

*APPROVED BY*

*Director of Institute of Cybernetics*  
 / D.M. Sonkin

***Introduction to Big Data***

***Field of Study:*** 09.04.04 Software Engineering

***Programme name:*** Big Data Solutions

***Level of Study:*** Master Degree Programme

***Year of admission:*** 2019

***Semester, year:*** 1,1

***ECTS:*** 3

***Total Hours:*** 108

***Contact Hours:*** 48

- ***Lectures:*** 24
- ***Labs:*** 24


***Assessment:*** credit test

***Department:*** Software Engineering

***Head of Department***

  
\_\_\_\_\_  
V.S. Sherstnev

***Instructor(s)***

  
\_\_\_\_\_  
E.I. Gubin



## *Introduction to Big Data*

### *Course Overview*

<b>Course Objectives</b>	Course is aimed to formation of student's skills and abilities for Big Data area.
<b>Learning Outcomes</b>	<p>As a result of mastering the discipline, the student must achieve the following results:</p> <ul style="list-style-type: none"> <li>- Ability to collect and clean up the raw data from multiple data sources.</li> <li>- Ability to create NoSQL store of data.</li> <li>- Ability to interpret and compute huge amount of data.</li> <li>- Understanding the basics of the following technologies: Map Reduce, CRISP-DM, Data mining, Hadoop.</li> </ul>
<b>Course Outline</b>	<ol style="list-style-type: none"> <li>1. <b>Big Data: Why and Where</b></li> <li>2. <b>Characteristics of Big Data and Dimensions of Scalability</b></li> <li>3. <b>Data Science: Getting Value out of Big Data</b></li> <li>4. <b>Foundations for Big Data Systems and Programming</b></li> <li>5. <b>Systems: Getting Started with Hadoop</b></li> </ol>
<b>Prerequisites (if available)</b>	<ol style="list-style-type: none"> <li>1. Theory of Probability and Mathematical Statistics.</li> <li>2. Informatics and Programming Basics.</li> </ol>
<b>Course Structure</b>	<p><b>Big Data: Why and Where</b> Data it's been around (even digitally) for a while. What makes data "big" and where does this big data come from?</p> <p><b>Characteristics of Big Data and Dimensions of Scalability</b> Different values of BigData</p> <p><b>Data Science: Getting Value out of Big Data</b> 5 step process for approaching data science problems.</p> <p><b>Foundations for Big Data Systems and Programming</b> Grounding in some of the key concepts in programming frameworks and systems for BigData.</p> <p><b>Systems: Getting Started with Hadoop</b> Introduction to Hadoop and MapReduce. Simple MapReduce task in the Cloudera VM.</p>
<b>Facilities and Equipment</b>	<ul style="list-style-type: none"> <li>- Lecture room (computer, projector) – 634034 Tomsk Region, Tomsk, Sovetskaya Street, 84/3, room 313.</li> <li>- Computer class (12 computers) – 634034 Tomsk Region, Tomsk, Sovetskaya Street, 84/3, room. 212.</li> </ul>
<b>Grading Policy</b>	<p>In accordance with TPU rating system we use:</p> <ul style="list-style-type: none"> <li>- Current assessment which is performed on a regular basis during the semester by scoring the quality of mastering of theoretical material and the results of practical activities (performance tests, perform tasks, problem solving). Max score for</li> </ul>



	<p>current assessment is 60 points, min – 40 points.</p> <ul style="list-style-type: none"> <li>- Course final assessment (exam/ credit test) is performed at the end of the semester. Max score for course final assessment is 40 points, min – 22 points.</li> </ul> <p>The final rating is determined by summing the points of the current assessment during the semester and exam (credit test) scores at the end of the semester. Maximum overall rating corresponds to 100 points, min pass score is 65.</p>
<b>Course Policy</b>	<p>Class attendance will be taken into consideration when evaluating students' participation in the course.</p> <p>Students are expected to actively engage in class discussions about the assigned readings.</p>
<b>Teaching Aids and Resources</b>	<p>Compulsory Readings:</p> <ol style="list-style-type: none"> <li>1. Big Data Analytics: A Literature Review Paper [Electronic resource]. - Access mode: <a href="https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper">https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper</a>, free.</li> <li>2. Big Data Tutorial [Electronic resource]. - Access mode: [Electronic resource]. - Access mode: <a href="https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper">https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper</a>, free., free.</li> <li>3. Overview of Big Data Analytics [Electronic resource]. - Access mode: <a href="https://www.ee.columbia.edu/~cylin/course/bigdata/EECS6893-BigDataAnalytics-Lecture1.pdf">https://www.ee.columbia.edu/~cylin/course/bigdata/EECS6893-BigDataAnalytics-Lecture1.pdf</a>, free.</li> </ol> <p>Additional Readings:</p> <ol style="list-style-type: none"> <li>4. Introduction to BigData [Electronic resource]. - Access mode: <a href="https://ru.coursera.org/learn/big-data-introduction">https://ru.coursera.org/learn/big-data-introduction</a>, free.</li> <li>5. Big Data Fundamentals [Electronic resource]. - Access mode: <a href="https://www.edx.org/course/big-data-fundamentals-adelaidex-bigdatax">https://www.edx.org/course/big-data-fundamentals-adelaidex-bigdatax</a>, free</li> <li>6. Big Data: from Data to Decisions [Electronic resource]. - Access mode: <a href="https://www.futurelearn.com/courses/big-data-decisions?utm_campaign=Courses+feed&amp;utm_medium=courses-feed&amp;utm_source=courses-feed">https://www.futurelearn.com/courses/big-data-decisions?utm_campaign=Courses+feed&amp;utm_medium=courses-feed&amp;utm_source=courses-feed</a>, free</li> </ol>
<b>Instructor (-s)</b>	<p>Gubin Evgeni gubine@tpu.ru 8 906 958 7250</p>