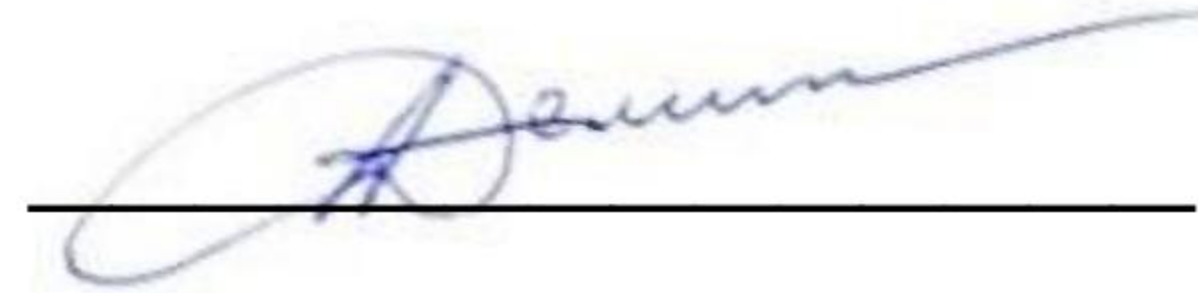


APPROVED BY

Director of Institute of Cybernetics

 / D. M. Sonkin

Large Scale Data Bases

Field of Study: Big Data

Programme name: 09.04.04 Software Engineering

Level of Study: Master Degree Programme

Year of admission: 2020

Semester, year: 2

ECTS: 3

Total Hours: 108

Contact Hours: 32

- **Lectures:** 16
- **Labs:** 16
- **Practical experience:** 0

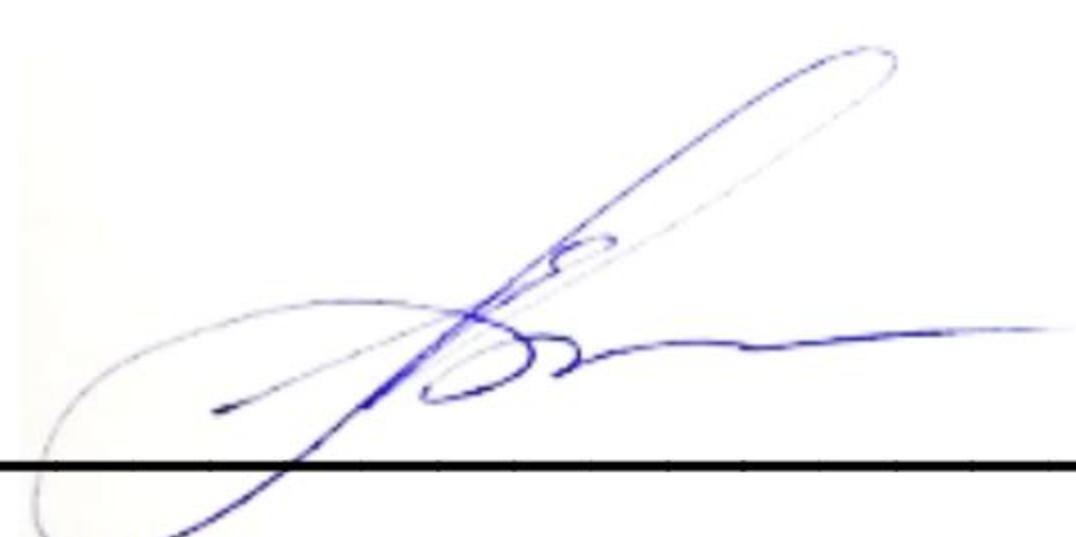
Assessment: exam

Department: Department of Software Engineering

Head of Department

 / V.S. Sherstnev

Instructor(s)

 / Mokina E.E.

Large Scale Data Bases

Course Overview

Course Objectives	The course begins with an introduction to NoSQL, and then delves into the specifics of document, wide-column, and graph databases. Learn key details for performing data preparation, exploration, and extraction for each type of NoSQL database. Review case studies that show how to use various NoSQL databases with popular data science tools, including the document database MongoDB, the wide-column database Cassandra, and the graph database Neo4j.
Learning Outcomes	The student should know and understand: • Define, compare and use the four types of NoSQL Databases (Document-oriented, Key/Value Pairs, Column-oriented and Graph). • Demonstrate an understanding of the detailed architecture, define objects, load data, query data and performance tune Column-oriented NoSQL databases. • Explain the detailed architecture, define objects, load data, query data and performance tune Document-oriented NoSQL databases. • Demonstrate an understanding of the detailed architecture, define objects, load data, query data and performance tune Key-Value Pair NoSQL databases. • Explain the detailed architecture, define objects, load data, query data and performance tune Graph NoSQL databases. • Evaluate NoSQL database development tools and programming languages. • Perform hands-on NoSQL database lab assignments that will allow students to use the four NoSQL database types via products such as Cassandra, Hadoop Hbase, MongoDB, Neo4J.
Course Outline	The course consists of 16 lectures and 16 labs covering the following main topics: Mongo DB, Hadoop unstructured data concepts (key-value), Map Reduce technology, related tools that provide SQL-like access to unstructured data: Pig and Hive, NoSQL storage solutions like HBase, Cassandra, and Oracle NoSQL and analytics for big data.
Prerequisites (if available)	Programming languages (Python), Data Analysis Methods
Course Structure	The course consists of 3 parts. The first part, “Introduction to noSQL database”, The second part, “Document-oriented database”, The third part, “Key/Value Pairs database” And the last part “Column-oriented and Graph database”.
Facilities and Equipment	3 servers with Big Data processing software (HP DL385p Gen8, 2 processors 6320 (2.8GHz-16MB) 8-Core Processor Option Kit, 6 Memory modules 8GB 2Rx4 PC3L-10600R-9 , RAID controller P420i (512MB) FBWC RAID 0,1,1+0,5,5+0, 11 HDD 500GB SC 6G 7.2K LFF SATA HotPlug Midline Drive 1y war, Flash drive 120GB 6G SATA VE 3.5in SCC EV G1 SSD) Hadoop cluster (Pig, Hive, Spark)
Grading Policy	In accordance with TPU rating system we use: • 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 (labs). Max score for current assessment is 60 points, min – 30 points. • Course final assessment (exam) is performed at the end of the semester. Max score for course final assessment is 40 points, min – 25 points. • Programming project 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 55 points.
Course Policy	Class attendance will be taken into consideration when evaluating students' participation in the course.
Teaching Aids and Resources	Compulsory Readings: • Pramod J. Sadalage Martin Fowler. NoSQL Distilled A Brief Guide to the Emerging World of Polyglot Persistence, Williams, ISBN: 978-5-8459-1829-1, 2013. • Joe Celko. Joe Celko's Complete Guide to NoSQL: What Every SQL Professional Needs to Know about Non-Relational Databases CRC Press, 2013. Additional Readings: • Brad DayLey NoSQL with MongoDB in 24 Hours, Sams Teach Yourself, ISBN-13: 978-0672337130, 2015 • Jeff Carpenter and Eben Hewitt .Cassandra: The Definitive Guide: Distributed Data at Web Scale, ISBN-13: 978-1491933664, 2016 • Graph Databases: New Opportunities for Connected Data • ISBN-13: 978-1491933664,2015
Instructor (-s)	Elena E. Mokina, +79059928662, alisandra@tpu.ru