

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Алтайский государственный технический университет им. И.И. Ползунова»

СОГЛАСОВАНО

Декан ФИТ

А.С. Авдеев

Рабочая программа дисциплины

Код и наименование дисциплины: **Б1.В.ДВ.2.2 «Технологии хранения и обработки данных в распределенных системах»**

Код и наименование направления подготовки (специальности): **09.04.04**

Программная инженерия

Направленность (профиль, специализация): **Разработка программно-информационных систем**

Статус дисциплины: **элективные дисциплины (модули)**

Форма обучения: **очная**

Статус	Должность	И.О. Фамилия
Разработал	доцент	В.С. Троицкий
Согласовал	Зав. кафедрой «ПМ»	Е.Г. Боровцов
	руководитель направленности (профиля) программы	С.М. Старолетов

г. Барнаул

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция	Содержание компетенции	Индикатор	Содержание индикатора
ПК-1	Владение методами программной реализации распределенных информационных систем	ПК-1.1	Осуществляет выбор методов программной реализации распределенных информационных систем
		ПК-1.2	Создает программное обеспечение распределенных информационных систем
ПК-5	Понимание существующих подходов к верификации моделей программного обеспечения	ПК-5.1	Анализирует существующие подходы к верификации моделей программного обеспечения
		ПК-5.2	Применяет в профессиональной деятельности существующие подходы к верификации моделей программного обеспечения
ПК-10	Владение навыками организации промышленного тестирования создаваемого программного обеспечения	ПК-10.1	Анализирует и выбирает методы тестирования создаваемого программного обеспечения
		ПК-10.2	Осуществляет тестирование создаваемого программного обеспечения

2. Место дисциплины в структуре образовательной программы

Дисциплины (практики), предшествующие изучению дисциплины, результаты освоения которых необходимы для освоения данной дисциплины.	Программирование параллельных процессов
Дисциплины (практики), для которых результаты освоения данной дисциплины будут необходимы, как входные знания, умения и владения для их изучения.	Преддипломная практика, Распределенные системы обработки информации, Функциональные языки разработки распределенных систем

3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающегося с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающегося

Общий объем дисциплины в з.е. /час: 4 / 144

Форма промежуточной аттестации: Экзамен

Форма обучения	Виды занятий, их трудоемкость (час.)				Объем контактной работы обучающегося с преподавателем (час)
	Лекции	Лабораторные работы	Практические занятия	Самостоятельная работа	
очная	16	32	0	96	57

4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Форма обучения: очная

Семестр: 2

Лекционные занятия (16ч.)

1. Введение в дисциплину. Методы программной реализации распределенных информационных систем. {лекция с разбором конкретных ситуаций} (2ч.)[1,3,4,5] Основные понятия. Принципы и методы построения распределенных систем хранения и обработки информации. Тенденции, этапы развития, требования. Распределенные базы данных (РБД), общий обзор. Концепции облачных вычислений, программное обеспечение, как сервис (Software as a service, SaaS), инфраструктура, как сервис (Infrastructure as a service, IaaS), платформа, как сервис (Platform as a service, PaaS). Что и когда нужно переводить в облако. Особенности облачных решений, в том числе выбор методов тестирования облачного программного обеспечения .

2. Архитектура СУРБД {лекция с разбором конкретных ситуаций} (2ч.)[1,3,4,5] Общие принципы организации распределённых баз данных. Критерии распределенности. Преимущества и недостатки РБД. Функции СУРБД. Архитектура СУРБД. Методы поддержки распределенных данных. Методы проектирования распределенных БД. Управление распределенными транзакциями. Оптимизация распределенных запросов. Глобальная оптимизация. Исследовательские проблемы в области РБД.

3. Распределенные вычисления, модель вычислений MapReduce. Создание программного обеспечения распределенных информационных систем на платформе Apache Hadoop. {лекция с разбором конкретных ситуаций} (4ч.)[1,3,4,5] Распределенные вычисления. Цели применения параллельных вычислений. Пути достижения параллелизма. Оценки времени выполнения параллельного алгоритма. Характеристики параллельных алгоритмов: ускорение, эффективность, стоимость. Принципы разработки параллельных алгоритмов. основополагающие принципы MapReduce, общая модель программирования и принципы параллельной реализации вычислений, применение, преимущества и недостатки. Архитектура платформы Apache Hadoop и ее основные элементы. Создание программного обеспечения распределенных информационных систем на платформе Apache Hadoop. Интерфейс прикладного программирования и общие принципы реализации приложений для Hadoop на языке Java. Подходы к верификации моделей программного обеспечения распределенных

информационных систем.

