוד (כגון כונני פלש או SAS), יש מידע שיכול לשבת על אמצעים אחסון בינוניים (כגון כונני SATA) ויש מידע שיכול לשבת אפילו על אמצעי גיבוי כגון קלטות. המידע בטוח באותה מידה על כל אחד מהאמצעים האלו - אבל מהירות הגישה למידע משתנה מאמצעי לאמצעי.

ניקח לדוגמה חברת תקשורת או מוסד רפואי המטפלים במאגר לקוחות גדול ובכמויות גדולות של תנועות יומיות. התנועות היומיות האחרונות דורשות כמובן את הזמינות הגבוה ביותר - לצורך עיבוד שוטף, התנועות של השנה האחרונה יכולות להיות בעלות זמינות נמוכה יותר. תנועות משנים קודמות יכולות להסתדר עם הזמינות הנמוכה ביותר. לא תמיד רמת הזמינות נקבעת רק לפי גיל המידע - אבל זו דוגמה מצוינת ונפוצה. המוטיבציה להעביר מידע לאמצעים עם זמינות נמוכה יותר היא בעיקר כלכלית. ככל שאמצעי האחסון מהיר וזמין יותר - הוא גם יקר יותר, משמעותית.

חלק גדול מהאינפורמציה האגורה הופך במהרה למידע קר. אז מה זה מידע קר ומתי בדיוק התחלנו למדוד חום למידע?

מידע קר הינו מידע שערך הזמינות שלו נמוך מאוד, אך הוא עדיין נדרש במערכת ולא ניתן להוריד אותו לאמצעים חיצוניים (offline data) לדוגמה - צילום רנטגן או אולטראסאונד שנעשה לפני שנה או שנתיים.

