

נושאים מתקדמים במערכות אחסון ל Big Data (קוד : 18103701)		שם הקורס בעברית
Advanced Storage Systems for Big Data		שם הקורס באנגלית
קורס בחירה 3 נ"ז	תואר שני פתוח גם למצטיינים מתואר ראשון	תשפ"ב סמסטר א' שלישי 17:15-19:45
שעת קבלה בתיאום מראש	מייל dalit.naor@mta.ac.il	שם המרצה פרופ' דלית נאור

פרטי הקורס

<p>התפוצצות המידע הדיגיטאלי – הן הארגוני והן האישי – הינו אחד מהמאפיינים הבולטים של שני העשורים האחרונים. מערכות מחשוב, ומערכות אחסון נתונים וניהול מידע בפרט, נדרשו לתכנון וחשיבה מחדש על מנת להתמודד עם הגידול, אופי המידע השונה והשימושים החדשים במידע בעקבות מהפיכת ה AI. בקורס זה נסקור את העקרונות הבסיסיים של מערכות אחסון (storage) ונעסוק במגמות שהתפתחו בשיטות אחסון ובאחסון ענן לטיפול במערכות BigData. מגמות אלה חוללו מהפיכה ארכיטקטונית באופן שבה נבנות מערכות כיום. נתמקד במבנה הארכיטקטוני של מערכות אלה והמודלים החישוביים המבוזרים הרלוונטיים לבנייתם. כמו כן נלמד כיצד הופעתן של חומרות חדשות לשימור ביטים (כגון flash, NVM ו DNA – Storage) השפיעו וישפיעו על האופן שבו מערכת ה Storage נבנית. תחום אחסון המידע הדיגיטלי הינו נושא מאד מרכזי וחשוב בעולם המחשוב, אך כמעט ולא נלמד במערכת האקדמית למרות מורכבותו הטכנית והאלגוריתמית. מטרת הקורס לגשר בין העולמות ולהכשיר את הסטודנטים בכלים הבסיסיים הנדרשים להבנת מערכות אלה.</p>	<p>תקציר בעברית</p>
<p>The explosion of business and personal data is one of the most fundamental trends we have experienced over the last two decades. IT systems, and storage in particular, had to be re-designed in order to address the new challenges imposed by the data explosion and the AI revolution. This course will review the basic principles of storage systems, and how they evolved over the past 20 years in order to handle BigData workloads in Cloud environments. We will focus on the architectural principles of these new distributed systems and the algorithms behind them. In addition, we will learn about new types of media that are continuously invented to store bits of information (e.g. flash, NVM and DNA storage) and how the introduction of new media types changes the software stack of the storage system that is built above it.</p> <p>Storing digital information is critical and complex in today's computing environments, but receives little attention in computer science academic curriculum. The purpose of the course is to bridge the gap between the two worlds, and equip students with tools needed to understand storage systems.</p>	<p>תקציר באנגלית</p>

<p>"תכנות מתקדם בשפת C", "תכנות מכוון עצמים ושפת ++C", מערכות הפעלה.</p>	<p>דרישות קדם</p>
--	-------------------

ידע קודם: מבני נתונים, מערכות הפעלה ומערכות קבצים (מומלץ)	
<p>הקורס יינתן באנגלית (הרצאות, שקפים ומטלות) כחלק מפיילוט שעוברת האקדמיה להנגשת תכנים אקדמיים מקצועיים באנגלית. מכיוון שרוב העבודה הטכנית בתעשייה מתרחשת באנגלית, הקניית הידע בשפה האנגלית תביא ערך מוסף לסטודנט.ית.</p> <p>הקורס יוגש בשיעור פרונטלי של 3 שעות שבועיות, וילווה בשקפים. יתכנו מספר הרצאות אורח של מומחי תוכן מהתעשייה.</p>	שיטת ההוראה
<p>במהלך הקורס יינתנו כ- 3-4 תרגילים, חלקם עיוניים וחלקם מעשיים. אחד מהתרגילים ידרוש לימוד עצמי ו/או העמקה במאמר או קוד פתוח. כל התרגילים הם חובה ואישיים.</p> <p>בסיום הקורס תתקיים בחינה על כל החומר שנלמד בשיעור ובתרגילי הבית.</p>	דרישות הקורס/מטלות
<ul style="list-style-type: none"> ● תרגילים : 50% ● בחינה : 40% ● השתתפות בשיעורים ותרומה : 10% 	הרכב הציון
<p>נוכחות חובה ב 10 מתוך 13 ההרצאות, אך חובת נוכחות בכל הרצאות האורח. הנוכחות גם ביתר ההרצאות מומלצת מכיוון שאין מקור מסודר שמנגיש את החומר ושעליו ניתן להתבסס למטרות לימוד עצמי.</p>	נוכחות

תכנית הקורס - נושאי הלימוד לפי שבועות

יתכנו שינויים שיפורסמו בהמשך

1. Overview of data persistence and storage: review of the basic notions of I/O systems, file systems, caching.
2. The basics of traditional storage media: hard drive (disk), tape, RAID
3. The evolution of storage architectures: The characteristics of data explosion, new access methods to storage, Block/File/Objects, the journey to the Cloud
4. The Cloud paradigm and Cloud storage.
5. Distributed storage architectures: CAP theorem, consistent hashing and the Chord system, cloud caching.
6. Data Reduction in storage systems: compression and deduplication
7. Solid State Disks (flash)
8. New types of media and implications on the storage system: DNA storage, Persistent memory (NVM).
9. Above the storage: Modern data management systems and Data Lakes

רשימת קריאה ומקורות נוספים חלקיים רשות :

1. *Operating Systems: Three Easy Pieces*. By Remzi H. Arpaci-Dusseau and Andrea C. Arpaci-Dusseau
2. *Information Storage and Management: Storing, Managing, and Protecting Digital Information in Classic, Virtualized, and Cloud Environments*, 2nd Edition. EMC Education Services (Editor)

3. **CAP theorem.** Eric A. Brewer. *Towards robust distributed systems*, PODC 2000. *CAP Twelve Years Later: How the "Rules" Have Changed*, Eric A. Brewer. Computer 2012.
4. **File systems.** *The Hadoop Distributed File System (Yahoo!)*, MSST 2010. *The Google File System (Gobioff)*, SOSP 2003. *Case Study GFS: Evolution on Fast-forward*, ACM 2009
5. **Chord peer-to-peer and DHT.** *Scalable Peer-to-Peer Lookup Service for Internet Applications*. Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan. ACM SIGCOMM Computer Communication Review, August 2001
6. **Content Delivery.** *Algorithmic Nuggets in Content Delivery*. Bruce M. Maggs and Ramesh K. Sitaraman. ACM SIGCOMM Computer Communication Review, 2015.
7. **DNA Storage.** *A DNA-Based Archival Storage System*, ASPLOS 2016.