12. Распределенная файловая система HDFS {лекция с разбором конкретных ситуаций} (2ч.)[1,3,4,5] HDFS. Сервер файлов контрольных точек (CheckpointNode), сервер резервных копий (BackupNode). Обновления и снимки файловой системы. Чтение и запись файлов. Размещение блоков. Управление репликацией. Балансировщик. Сканер блоков. Копирование данных между кластерами.

Долговечность хранения данных. Возможности совместного использования ресурсов HDFS. Масштабирование и объединение файловой системы. Конфигурация Hadoop для распределенной работы. □ Дистрибутивы Hadoop. Аппаратные требования. Установка и настройка. Пример кластера ClouderaHadoop. Потребности приложений Hadoop. Работа с HDFS из Java. Разделение Hadoop на главный и подчиненные узлы. Конфигурация кластера Hadoop. Обновление конфигурации Hadoop. Записи для узлов Hadoop в файле. Определение хозяина HDFS в файле core-site.xml Запуск демонов MapReduce. Получение информации о запущенных процессах на одном из подчиненных узлов. Тестирование HDFS. Тестирование файловой системы HDFS □ Проверка файловой системы HDFS. Параллельная обработка больших массивов данных. Распределенные хранилища данных. Распределенное программирование.

14. Azure Services Platform {лекция с разбором конкретных ситуаций} (4ч.)[1,2,3,4,5] Установка и настройка программного обеспечения, знакомство с инструментарием Azure Services Platform, создание первого облачного приложения; создание и настройка простого облачного приложения, свойства рабочей роли и веб-роли, запуск приложения в режиме эмуляции

Структурированное хранилище данных Windows Azure Table: модель данных Windows Azure Table и принципы секционирования, подключение к хранилищу разработки, создание хранилища с простой структурой данных. Разработка хранилища Windows Azure Table с реляционной структурой: базовые операции таблиц и сущностей в Windows Azure Table. Манипулирование данными в Windows Azure Table через web-интерфейс: создание классов и веб-интерфейса облачного приложения, для добавления, просмотра, редактирования и удаления данных. Выборка данных в Windows Azure Table через web-интерфейс: изучение способов выборки данных из хранилища Windows Azure Table и способов представления данных посредством веб-приложений в среде Visual Studio 2010; усовершенствование веб-интерфейса облачного приложения для выборки данных по заданному критерию.

Работа с Windows Azure Blob: модель данных Windows Azure Blob, REST-интерфейс Blob-объектов, блоки и Blob-страницы, Windows Azure Blob, как набор блоков, REST-запросы. Работа с Windows Azure Queue: модель данных и REST-интерфейс Windows Azure Queue, примеры использования и REST-запросы. Организация тестирования создаваемого программного обеспечения.

16. Текущие задачи в области распределенной обработки данных {лекция с разбором конкретных ситуаций} (2ч.)[1,3,4,5] Текущее состояние в области распределенной обработки данных, примеры технологий, задач и объемов

обрабатываемых данных. Обзор актуальных задач в области распределенной обработки данных.

Лабораторные работы (32ч.)

