Course Content
COSI 132a will convey knowledge of the principles and practices underlying today’s distributed and parallel data management systems. The course will first review fundamental distributed computing mechanisms, such as distributed databases, distributed and parallel data processing. It will investigate the issues that arise in the design and implementation of today’s web-based systems through abstract models, protocols, and case studies of widely used web applications and services. Finally, the course will address how these mechanisms and technologies fit together to realize existing and newly emerging web-based data management systems by discussing cloud databases, cloud data storage and computing services.

Learning Goals
The learning objectives of the class is to:

- Cover the basic concepts of distributed and web-based data management
- Provide the foundations of the development of parallel and distributed data processing systems
- Discuss the state of the art in distributed data storage and processing
- Provide students with hands-on experience in designing and building distributed and parallel data management services.

Outline of topics
- Databases in a Nutshell
- Distributed Databases
- Parallel Databases
- MapReduce
- Cloud computing

Audience
The course is addressed to upper-level undergraduate students as well as to graduate students that have solid background in programming, data structures and computer systems organization. Students are required to have taken COSI 131a (Operating Systems).
Prerequisites
COSI 131a or the equivalent. Background in database systems is also recommended but not required. First lectures will provide relevant database background.

Required Readings
Many lectures will be based on the textbook “Principles of Distributed Database Systems”, M.T. Oszu, P. Valduriez, Springer, 3rd edition. This is not a required textbook and an available pdf is posted on LATTE. The course will also rely on published papers and online resources. The instructor will also make available lecture notes/slides on the topics covered on class.

Schedule
Please see LATTE for tentative schedule.

Lecture Slides
Lecture slides will be posted on LATTE

Assignments
Please refer to LATTE for assignments, due dates, and submit all work via LATTE.

Logistics
Where: Abelson-Bass-Yalem Physics 126
When: Tuesday and Friday 11:00am-12:20pm

Professor
Olga Papaemmanouil
Office: Volen 214, x62716
Email: olga@cs.brandeis.edu
Office Hours: Tuesday 12:30-13:30pm and by appointment

TAs
TA: Chi Zhang, email: chizhang@brandeis.edu
Course Methodology

- Homework problem sets will test and improve your knowledge of the material. The intent is for you to use these as mini exams to self-assess your knowledge of the course material and evaluate your performance.
- Perhaps the most valuable part of this class will be the labs. They will require you to apply in practice the concepts covered in the lectures. You will gain experience with implementing in Java parallel processing algorithms (Map Reduce) as well as get hands on experience on using a interacting with a distributed data management system (Vertica). The labs will be done individually.
- There will be three in-class quizzes testing the mastery of the material covered in the course. Unless stated otherwise, the quizzes will be closed book and will cover material from lectures, readings, homework sets and labs.

Grading

15% Homeworks
40% Labs
45% Quizzes