MIRI Master (High Performance **Computing)**

masters.fib.upc.edu/masters/miri-high-performance-computing

Visiting Professors' Seminar, Spring 2017

SEMINAR

Performance Engineering (2 ECTS)

Prof. Emery Berger (University of Massachusets Amherst, USA)

Dates: July 3rd - July 14th , 2017

Place: UPC, Campus Nord, C6-E101

Registration: Please contact Xavier Martorell <xavim@ac.upc.edu> to indicate your interest on attending this seminar.

Description:

In this two-week course, students will learn the fundamentals behind large-scale data intensive systems often used in the context of data science. We will cover the abstractions and discuss the implementations of these systems, ranging from traditional SQL-based databases to "NoSQL" key-value stores like MongoDB and Redis, map-reduce frameworks like Hadoop, and analytics platforms like Apache Spark and Pregel.

Week 1:

- Databases
  - SQL
  - Query planning and optimization
  - Indexing, join implementations
  - Optimistic concurrency
  - Consistency guarantees
- NoSQL
  - Abstraction
  - Example implementations: Redis, MongoDB
  - Implementation, DHTs
  - Consistency guarantees or lack thereof

Week 2:
- Large-scale data analytics
  - Performance considerations vs. single-node
- Map-Reduce frameworks
  - Hadoop
  - HDFS
- Spark
  - RDDs, Apache Spark
  - SparkSQL
- Graph analytics
  - Pregel / Giraph
  - Arabesque

Teaching hours

July 3: 11:00 to 13:00. Lecture
July 4: 11:00 to 13:00. Lecture
July 5: 11:00 to 13:00. Lecture
July 6: 11:00 to 13:00. Lecture
July 7: 11:00 to 13:00. Lecture

July 10: 11:00 to 13:00. Lecture
July 11: 11:00 to 13:00. Lecture
July 12: 11:00 to 13:00. Lecture
July 13: 11:00 to 13:00. Lecture
July 14: 11:00 to 13:00. Lecture

Evaluation

Students will be evaluated, based on a research report on a specific topic agreed upon with the professor of the course.