Course Big Data

Basic Information

This is a course, which contributes to MSc award in Mathematics

Duration	**108 hours (3 ECTS)**
Starting date	September 1st
Study credits	3 ECTS credits
Language of instruction	English level B1(European Framework of Reference of Communicative Skills)
Academic requirements	– BSc degree in Mathematics, Physics, Computer Science, Engineering or equivalent (a copy of your diplomas from previous university studies and transcripts of completes courses and grades) – Skype interview

Course Description

The discipline "Big Data" is a logical continuation of the cycle of disciplines in data analysis, which introduces students to existing methods and systems for storing, processing and analyzing big data.

The main goal of the course is to study the principles of working with big data, to get acquainted with the components of the Hadoop ecosystem and the computational paradigm MapReduce. During the educational process, students in practice implement the main Big Data solutions in cloud platforms and master the standard tools for working with Big Data.

Special Features of the Course

The course focuses on the practical application of the MapReduce distributed computing model. For the implementation of the algorithms, a freely distributed set of Hadoop utilities has been selected, which is used to implement search and contextual mechanisms of many highly loaded websites during mass-parallel data processing. Currently, Hadoop is considered one of the fundamental technologies when working with Big Data and is used in many industries: healthcare, telecommunications, trade, logistics, financial companies, as well as in public administration.

Course Aim

- Learn how to work with big data.
- Build skills for implementing Big Data solutions in cloud platforms.
- Provide students with knowledge of modern technical implementations of big data processing and analysis methods.

Course Objectives

- Master the standard tools for working with big data.
- Learn how to batch and stream data.
- Build hands-on skills with the components of the Hadoop ecosystem.
- Implementation of the solution of practical problems in the architecture of "Hadoop MapReduce".

Learning Outcomes of the Course

By the end of the course, students will be able to:

- use the Hadoop architecture for working with big data;
- implement models of streaming processing of big data;
- perform cloud computing on data of the order of several petabytes.

Course (module) Structure

Learning Activities	Hours
Lectures	18
Practice sessions / Seminars,	18
Self-study Assignments	72
Total study hours	**108**

Course Outline

Week	Lectures	Practice session / Assignments	Hours[1]
1-6	Batch Processing	– Distributed file systems – MapReduce Computation Model – Beyond MapReduce, Spark	36
7-12	Stream Data Processing	– Schemes for receiving and processing sweat data – Stream processing models and their characteristics – Streaming frameworks (open source Apache products) – Essay	36
13-18	Big Data Storage	– Types of data warehouses (Big Table-like, Dynamo-like, SQL over Big Data) – Architecture of data processing systems – The project “Cloud computing for Big Data analysis”	36

[1] Hours designed for Classroom sessions, Web-sessions, Home Assignments etc.

Assessment

Grade policy for project, essay and interview by discipline:

- A (excellent work) 91–100 points
- B (above average work) 81–90 points
- C (average work) 71–80 points
- D (below average work) 50–70 points
- F (failed work) < 50 points

Final standings are conducted in the form of an interview on sections of the discipline
The final grade includes grade for homework and interview in a ratio of 50:50.

Attendance Policy

Students are expected to attend classes regularly. However, occasional skipping classes is permissible if a student does all necessary in-class work at home.

Lecturer(s) and Tutors, Contact Information

Roman ESIN
Senior Lecturer at Institute of Space and Information Technologies, Siberian Federal University
(room 311) 26-1, Kirensky st, Krasnoyarsk, Russia
Tel: +7 391 291 2790,
resin@sfu-kras.ru, http://ikit.sfu-kras.ru/e/207

Core Reading

All the books are available on-line in Elsevier at SFU library.

1. **Rajkumar Buyya, Rodrigo Calheiros, Amir Vahid Dastjerdi**, Big Data, 1st Edition, Principles and Paradigms, Paperback ISBN: 9780128053942, eBook ISBN: 9780128093467, Imprint: Morgan Kaufmann, Published Date: 3rd June 2016, Page Count: 494
2. **David Loshin**, Big Data Analytics, 1st Edition, From Strategic Planning to Enterprise Integration with Tools, Tech-niques, NoSQL, and Graph, Paperback ISBN: 9780124173194, eBook ISBN: 9780124186644, Imprint: Morgan Kaufmann, Published Date: 30th August 2013, Page Count: 142