

Белорусский государственный университет

УТВЕРЖДАЮ

Проректор по учебной работе
и образовательным инновациям

О.И. Чуприс



Регистрационный № УД-5636/уч.

СИСТЕМЫ ХРАНЕНИЯ ДАННЫХ

Учебная программа учреждения высшего образования
по учебной дисциплине для специальности второй ступени высшего
образования (магистратуры) с углубленной подготовкой специалиста:

1-31 81 09 Алгоритмы и системы обработки больших объемов
информации

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 81 09-2014 и учебного плана G31-254/уч. от 26.05.2017 г.

Составители:

А.А. Толстик – старший преподаватель кафедры вычислительной математики факультета прикладной математики и информатики Белорусского государственного университета.

Рекомендована к утверждению:

Кафедрой дискретной математики и алгоритмики Белорусского государственного университета (протокол № 14 от 26 апреля 2018 г.);

Научно-методическим советом Белорусского государственного университета (протокол № 7 от 13 июля 2018 г.).

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Учебная дисциплина «Системы хранения данных» знакомит студентов магистратуры с основными направлениями в разработке и использовании систем хранения данных. В дисциплину включены связанные с практическим использованием систем хранения данных разделы. Особое внимание уделяется распределенным системам, которые чаще всего строятся на базе облачной инфраструктуры. Эффективность полученной системы сильно зависит от накопителей, которые используются при построении хранилища данных, поэтому отдельный раздел курса посвящен архитектуре и особенностям различных видов накопителей.

Цель преподавания учебной дисциплины «Системы хранения данных»: создание базы для применения современных методов сбора и анализа данных для решения практических задач и формирование у студентов магистратуры умения создавать необходимую архитектуру хранилища данных для анализа больших массивов данных с целью получения агрегированной информации. При изложении материала учебной дисциплины важно показать особенности работы с распределенной инфраструктурой, уделить особое внимание природе возникновения различий в характеристиках различных накопителей.

В рамках поставленной цели **задачи** учебной дисциплины состоят в следующем:

- изучение подходов к созданию хранилищ данных;
- изучение методов обработки и анализа многомерных данных;
- изучение различий в архитектуре накопителей данных, особенностей файловых систем.

Учебная дисциплина «Системы хранения данных» относится к компоненту учреждения высшего образования цикла дисциплин специальной подготовки.

Учебная программа составлена с учетом межпредметных связей с учебными дисциплинами. Так, основой для изучения учебной дисциплины «Системы хранения данных» являются следующие учебные дисциплины I ступени высшего образования: «Архитектура компьютеров» и «Базы данных», и II ступени высшего образования: «Технологии и компьютерные системы обработки больших объемов информации» и «Операционные системы семейства Unix». Знания, полученные в результате изучения дисциплины, будут использованы при изучении дисциплины II ступени высшего образования «Машинное обучение на больших массивах данных», а также способствовать успешному прохождению производственной практики по специальности и подготовки магистерской диссертации.

В результате освоения учебной дисциплины «Системы хранения данных» студент магистратуры должен:

знать:

- особенности построение распределенных хранилищ данных;
- основные подходы к анализу многомерных данных;
- полный цикл сбора, хранения и обработки данных;

уметь:

- подбирать подходящие решения для создания хранилищ данных, соответствующих заданным требованиям;
- подбирать подходящие инструменты для обработки данных;

владеть:

- программными средствами для моделирования и настройки распределенных систем в облачной инфраструктуре;
- навыками анализа требований к созданию хранилища данных и преобразования требований в соответствующие решения;
- навыками разработки инструментов обработки данных.

Освоение учебной дисциплины «Системы хранения данных» должно обеспечить формирование следующих социально-личностных и профессиональных компетенций:

социально-личностные компетенции:

СЛК-1. Учитывать социальные и нравственно-этические нормы в социально-профессиональной деятельности.

СЛК-2. Быть способным к сотрудничеству и работе в команде.

СЛК-3. Владеть коммуникативными способностями для работы в междисциплинарной и международной среде.

СЛК-4. Совершенствовать и развивать свой интеллектуальный и общекультурный уровень, добиваться нравственного и физического совершенствования своей личности.

СЛК-6. Логично, аргументировано и ясно строить устную и письменную речь, использовать навыки публичной речи, ведения дискуссии и полемики.

