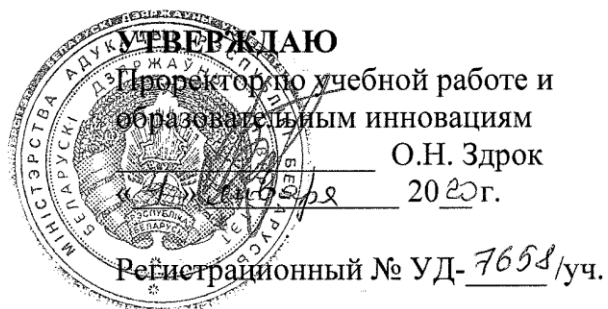


Белорусский государственный университет



**МАСШТАБИРУЕМЫЕ СТРУКТУРЫ И ХРАНИЛИЩА ДАННЫХ
ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ**

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности:**

**1-98 01 01 Компьютерная безопасность (по направлениям)
направление специальности**

**1-98 01 01-02 Компьютерная безопасность (радиофизические методы и
программно-технические средства)**

2020 г.

Учебная программа составлена на основе ОСВО 1-98 01 01-2013 и учебного плана УВО № Р98-139/уч., Р98и-140/уч. от 30.05.2013 г.

СОСТАВИТЕЛИ:

Василий Сергеевич САДОВ, профессор кафедры интеллектуальных систем, факультет радиофизики и компьютерных технологий, Белорусский государственный университет, кандидат технических наук, доцент;

Александр Васильевич КУРОЧКИН, старший преподаватель кафедры интеллектуальных систем, факультет радиофизики и компьютерных технологий, Белорусский государственный университет;

Екатерина Александровна ГОЛОВАТАЯ, старший преподаватель кафедры интеллектуальных систем, факультет радиофизики и компьютерных технологий, Белорусский государственный университет;

РЕЦЕНЗЕНТ:

Владимир Васильевич ГОЛЕНКОВ, профессор кафедры интеллектуальных информационных технологий УО «Белорусский государственный университет информатики и радиоэлектроники», доктор технических наук, профессор;

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой интеллектуальных систем факультета радиофизики и компьютерных технологий Белорусского государственного университета

(протокол № 5 от 20.11.2019);

Научно-методическим Советом БГУ

(протокол № 3 от 03.01.2020)

Заведующий кафедрой
интеллектуальных систем _____

Е.И. Козлова

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины – систематизация знаний в области методов оперативного и долгосрочного хранения информации в интеллектуальных информационных системах различного назначения.

Задачи учебной дисциплины:

1. Рассмотреть основополагающие принципы проектирования систем оперативного и долгосрочного хранения информации в информационных системах, а также требования к таким системам;
2. Сформировать представление об особенностях хранения данных в различных сценариях использования информационных систем;
3. Ознакомить с актуальными решениями и проблемами в области оперативного и долгосрочного хранения данных;
4. Предоставить навыки анализа пригодности существующих решений для различных задач автоматизации и масштабирования.

Учебная дисциплина относится к циклу дисциплин специализации компонента учреждения высшего образования.

Для успешного усвоения данной учебной дисциплины необходимы знания по дисциплинам «Программирование», «Дискретная математика и математическая логика», «Численные методы» в объеме программы высшей школы.

Требования к компетенциям

Освоение учебной дисциплины «Масштабируемые структуры и хранилища данных интеллектуальных систем» должно обеспечить формирование следующих *профессиональных* компетенций:

- ПК-3. Разрабатывать модели явлений, процессов или систем при организации защиты информации;
- ПК-4. Выбирать необходимые методы исследования, модифицировать существующие, разрабатывать новые методы и применять их для решения поставленных задач при организации защиты информации;
- ПК-15. Организовывать процесс создания, оценки и эксплуатации средств и систем защиты информации, поддерживать и повышать их безопасность; осуществлять контроль за их использованием.

В результате освоения учебной дисциплины студент должен:

знать:

- теоретические принципы, лежащие в основе работы систем хранения информации;
- основные сценарии использования различных средств хранения информации;

уметь:

- проектировать системы долговременного и оперативного хранения информации для различных частей информационных систем;

- выявлять и устранять основные угрозы информационной безопасности при хранении данных;

владеть:

- навыками анализа нагрузки, масштабирования и расширения систем хранилищ данных;
- инструментами развертывания, мониторинга, резервного копирования, восстановления, отказоустойчивой репликации оперативных и долговременных хранилищ информации.

