Introduction to Big Data

Bachelor's program 38.03.05 «Business Informatics»

Author: O.A.Tsukanova, lecturer (otsukanova@hse.ru)

Approved at the Meeting of the Department
of Innovation and Business in Information Technologies
“___” _____________ 2017.

Head of Department
________________ / S.V.Maltseva/

Approved by the Academic Council of Faculty of Business and Management
“___” _____________ 2017.

Chairman
________________ / A.V.Dmitriev/

Moscow, 2017

The document may not be used by other Departments of the University as well as other educational institutions without permission of the Authors
1. Applicability and Normative References

The present course syllabus sets minimal requirements for the theoretical background and practical skills obtained by students; provides the contents of the course and describes assessment of students' performance.

The program is designed for instructors teaching this course, teaching assistants and students of educational track 38.03.05 “Business Informatics”, bachelor's level.

The program is designed in compliance with:
- the educational standard of the Federal State Autonoumous Educational Institution of Higher Professional Education National Research University – Higher School of Economics, educational track 38.03.05 “Business Informatics”, level of training: Bachelor’s degree, as of 26.12.2014;
- the working curriculum of the University, educational track 38.03.05 “Business Informatics”, 4th year, bachelor’s program.

2. Course Objectives

The course is to introduce students to the core concepts of Big Data analysis and application.

3. Course Description

“Introduction to Big Data” is an elective “blended” course taught in the 4th year of the bachelor’s program. The course consists of the on-line part provided by www.coursera.org (course title – Introduction to Bid Data, https://www.coursera.org/learn/big-data-introduction) and the off-line part described below. The students are supposed to study the on-line part on their own using the materials available at www.coursera.org. The off-line part of the course helps students better understand the basics of Big Data by communicating with instructors. The coverage of the off-line part is not limited to the topics of the on-line part and makes special emphasis on the issues which may be hard for self-study. The duration of the course is one module. The course is worth 4 credits.

4. Learning Outcomes and Competences

By the end of the module, students should
- **Know** the fundamental concepts, principles and approaches to description of the Big Data Landscape.
- **Be able to** understand the main problems of the Big Data Analysis, get acquainted to the architectural components and programming models used for scalable data analysis.
- **Learn how to** use one of the most common frameworks, Hadoop.
- **Have the following competences:**
### 5. Role of the Course in the Curriculum

The course is a part of major (professional) block of disciplines. It is an elective course. The course is based on a number of preceding disciplines:

- Information Processes, Systems and Networks
- Data Management

To successfully study the current course, students should know major definitions and theorems of the above disciplines and be able to solve the typical problems. The concepts of the current course may be subsequently used in studying the following courses:

- Business and Innovation in ICT
6. Course Structure and Contents

<table>
<thead>
<tr>
<th>№</th>
<th>Topic</th>
<th>In-class hours (off-line phase)</th>
<th>Self-study</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lectures</td>
<td>Seminars</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>ON-LINE PHASE (Introduction to Big Data, <a href="https://www.coursera.org/learn/big-data-introduction">https://www.coursera.org/learn/big-data-introduction</a>)</td>
<td></td>
<td></td>
<td>120</td>
</tr>
<tr>
<td>2.</td>
<td>OFF-LINE PHASE Selected topics related to Big Data Analytics</td>
<td>10</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>10</td>
<td>10</td>
<td>132</td>
</tr>
</tbody>
</table>

Course Outline (Refers to the material of both the on-line and off-line parts of the course)

1. ON-LINE PHASE
   - Introduction.
   - Characteristics of Big Data and Dimensions of Scalability.
   - Data Science: Getting Value out of Big Data.
   - Foundation for Big Data Systems and Programming.
   - Systems: Getting Started with Hadoop

2. OFF-LINE PHASE
   Selected topics related to Big Data in:
   - Social Network Analysis
   - E-Learning
   - Healthcare

7. Assessment of Student's Performance

<table>
<thead>
<tr>
<th>Type of assessment</th>
<th>Means of assessment</th>
<th>4th year</th>
<th>Department</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Current (last weeks of module 3)</td>
<td>Tests “Introduction to Big Data”</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>Exam</td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

7.1. Criteria of assessment

To qualify the current assessment, the students should be able to solve the tests and problems suggested by the on-line part of the course as well as those discussed in class.
7.2. Topics tested at current assessment


7.3. Questions for assessing student’s performance

Basic concepts, principles and algorithms studied in the course “Introduction to Big Data”

7.4. Sample exam questions

The final exam is paper based. The duration of the exam is 90 minutes. The final exam covers the material of both the on-line and off-line parts of the course.

8. Grading

The final grade $O_{fin}$ is comprised of the grade $O_{accm}$ accumulated over the module and the grade $O_{exam}$ obtained on the final in-class exam:

$$O_{fin} = 0.7 \cdot O_{accm} + 0.3 \cdot O_{exam}$$

The accumulated grade $O_{accm}$ includes both the results of the on-line tests (at www.coursera.org) and the in-class activities in the off-line phase.