СЛК-7. Проявлять инициативу и креативность, в том числе в нестандартных ситуациях);

профессиональных компетенции:

ПК-3. Разрабатывать эффективные численные алгоритмы и интегрировать их в компьютерные системы.

Структура содержания учебной дисциплины включает такие дидактические единицы, как разделы (темы), в соответствии с которыми разрабатываются и реализуются соответствующие лекционные и практические занятия. Примерная тематика занятий приведена в информационно-методической части.

Дисциплина изучается в 4 семестре. Всего на изучение учебной дисциплины «Системы хранения данных» отведено 104 часа, в том числе 40 аудиторных часов, из них: лекции – 20 часа, практические занятия – 20 часов

Трудоемкость учебной дисциплины составляет 3 зачётные единицы.

Форма текущей аттестации – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел I. Архитектура систем хранения данных

Тема 1.1. Введение

Что такое данные. Какие данные необходимо хранить. Методы использования накопленных данных. Примеры использования данных.

Тема 1.2. Хранилища данных. Технологии хранения данных

Понятие хранилища данных. Распределенные и централизованные хранилища данных. Репликация. Сжатие данных. Методы восстановления поврежденных данных.

Тема 1.3. Хранение данных в многомерной модели

Многомерное представление данных. Преимущества, основные операции многомерного анализа данных (OLAP). OLAP кубы. Построение OLAP срезов.

Раздел II. Распределенные системы хранения данных

Тема 2.1. Логическая и физическая модели распределенной системы хранения данных.

Архитектура распределенной системы хранения данных. Слой представления данных. Слой бизнес-логики. Слой хранения данных. Объектно-реляционное отображение. Топологии и расширения сетей. Увеличение пропускной способности сети.

Тема 2.2. Распределенные системы хранения данных в виртуальном окружении

Облачные сервисы: особенности, преимущества, недостатки. Настройка облачного окружения. Создание и настройка тестовых виртуальных окружений.

Раздел III. Архитектура устройств хранения данных

Тема 3.1. Файловая система

Понятие файловой системы. Отображение логических и физических адресов. Ограничения файловой системы. Файловая система для распределенных систем.

Тема 3.2. Накопители на жестких дисках (HDD)

Архитектура жесткого диска. Кодирование и декодирование данных. Технология RAID. Стандартные характеристики. Дефрагментация.

Тема 3.3. Твердотельный накопитель (SSD)

Архитектура твердотельного накопителя: SLC, MLC, TLC. Преимущества и недостатки. Сборка мусора. Контроль износа памяти. Стандартные характеристики.

Тема 3.4. Виртуальный жесткий диск. Диск в памяти

Архитектура виртуального жесткого диска. Цели создания виртуальных дисков. Стандартные характеристики. Архитектура дисков в памяти. Цели создания и преимущества.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

№ п/п	Название раздела, темы	Количество аудиторных часов		Форма контроля знаний
		Лекции	Практические занятия	
1	Архитектура систем хранения данных	6	6	
1.1	Введение	1		
1.2	Хранилища данных. Технологии хранения данных	3		Устный опрос
	<i>Разработка схемы сбора, хранения и использования данных</i>		2	Отчёт по заданию.
1.3	Хранение данных в многомерной модели	2	2	Устный опрос. Контрольная работа №1.
	<i>Создание многомерного хранилища данных</i>		2	Отчёт по заданию.
2	Распределенные системы хранения данных	4	6	
2.1	Логическая и физическая модели распределенной системы хранения данных	2		Устный опрос
	<i>Разработка логической модели и анализ физической модели для проектируемой системы хранения данных</i>		2	Отчёт по заданию.
2.2	Распределенные системы хранения данных в виртуальном окружении.	2		Коллоквиум.
	<i>Создание системы сбора и хранения данных в облачной инфраструктуре</i>		4	Защита выполненного проекта
3.	Архитектура устройств хранения данных	10	8	
3.1	Файловая система	2		Устный опрос
3.2	Накопители на жестких дисках (HDD)	2		Устный опрос
	<i>Исследование скорости</i>		4	Отчёт по заданию.