Структура учебной дисциплины

Дисциплина изучается в 8 семестре. Всего на изучение учебной дисциплины «Масштабируемые структуры и хранилища данных интеллектуальных систем» отведено 130 часов, в том числе 62 аудиторных часа, из них: лекции – 34 часа, лабораторные занятия – 24 часа, управляемая самостоятельная работа – 4 часа.

Трудоемкость учебной дисциплины составляет 3,5 зачетные единицы.
Форма текущей аттестации в 8-м семестре – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Основы систем хранения информации

Тема 1.1. Задача хранения информации. Оперативное и долговременное хранение.

Тема 1.2. Представление информации в памяти. Основные алгоритмы и структуры данных для быстрого доступа к информации.

Тема 1.3. Понятие схемы данных. Отношения и связи. Виды связей.

Тема 1.4. Реляционные и нереляционные хранилища информации. **Тема 1.5.** Основные свойства хранилищ информации. CAP-теорема.

Тема 1.6. Поддержка транзакций. ACID.

Тема 1.7. Администрирование и поддержка систем хранения информации. Обновления и миграции.

Тема 1.8. Мониторинг и оперативное оповещение при работе с системами хранения информации. Резервное копирование и обслуживание хранилищ.

Раздел 2. Нереляционные и специализированные хранилища информации

Тема 2.1. Столбцовые и широко-столбцовые хранилища на примере Apache Cassandra.

Тема 2.2. Хранилища «ключ-значение» на примере Redis.

Тема 2.3. Документно-ориентированные базы данных на примере MongoDB.

Тема 2.4. Событийные хранилища и очереди сообщений на примере Apache Kafka и RabbitMQ. **Тема 2.5.** Графовые базы данных на примере OrientDB.

Раздел 3. Масштабирование и распределение нагрузки в хранилищах информации

Тема 3.1. Стратегии горизонтального и вертикального масштабирования хранилищ данных. Масштабирование индексов. Шардинг и репликация.

Тема 3.2. Отказоустойчивость хранилищ информации. Отказоустойчивые кластеры и алгоритмы выбора ведущего.

Тема 3.3. Высоконагруженные системы хранения информации. Стратегии оптимизации доступа к высоконагруженным хранилищам.

Тема 3.4. Хранилища информации в облачном окружении. Масштабирование при использовании PaaS и IaaS провайдеров.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дневная форма получения образования с применением дистанционных образовательных технологий

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСП	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8	9
1.	Основы систем хранения информации	16			2			
1.1.	Задача хранения информации. Оперативное и долговременное хранение.	2						выборочный опрос
1.2.	Представление информации в памяти. Основные алгоритмы и структуры данных для быстрого доступа к информации.	2			2			отчёт по лабораторной работе
1.3.	Понятие схемы данных. Отношения и связи. Виды связей.	2						выборочный опрос
1.4.	Реляционные и нереляционные хранилища информации.	2						выборочный опрос
1.5.	Основные свойства хранилищ информации. CAP-теорема.	2						выборочный опрос
1.6.	Поддержка транзакций. ACID.	2						тест
1.7.	Администрирование и поддержка систем хранения информации. Обновления и миграции.	2						выборочный опрос
1.8.	Мониторинг и оперативное оповещение при работе с системами хранения	2						выборочный опрос

	информации. Резервное копирование и обслуживание хранилищ.							
2.	Нереляционные и специализированные хранилища информации	10			22		4	
2.1.	Столбцовые и широко-столбцовые хранилища на примере Apache Cassandra.	2			4			отчёт по лабораторной работе
2.2.	Хранилища «ключ-значение» на примере Redis.	2			6		2	отчёт по лабораторной работе
2.3.	Документно-ориентированные базы данных на примере MongoDB.	2			4		2	отчёт по лабораторной работе
2.4.	Событийные хранилища и очереди сообщений на примере Apache Kafka и RabbitMQ.	2			4			отчёт по лабораторной работе
2.5.	Графовые базы данных на примере OrientDB.	2			4			отчёт по лабораторной работе
3.	Масштабирование и распределение нагрузки в хранилищах информации	8						
3.1.	Стратегии горизонтального и вертикального масштабирования хранилищ данных. Масштабирование индексов. Шардинг и репликация.	2						тест
3.2.	Отказоустойчивость хранилищ информации. Отказоустойчивые кластеры и алгоритмы выбора ведущего.	2						выборочный опрос
3.3.	Высоконагруженные системы хранения информации. Стратегии оптимизации доступа к высоконагруженным хранилищам.	2						выборочный опрос
3.4.	Хранилища информации в облачном окружении. Масштабирование при	2						выборочный опрос на лекции

