


DISTRIBUTED SYSTEMS AND CLOUD COMPUTING


Technical Teaching
Clouds
APPUSWAMY Raja

ABSTRACT

The goal of this course is to provide a comprehensive view on recent topics and trends in distributed systems and cloud computing. We will discuss the software techniques employed to construct and program reliable, highly-scalable systems. We will also cover architecture design of modern datacenters and virtualization techniques that constitute a central topic of the cloud computing paradigm. The course is complemented by a number of lab sessions to get hands-on experience with Hadoop and the design of scalable algorithms with MapReduce.

Teaching and Learning Methods: Lectures and Lab sessions (group of 2 students)

Course Policies: Attendance to Lab session is mandatory.

BIBLIOGRAPHY

- Learning Spark, by Holden Karau, Andy Konwinski, Patrick Wendell and Matei Zaharia, O'Reilly
- Data-intensive Text Processing with MapReduce, by Jimmy Lin and Chris Dyer
- Hadoop, The Definitive Guide, by Tom White
- Hadoop Operations, by Eric Sammer
- HBase, The Definitive Guide, by Lars George

REQUIREMENTS

Knowledge of data structures, algorithm design, distributed algorithms. Being fluent with Java and at least another programming language (python is recommended, scala is required for advanced topics) is highly desirable.

DESCRIPTION

- Introduction
- Scalable algorithm design
- Apache Hadoop internals
- Apache Spark internals
- Cluster and datacenter schedulers
- Relational algebra
- Apache Pig and Pig Latin
- Distributed Storage Systems
- Coordinating distributed systems with Apache Zookeeper
- Selected topics in Cloud Computing

Learning outcomes:

- A thorough understanding of the fundamentals of cloud computing.
- Gain proficiency in programming paradigms and run time systems developed for the cloud.
- Acquire the ability to analyze, design, and develop algorithms for solving several distributed systems problems.
- Explore a wide range of system design alternatives for various aspects of cloud-native application development and understand their tradeoffs.

Nb hours: 42.00, at least 5 Lab sessions (15 hours)

Grading Policy: Lab reports (50%), Final Exam (50%)