	<i>работы HDD в зависимости от операционной системы и файловой системы</i>			
3.3	Твердотельный накопитель (SSD)	4	4	Устный опрос. Контрольная работа №2.
3.4	Виртуальный жесткий диск. Диск в памяти	2		Устный опрос
ИТОГО		20	20	

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Рекомендуемая литература

Основная

1. Paulraj Ponniah, Data Warehousing Fundamentals for IT Professionals, John Wiley & Sons, 2011. – 608 p. – ISBN 978-0-470-46207-2.
2. Rino Michelsoni, Solid-State-Drives (SSDs) Modeling: Simulation Tools & Strategies, Springer, 2017. – 170 p. – ISBN 9783319517353.
3. Faithe Wempen, Computing Fundamentals: IC3 Edition, John Wiley & Sons, 2014. – 752 p. – ISBN 9781118910160.

Дополнительная

4. Mark Humphries, Michael W Hawkins, Michelle C Dy Data warehousing: architecture and implementation, Upper Saddle River, N.J.: Prentice Hall PTR, 1999 – 365 p. – ISBN: 0-13-080902-0.
5. Sudharsan Seshadri, Software Defined SSD: An Architecture to Generalize Data Processing in SSDs, 2013 – ISBN 9781303457647.
6. Yuan Xie, Emerging Memory Technologies: Design, Architecture, and Applications, Springer Science & Business Media, 2013. – 322 p. – ISBN 9781441995513.
7. Paul Rogers, Robert Hering, IBM Redbooks, z/OS Distributed File Service zSeries File System Implementation z/OS V1R13, IBM Redbooks, 2012. – 518 p. – ISBN 9780738436401.

Рекомендуемая тематика контрольных работ

- 1) Контрольная работа №1. *Анализ многомерных данных.*
- 2) Контрольная работа №2. *Анализ работы сборщика мусора.*

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы магистрантов по учебной дисциплине следует использовать современные информационные технологии: разместить в сетевом доступе комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, презентации лекций,

методические указания к практическим занятиям, электронные версии домашних заданий, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к экзамену, задания, тесты, вопросы для самоконтроля, список рекомендуемой литературы, информационных ресурсов и др.).

Эффективность самостоятельной работы студентов магистратуры проверяется в ходе текущего и итогового контроля знаний. Для общей оценки качества усвоения студентами учебного материала рекомендуется использование рейтинговой системы.

Перечень рекомендуемых средств диагностики

Для текущего контроля качества усвоения знаний студентами магистратуры используется следующий диагностический инструментарий:

1. Устная форма: устные опросы, устные защиты домашних заданий, проведение коллоквиума «Построение системы хранения данных в виртуальном окружении».

2. Письменная форма: отчеты по самостоятельным заданиям, письменные контрольные работы по отдельным темам учебной дисциплины.

3. Техническая форма: система AnyTask (<https://anytask.org/school/bsu>) – инструменты с эффективной функциональностью контроля, тренинга и самостоятельной работы.

4. Система Insight Runner 2.0 (<https://acm.bsui.by>) – автоматизированная система проверки заданий с проверкой на плагиат.

Методика формирования итоговой оценки

Формой текущей аттестации по дисциплине «Системы хранения данных» учебным планом предусмотрен экзамен.

Рекомендуется использовать рейтинговую оценку знаний студента магистратуры, дающую 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. Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний в рейтинговую оценку:

1) работа на практических занятиях – 0.5;

2) контрольные работы – 0.3;

3) коллоквиумы – 0.2.

Рейтинговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости и экзаменационной оценки с учетом их весовых коэффициентов. Оценка по текущей успеваемости составляет: 40 %, экзаменационная оценка – 70 %.

Итоговая оценка формируется на основе:

- 1) Правил проведения аттестации студентов (Постановление Министерства образования Республики Беларусь № 53 от 29 мая 2012 г.);
- 2) Положение о рейтинговой системе оценки знаний по дисциплине в БГУ (Приказ ректора БГУ от 18.08.2015 № 382-ОД);
- 3) Критериев оценки знаний студентов (письмо Министерства образования от 22.12.2003).

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)
Машинное обучение на больших массивах данных	Дискретной математики и алгоритмики	нет	Оставить содержание учебной дисциплины без изменения, протокол № 14 от 26.04.2018 г.

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ

на ____ / ____ учебный год

№№ ПП	Дополнения и изменения	Основание

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

Заведующий кафедрой

(ученая степень, звание)

(подпись)

(И.О. Фамилия)

УТВЕРЖДАЮ

Декан факультета

(ученая степень, звание)

(подпись)

(И.О. Фамилия)