שם הקורס: סמינר במערכות אחסון מידע
Seminar on Storage Systems: Deduplication
מספר הקורס 236834
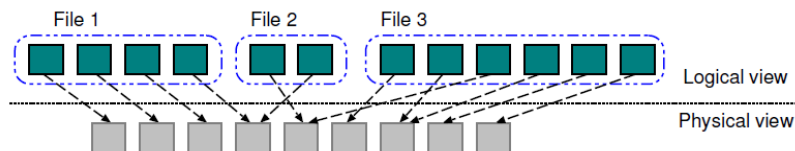
מרצים:		דר' גלה ידגר
שעות הרצאה:		יום ב' 14:30-12:30
דרישות קדם:		מערכות הפעלה (234123) - חובה מערכות אחסון (236322) ו/או תכנות מקבילי ומבוזר (236370) - יתרון

תאור הקורס

> *"The **Global Datasphere** is a measure of all new data that is captured, created, and replicated in any given year across the globe… IDC predicts that the Global Datasphere will grow from 33 Zettabytes in 2018 to **175 Zettabytes by 2025**... Over **22ZB of storage capacity must ship** across all media types from 2018 to 2025 to keep up with storage demands."*
>
> <u>Data Age 2025, by Seagate</u>

Data deduplication is one for the most effective ways to reduce the size of data stored in large scale systems, and is widely used to date. The process of deduplication consists of identifying duplicate chunks of data in different files (including backup versions, virtual machine images, etc.), storing a single copy of each unique chunk, and replacing the duplicate chunks with pointers to this copy.

The goal of this seminar is to study the basic concepts of deduplication, the challenges and tradeoffs it introduces in the design of large scale storage systems, and the most recent advances in addressing these challenges. We will cover the various aspects of system design, including chunking, duplicate identification, metadata and reference management, inline vs. background deduplication, centralized vs. distributed deduplication, hard disk vs. SSD storage, and more.

The list of papers will be published shortly before the beginning of the semester.

דרישות הקורס

The course will combine lectures by the instructors with independent reading in a seminar format. The students will read important papers in the field and will present them in class along with their background. Students will be expected to familiarize themselves with the each week's paper prior to class.

רישום

Please email your name and ID to Gala Yadgar. Please state whether you are a graduate or undergraduate student (which semester), related courses you took, and other relevant information.