	использовании PaaS и IaaS провайдеров.							
	Всего	34			24		4	

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Encyclopedia of Database Systems / L. Liu [et. al.] ; ed. L. Liu, M. T. Ozs. – 2nd ed. – Georgia Institute of Technology College of Computing, Atlanta, GA, USA, 2018. – 4866 p.
2. Harrison, G. Next Generation Databases: NoSQLand Big Data / G. Harrison // Apress, 2015. – 260 p.
3. Wiese, L. Advanced Data Management: For SQL, NoSQL, Cloud and Distributed Databases / L. Wiese // 1st ed. – De Gruyter, 2015. – 374 p.
4. Ozs. M. T. Principles of Distributed Database Systems // M. T. Ozs., V. Patrick // 4th ed. – Springer International Publishing, 2020. – 674 p.
5. Kroenke, D. M. Database Concepts / D. M. Kroenke, D. Auer, S. L. Vandenberg, R. C. Yoder // 8th ed. – Pearson, 2019. – 576 p.
6. Kleppmann, M. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems / M. Kleppmann // O'Reilly Media, 2017. – 616 p.

Перечень дополнительной литературы

1. Abadi, D. The Design and Implementation of Modern Column-Oriented Database Systems // D. Abadi, P. Boncz, S. Harizopoulos // Now Publishers Inc, 2013. – 102 p.
2. Celko, J. Joe Celko's Complete Guide to NoSQL: What Every SQL Professional Needs to Know about Non-Relational Databases / J. Celko // Morgan Kaufmann, 2013. – 244 p.

Перечень рекомендуемых средств диагностики и методика формирования итоговой оценки

Оценка за ответы на лекциях (опрос) включает в себя полноту ответа, наличие аргументов, примеров из практики и т.д.

Оценка за выполнение лабораторных работ включает в себя полноту выполнения перечисленных заданий по лабораторной работе, а также владение теоретическим материалом по теме работы.

При оценивании УСР обращается внимание на полноту раскрытия темы, корректность выполнения поставленного задания и выполнение условий по установленной форме контроля.

Формой текущей аттестации по дисциплине «Масштабируемые структуры и хранилища данных интеллектуальных систем» учебным планом предусмотрен экзамен.

При формировании итоговой оценки используется рейтинговая оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине.

Формирование оценки за текущую успеваемость:

- защита отчетов по лабораторным работам – 50 %;
- УСР – 25%;
- выполнение тестов – 25%.

Рейтинговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости и экзаменационной оценки с учетом их весовых коэффициентов. Вес оценки по текущей успеваемости составляет 40 %, экзаменационная оценка – 60 %.

Примерный перечень заданий для управляемой самостоятельной работы студентов

Тема 2.2. Хранилища «ключ-значение» на примере Redis. (2ч.)

Проектирование программных кэшей данных на основе Redis.

Спроектировать программный кэш для получения сущностей через REST API по уникальному ключу. Реализовать методы инвалидации по ключу и по всем записям.

(Форма контроля – прохождение автоматизированного теста на согласованность данных при комбинации операций чтения и модификации).

Тема 2.3. Документно-ориентированные базы данных на примере MongoDB. (2ч.)

Встроенные средства обеспечения отказоустойчивости сервера MongoDB.

Создать сценарий развертывания отказоустойчивого кластера из 3 экземпляров сервера MongoDB с репликацией. Убедиться в выполнении свойства согласованности в конечном счёте.

(Форма контроля – демонстрация работоспособности тестового приложения с записью и чтением через фиксированные промежутки времени при ручном отключении ведущего и ведомого экземпляра сервера).

Примерная тематика лабораторных занятий

Лабораторная работа № 1. Алгоритмы и структуры данных для поддержки индексирования и поиска информации.

Лабораторная работа № 2. Использование широко-столбцовых хранилищ для работы с большими объемами данных. Оптимизация производительности в широко-столбцовых хранилищах.

