

Resilient, Parallel, Big Data Application Frameworks in X10

Vijay Saraswat¹

¹IBM T. J. Watson research center
Email: vijay@saraswat.org

Timetable: 12 hrs. lectures in two weeks, spring 2014, Torre Archimede, Room 2BC/30.

Course requirements: Java and/or C++, programming experience, introduction to concurrency

Examination and grading: Two in-class quizzes, one two-week programming assignment (code to be run on cluster)

SSD: INF/01

Aim: To introduce students to programming with large data sets using application frameworks developed in (Resilient) X10.

Course contents:

This course will introduce students to the use and design of resilient parallel application frameworks in X10, such as Map Reduce, Bulk Synchronous Parallel (BSP) programming, Streaming, Global Load Balancing etc. These frameworks are characterized by offering a very simple programming model for the user – basically, sequential programming – while taking care of distribution, scale-out, synchronization and parallelism. Hitherto such frameworks have typically been developed from the ground up in languages such as Java or C++. We show how these frameworks can be built in a small amount of code (often just a few hundred lines of code) in X10, a programming language built explicitly for productivity and performance at scale. We will also discuss programming in Resilient X10, an extension of X10 that permits programs to continue executing even when nodes fail.

Topics covered: Introduction to Big Data and data analysis problems. Introduction to X10 and the APGAS (Asynchronous, Partitioned Global Address Space model), and basic multi-place programming idioms. Map Reduce programming – Hadoop and the X10 Map Reduce engine. Graph analytics – between-ness centrality, k-clique, using Bulk Synchronous Parallelism. State-space search – unbalanced tree search problem. Global Matrix Library – sparse, distributed matrix computations.