

MIE1512H: Data Management for Analytics

Instructor: Mariano P. Consens (Email: consens@mie.utoronto.ca)

Web: <http://www.mie.utoronto.ca/staff/profiles/consens/>

Office: BA8128

Office hours: Tue 3pm-4pm (send email to make an appointment)

TA

TBD

Outline

This course is a research seminar that focuses on recent developments in the area of Data Management for Analytics. Science, businesses, society, and government are being revolutionized by data-driven methods that benefit heavily from scalable data management techniques.

The course provides an overview of data management concepts applied to analytics, covering methods and techniques, including distributed computations on massive datasets and frameworks for enabling large-scale parallel data processing on clusters of commodity servers. Emphasis is given to data management techniques for analyzing Web Data and Open Datasets. The course evaluation is based on student presentations, a focused bibliography survey, a hands on invigilated lab, and a course project (the last two using computational notebooks on scalable platforms). The project goal is to reproduce high quality published research in the area of data analytics, emphasizing data management aspects.

Pre-requisites: An undergraduate level course in Databases, such as MIE253 Data Modelling, or equivalent (including hands-on knowledge of SQL).

Course material (lecture slides and lab materials) are available on Quercus. The course slides posted on portal contain the references to the course textbooks and to the assigned reading materials covered each week.

Grading Scheme

<i>Invigilated Lab Test (notebooks)</i>	20%
<i>Class presentation and participation</i>	15%
<i>Course project and bibliography</i>	65%

Course Schedule

	Day	Time	Location
Lectures and Labs	TBD	TBD	RS303