Modern Web-scale applications (e.g., Facebook, Twitter, Google Docs) must face technical challenges that derive from their scale:

- Scalability: the possibility to grow as the user-base and data-size handled by the application grows to hundreds of millions of users and petabytes of data.
- High-Availability: the capacity to provide service to users even when part of the infrastructure (CPUs, Networks, Disks) become unaccessible in an intermittent or permanent manner.

The way to address these requirements is to develop loosely distributed applications that can operate in a "cloud-like" runtime environment. This course introduces basic theory behind such massively distributed applications and modern programming tools that constitute an emerging infrastructure for distributed applications.
2

2.1 Caching

3.1 Map-Reduce

3.2 Join

3.3 Latent Semantic Indexing

4.1 Latent Semantic Indexing

Cloud

A.T.A.

K.-M.
לrespect to Zion the course

Activities:

Exercises: 30%
Project: 30%
Exam: 20%

Books:


Yoav Goldberg, Meni Adler and Michael Elhadad, EM Can Find Pretty Good HMM POS-Taggers (When Given a Good Start), ACL 2008.


Ryan McDonald, Keith Hall and Gideon Mann, Distributed Training Strategies for the Structured Perceptron, NAACL, 2010

Jing Gao and Jun Zhang, Clustered SVD strategies in latent semantic indexing, Information Processing and Management, 2005

Yang Liu, Maozhen Li, Mukhtaj Khan and Man Qi, A MAPREDUCE BASED distributed LSI for scalable information retrieval, Computing and Informatics, 33:2, 2014