Ministry of Science and Higher Education of the Russian Federation
NATIONAL RESEARCH
TOMSK STATE UNIVERSITY (NR TSU)

Institute of Applied Mathematics and Computer Science

APPROVE:
Director
_____ A. V. Zamyatin

Work program of the discipline

Post-relational Data Models and Industrial DBMS - II

in the major of training

01.04.02 Applied mathematics and informatics

Orientation (profile) of training:

Big Data and Data Science

Form of study
full-time

Qualification
Master

Year of admission
2023

Code of discipline in the curriculum: B1.P.V.07

AGREED:
Head of EP
_____ A.V. Zamyatin

Chairman of the EMC
_____ S.P. Sushchenko

Tomsk – 2023

1. Purpose and planned results of mastering the discipline

The purpose of mastering the discipline is the formation of the following competencies:

GPC-4 - the ability to combine and adapt existing information and communication technologies to solve problems in the field of professional activity, taking into account the requirements of information security;

PC-5 - the ability to choose methods, draw up terms of reference and develop algorithms for solving problems of industrial data analysis.

The results of mastering the discipline are the following indicators of the achievement of competencies:

IOPC-4.2 Takes into account the basic information security requirements.

IPK-5.1 Uses modern information processing technologies, computer technology in solving problems of industrial data analysis.

IPK-5.2 Knows how to collect industrial data, knows the specifics of such data.

IPK-5.3. Draws up the terms of reference for the task of the professional area.

2. Tasks of mastering the discipline

- Know modern tools for long-term data storage;
- Choose tools for long-term data storage for their practical application;
- Have the skills to use theoretical and practical knowledge in the field of building post-relational databases;
- Design models of relational and post-relational databases.

3. The place of discipline in the structure of the educational program

Discipline refers to a part of the educational program formed by participants in educational relations.

4. Semester of mastering and form of intermediate certification in the discipline

Third semester, credit.

5. Entrance requirements for mastering the discipline

For the successful mastering of the discipline, training outcomes are required in the following disciplines: "Mathematical Methods and Models for Computer Science", "Algorithms and Data Structures", "Post-relational Data Models and Industrial DBMS - I", knowledge of Python .

6. Implementation language

English.

7. Scope of discipline

The total labor intensity of the discipline is 3 credits, 108 hours, of which:

- lectures: 10 hours
- laboratory: 20 hours

including practical training: 0 h.

The volume of independent work of the student is determined by the curriculum.

8. The content of the discipline, structured by topics

Topic 1. The history of the development and formation of the concept of databases. Relational databases.

Historical development of data storage systems. Basic concepts of databases, data structures and database management systems. relational model. Basic concepts and terms of the relational model. Relational DBMS. The concept of normal form. SQL is the standard query language for relational DBMSs. The concept of data integrity. Transaction processing. Disadvantages of the relational data model.

Topic 2. Object-oriented and postrelational DBMS. Post-relational databases (NoSQL). Temporal databases.

Object-oriented DBMS. Post-relational databases (NoSQL). The CAP theorem. Basic types of NoSQL databases. Advantages and disadvantages of NoSQL databases. Application area. Basic principles underlying temporal data models. The concept of time in temporal data models. Models used in temporal databases (TRM, HDM). Temporality in relational DBMS.

Topic 3. Document-oriented databases.

Document-oriented data model. Advantages and disadvantages of the model, areas of application. Methods for working with data. Data storage structures. MongoDB database. Elasticsearch search engine.

Topic 4. Graph databases

Graph theory. Graph data model. Advantages and disadvantages of the model, areas of application. Methods for working with data. Data storage structures. Neo4J database. Requests in the Cypher language.

Topic 5. Key-value databases.

Key-value data model. Advantages and disadvantages of the model, areas of application. Methods for working with data. Data storage structures. Database Riak, Redis.

Topic 6. Column databases.

Column databases. data model. Advantages and disadvantages of the model, areas of application. Methods for working with data. Data storage structures. Cassandra database.

9. Ongoing evaluation

The ongoing evaluation is carried out by monitoring attendance, conducting tests, tests on lecture material, performing laboratory work, and is recorded in the form of a checkpoint at least once a semester.

10. The procedure for conducting and criteria for evaluating the intermediate certification

The test in the first semester is carried out in writing in the form of a test, which contains at least 30 theoretical and practical questions. The duration of the test is 1.5 hours. A prerequisite for admission to the mastery test is the successful completion of all laboratory and practical work on the course.

Sample test question:

1. What type of noSQL database is MongoDB?

Answers:

a) Key-value database

b) Graph databases

c) Document-oriented databases

d) Column family repositories.

The results of the mastery test are determined in points and make up 30% of the rating.

11. Educational and methodological support

a) Evaluation materials of the current control are sets of tests on the topics of the theoretical block.

b) Guidelines for laboratory work are presented in the form of a description of laboratory work in pdf files. List of laboratory works:

Lab.1. Logical and physical design in Oracle DataModeler

Lab.2. Designing a temporal data model.

Lab.3. Working with document-oriented database MongoDB. Creation of requests to the data.

Lab.4. Creation of a graph data model. Creation of requests to the data.

Lab.5. Creating and working with a Redis (or Riak) database. Creation of requests to the data.

Lab.6. Creating and working with the Cassandra database. Creation of requests to the data.

12. List of educational literature and Internet resources

a) basic literature:

1.	MongoDB Basics	Hows, David.	2014
2.	Next Generation Databases NoSQL, NewSQL, and Big Data /	Harrison, Guy.	2015
3.	Practical Neo4j	Jordan, Gregory.	2014
4.	Quality-aware Scheduling for Key-value Data Stores	Xu, Chen.	2015

b) additional literature:

1.	Beginning Apache Cassandra Development	Mishra, Vivek.	2014
2.	Practical MongoDB Architecting, Developing, and Administering MongoDB /	Edward, Shakuntala Gupta.	2015
3.	The Definitive Guide to MongoDB A complete guide to dealing with Big Data using MongoDB /	Hows, David.	2015

c) Internet resources:

Documentation on working with databases, presented in the public domain on official websites.

1. https://www.mongodb.com/
2. http://cassandra.apache.org/
3. https://neo4j.com/
4. https://www.elastic.co/
5. http://www.oracle.com/

13. List of information technologies

a) licensed and freely distributed software::

– Microsoft Office Standard 2013 Russian: software package. Includes applications: MS Office Word, MS Office PowerPoint,

– publicly available cloud technologies (Google Docs, Yandex disk, etc.).

– Open databases (Mongodb, Neo4j, Redis, Cassandra)

– Oracle DataModeler relational database modeling tool

b) information reference systems:

– Electronic catalog of the TSU Scientific Library – http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system

– TSU electronic library (repository) – http://vital.lib.tsu.ru/vital/access/manager/Index

14. Logistics

Halls for lectures. Classrooms for seminars, individual and group work, ongoing evaluation and intermediate certification. Classrooms for independent work, equipped with computer technology and access to the Internet, to the electronic information and educational environment and to information reference systems.

All basic and additional literature necessary for independent work and preparation for the exam is available in the TSU scientific library. To perform laboratory work, databases deployed on the university's server are required.

15. Authors information

Mokina Elena Evgenievna, Senior teacher of the Department of Theoretical Foundations of Informatics, Institute of Applied Mathematics and Computer Science TSU.