



מחשוב ענן: זה לא בשמים

הגידול המואץ בכמות המידע והצורך לאחסן מידע זה כך שיהיה בטוח וזמין, מחייבים מחקר טכנולוגי נרחב. זהו אחד ממוקדי הפעילות של המרכז להנדסת מחשבים בטכניון (TCE) | אסף צור

חזור מהיר; בשיטה השנייה מגבילים את הגישה למספר קטן של דיסקים, קוראים אותם במלואם וכך מנסים לשחזר את המידע.

אחת ממשפחות הקודים היעילים היא "קוד די הזיג זג", שהומצאו על ידי פרופ' שוקי ברוק מאוניברסיטת קלטק, המשמש מרצה אורח בפקולטה להנדסת חשמל בטכניון; פרופ' יצחק תמו מאוניברסיטת תל-אביב ופרופ' זיינג ווינג מאוניברסיטת קליפורניה אירוויין. שיטה זו, אף שהדרך עוד ארוכה למימושה התעשייתי, היא דו-גמא למחקר תיאורטי אוניברסיטאי העשוי לפתור בעיות מעשיות מאוד בתעשייה.

אתגרים חדשים

נושא האחסון והגיבוי הוא גם תחום אקדמי חשוב עבור אנשי big data, שכן כאמור כמות המידע המצטבר היא עצומה ברמה יומית ופתרון נותן האחסון, שלא יכולים להמשיך לגדול באותה מידה, חייבים להתייעל. היכולת להבחין ולהחליט מהו מידע נכון ומהו מידע בעייתי עומדת בבסיס המחקר הזה, המסתכל על כמות נתונים שמגיעה לשרת במהירות ובכמויות אדירות, ויש לגבות אותה באותה מהירות כמעט.

המשך הגידול בכמויות המידע העצומות הנאגרות תמשיך ליצור אתגרים רבים ומעניינים בתחום אחסון המידע (storage), החל מהרמה הפיזית של ייצור זכרונות מהירים וזולים ועד להקמה וניהול של מערכות אחסון אלו. אין ספק שאתגרים אלו יפתחו דלתות להמשך מחקר פורה ופתרונות חדשניים הן באקדמיה והן בתעשייה.

של 'נפילות' דיסקים, מסביר פרופ' משנה איתן יעקובי, חבר סגל בפקולטה למדעי המחשב בטכניון. פרופ' משנה יעקובי, שהשלים את הדוקטורט שלו באוניברסיטת קליפורניה סאן דייגו, מתמחה בתחומי תורת האינפורמציה ותורת הצפינה ומתמקד באפליקציות לזיכרונות מחשב, בזיכרונות אסוציאטיביים, באחסון מידע ובאחזורו ובתורת ההצבעות.

"בין מהירות השיחזור לעלותו קיים טרייד-אוף ברור", מסביר פרופ' משנה יעקובי, "ושתי 'משפחות השיחזור' העיקריות הנחקרות כיום הן regenerating codes ו-locally repairable codes. השיטה הראשונה מאפשרת לנו לגשת למספר גדול של דיסקים בבת אחת (שיחזור מהיר), אבל מכל דיסק קוראים רק חלק קטן כך שסך כמות המידע שנקראת היא קטנה ומאפשרת שי-

רמז משמו, אלא פלטפורמה פיזית המתאימה לאחסון מידע רב. בדרך כלל מדובר במחסנים גדולים של data centers, המכילים אלפי דיסקים קשיחים וכן טייפים וכוונני מצב מוצק. לפיכך, מדובר בתשתיות שהקמתן ותחזוקתן דורשות השקעה כספית עצומה. המחסנים האמורים חייבים להיות אמינים לגמרי. כלומר, המידע חייב להיות מגובה, או לכל הפחות שמור באופן שיבטיח את שחזורו בזמן נפילות. אופציית הש"כפול - יצירה של דיסקים "כפילים" לכל דיסק - אינה כלכלית ואינה יעילה כיום, וזאת בשל כמות המידע העצומה.

שיחזור מידע מהיר

"בשנים האחרונות הוצעו פתרונות קידוד שונים, שמטרתם להפחית את זמן השחזור במקרה

מחקר מעשי ויישומי

המרכז להנדסת מחשבים (TCE) הוא מרכז מחקר יישומי משותף לפקולטות להנדסת חשמל ולמדעי המחשב בטכניון. שותפים בו חברי סגל משתי הפקולטות, מפקולטות נוספות בטכניון ואף מאוניברסיטאות אחרות. המרכז, שבו שותפות גם חברות תעשייתיות, מקדם מחקר מעשי ויישומי ומאפשר דיאלוג שוטף בין התעשייה לאקדמיה בנושאים המעניינים את שני הצדדים. בין היתר עוסק המרכז בהפצת ידע ועורך כנסים רבים, ובהם הכנס השנתי המתקיים במאי ועוסק בתחומים החמים ביותר בהנדסת מחשבים. הכנס האחרון עסק במערכות מבוזרות, באימות קוד ובעיבוד מידע רב. המרכז מקדם מחקר במגוון נושאים הקשורים לענן, ובהם רשתות מחשבים, סייבר, איחוד סונו ודחיסת מידע ועוד. לדברי רות בונה, מנהלת המרכז, "כל חברה הנתקלת באתגרים מחקריים, ומעוניינת להיעזר בכוח המחקרי החזק הנמצא בטכניון, מוזמנת לפנות אלינו".

המידע הדיגיטלי בעולם גדל בקצב מטר-רף; מדי יום נוספים למאגר המידע הווירטואלי הזה 2.5 קוונטיליון ביטים של נתונים, כלומר 2.5 כפול 1018. גידול הנתונים דרמטי כל כך, ש-90% מהמידע הקיים כיום נוצר בשנתיים האחרונות. לשם המחשה, אפשר להשוות על אוכלוסיית העולם הנוכחית (7.4 מיליארד איש) ולדמיין ש-90% מתוכה - 6.6 מיליארד בני אדם - הם תינוקות שנולדו בשנתיים האחרונות. המידע נולד באין-ספור מקומות ופלטפורמות - בחיילים שאוספים מידע על שינויי מזג אוויר, ברשתות החברתיות (טקסטים, תמונות וקטעי וידאו), ואפילו במערכות GPS המתעדות את מקומיהם של אנשים וכלי רכב - ונשמר במקומות שונים. אף שעלויות האחסון פוחות בהתמדה, עדיין יש להן משמעות רבה. זו הסיבה להתפתחותו של "מחשוב ענן" (Cloud Computing) - גישה שבה המידע עולה לענן, כלומר לרשת, וממנו הוא זמין למשתמשים.

בעשור האחרון משמש הענן לא רק לשמירת מידע, אלא גם כמרחב שבו פועלות אפליקציות ותוכנות. שינוי זה, המוריד מהעומס על המחשבים והסמארטפונים, כרוך באתגרים טכנולוגיים גדולים שעומדים במודדים בעיקר ספקי השירות הגדולים ובהם גוגל, אמזון, מיקרוסופט, פייסבוק, יבמ ודרופבוקס. חברות אלה נדרשות לספק שירותי ענן מהירים, אמינים ובטוחים, שיהיו גם זולים מספיק כדי להוות חלופה אטרקטיבית לרכישה של שרתים מקומיים. הענן, ראוי לציין, אינו מרחב וירטואלי כנ-