Лабораторная работа № 3. Сценарии использования хранилищ «ключ-значение» с реализацией в оперативной памяти. Основные типы данных и команды в Redis.

Лабораторная работа № 4. Использование MongoDB. Прямой доступ к данным. Работа со схемами при помощи паттерна Active Record на примере библиотеки Mongoose. Работа с миграциями.

Лабораторная работа № 5. Концепция очереди сообщений. Работа с сервером очередей RabbitMQ. Очереди сообщений в условиях наличия и отсутствия потребителей.

Лабораторная работа № 6. Работа с графовыми базами данных. Понятие вершины и связи. Глубокий обход и извлечение фактов в графовых базах данных.

Описание инновационных подходов и методов к преподаванию учебной дисциплины (эвристический, проективный, практико-ориентированный)

При организации образовательного процесса используется практико-ориентированный подход, который предполагает:

- освоение содержания образования посредством решения практических задач;
- приобретение навыков эффективного выполнения разных видов профессиональной деятельности;
- ориентацию на генерирование идей, реализацию групповых студенческих проектов, развитие предпринимательской культуры;
- использование процедур, способов оценивания, фиксирующих сформированность профессиональных компетенций.

При проведении выборочных опросов на лекциях используется метод учебной дискуссии, который предполагает участие студентов в целенаправленном обмене мнениями, идеями для предъявления и/или согласования существующих позиций по определенной проблеме. Использование метода обеспечивает появление нового уровня понимания

изучаемой темы, применение знаний (теорий, концепций) при решении проблем, определение способов их решения.

При проведении лабораторных занятий рекомендуется использовать метод группового обучения, который представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую функционирование разных типов малых групп, работающих как над общими, так и специфическими учебными заданиями.

Методические рекомендации по организации самостоятельной работы обучающихся

При изучении учебной дисциплины рекомендуется использовать следующие формы самостоятельной работы:

- поиск (подбор) и обзор литературы и электронных источников по индивидуально заданной проблеме курса;
- выполнение домашнего задания;
- работы, предусматривающие решение задач и выполнение упражнений, выдаваемых на практических занятиях;
- изучение материала, вынесенного на самостоятельную проработку;
- анализ статистических и фактических материалов по заданной теме, проведение расчетов, составление схем и моделей на основе статистических материалов;
- подготовка к участию в конференциях и конкурсах.

При составлении заданий УСР по учебной дисциплине необходимо предусмотреть возрастание их сложности: от заданий, формирующих достаточные знания по изученному учебному материалу на уровне узнавания, к заданиям, формирующим компетенции на уровне воспроизведения, и далее к заданиям, формирующим компетенции на уровне применения полученных знаний.

Таким образом, задания УСР по учебной дисциплине рекомендуется делить на три модуля:

- задания, формирующие достаточные знания по изученному учебному материалу на уровне узнавания;
- задания, формирующие компетенции на уровне воспроизведения;
- задания, формирующие компетенции на уровне применения полученных знаний.

Примерный перечень вопросов к экзамену

1. Представление информации в памяти. Оперативное и долговременное хранение.
2. Представление информации в памяти. Основные алгоритмы и структуры данных для быстрого доступа информации.
3. Понятие схемы данных. Отношения и связи.
4. Понятие схемы данных. Нормализация.
5. Реляционные и нереляционные хранилища.
6. Согласованность данных в хранилищах информации. CAP-теорема.
7. Доступность данных в хранилищах информации. CAP-теорема.
8. Устойчивость данных к разделению в хранилищах информации. CAP-теорема.
9. Понятие транзакции. Уровни изоляции транзакций.
10. Понятие транзакции. ACID.
11. Понятие миграции данных. Стратегии применения миграций.
12. Обслуживание систем хранения информации
13. Мониторинг и оперативное оповещение при работе с системами хранения информации.
14. Резервное копирование в системах хранения информации.
15. Столбцовые и широко-столбцовые хранилища.
16. Хранилища «ключ-значение».
17. Документно-ориентированные базы данных.
18. Событийные хранилища и очереди сообщений.
19. Графовые базы данных.
20. Масштабирование хранилищ данных.
21. Шардинг и репликация.
22. Отказоустойчивые кластеры и алгоритмы выбора ведущего.
23. Высоконагруженные системы хранения информации.
24. Хранилища информации в облачном окружении.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Согласование с другими дисциплинами не требуется			

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 201_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ
Декан факультета