- 1. Azure Services Platform {разработка проекта} (4ч.)[1,2,3,4,5]** Установка и настройка программного обеспечения, знакомство с инструментарием Azure Services Platform, создание и настройка простого облачного приложения, свойства рабочей роли и веб-роли, запуск приложения в режиме эмуляции.
- 2. Структурированное хранилище данных Windows Azure Table и принципы секционирования {разработка проекта} (4ч.)[1,2,3,4,5]** Подключение к хранилищу разработки, создание хранилища с простой структурой данных
- 3. Разработка хранилища Windows Azure Table с реляционной структурой {разработка проекта} (4ч.)[1,2,3,4,5]** Создать хранилище с реляционной структурой и продемонстрировать базовые операции таблиц и сущностей в Windows Azure Table
- 4. Манипулирование и выборка данных в Windows Azure Table через web-интерфейс {разработка проекта} (4ч.)[1,2,3,4,5]** Создание классов и веб-интерфейса облачного приложения, для добавления, просмотра, редактирования и удаления данных. Выборка данных через web-интерфейс. Изучение способов выборки данных из хранилища Windows Azure Table и способов представления данных посредством веб-приложений в среде Visual Studio; усовершенствование веб-интерфейса облачного приложения для выборки данных по заданному критерию
- 5. Работа с Windows Azure Blob и Azure Queue {разработка проекта} (4ч.)[1,2,3,4,5]** Модель данных Windows Azure Blob, REST-интерфейс Blob-объектов, блоки и Blob-страницы, Windows Azure Blob, как набор блоков, REST-запросы. Работа с Windows Azure Queue – модель данных и REST-интерфейс Windows Azure Queue, примеры использования и REST-запросы
- 6. Map и Reduce в платформе Apache Hadoop {разработка проекта} (4ч.)[1,3,4,5]** Практическое знакомство с платформой Apache Hadoop, реализация модели вычисления MapReduce для решения задач распределенной обработки данных.
- 7. HDFS {разработка проекта} (4ч.)[1,3,4,5]** Сконфигурировать HDFS, установив количество копий в 4 (репликация), и проверить работу файловой системы с этой конфигурацией.
- 8. Hadoop кластер {разработка проекта} (4ч.)[1,3,4,5]** Запустить Hadoop кластер из трех нод – 1 главной и 2 подчиненных

Самостоятельная работа (96ч.)

- 1. Подготовка к лекциям(24ч.)[1,3,4,5]**
- 2. Оформление отчетов по лабораторным работам и подготовка к их защите(36ч.)[1,3,4,5]**

3. Подготовка к сдаче экзамена(36ч.)[1,3,4,5]

5. Перечень учебно-методического обеспечения самостоятельной работы обучающихся по дисциплине

Для каждого обучающегося обеспечен индивидуальный неограниченный доступ к электронно-библиотечным системам: Лань, Университетская библиотека он-лайн, электронной библиотеке АлтГТУ и к электронной информационно-образовательной среде:

1. Старолетов С. М. Распределенные системы: Учебно-методическое пособие.— Барнаул:

АлтГТУ, 2016. — 77с. Прямая ссылка: http://elib.altstu.ru/eum/download/pm/Staroletov_ds_bi.pdf

2. Развертывание приложений в Windows Azure Троицкий В.С. (ПМ) 2021 Учебно-методическое пособие, 2.97 МБ Дата первичного размещения: 13.05.2021. Обновлено: 13.05.2021.

Прямая ссылка: http://elib.altstu.ru/eum/download/pm/Troickiy_WinAzure_ump.pdf

6. Перечень учебной литературы

6.1. Основная литература

3. Чак, Л. Nadoop в действии / Л. Чак. — Москва : ДМК Пресс, 2012. — 424 с. — ISBN 978-5-94074-785-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/39997> (дата обращения: 10.05.2021). — Режим доступа: для авториз. пользователей.

4. Теджасви, Р. Платформа Windows Azure / Р. Теджасви, Г. Тони. — Москва : ДМК Пресс, 2012. — 656 с. — ISBN 978-5-94074-654-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/39991> (дата обращения: 10.05.2021). — Режим доступа: для авториз. пользователей.

6.2. Дополнительная литература

5. Новиков, Б. А. Основы технологий баз данных / Б. А. Новиков ; под редакцией Е. В. Рогова. — Москва : ДМК Пресс, 2019. — 240 с. — ISBN 978-5-94074-820-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/123699> (дата обращения: 10.05.2021). — Режим доступа: для авториз. пользователей.

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

6. <https://azure.microsoft.com/ru-ru/> -Создание безопасных перспективных облачных решений в локальной, гибридной, многооблачной или пограничной среде с помощью Azure

7. <https://hadoop.apache.org/> -разработка программного обеспечения с открытым исходным кодом для надежных, масштабируемых распределенных вычислений.

8. Фонд оценочных материалов для проведения текущего контроля успеваемости и промежуточной аттестации

Содержание промежуточной аттестации раскрывается в комплекте контролирующих материалов, предназначенных для проверки соответствия уровня подготовки по дисциплине требованиям ФГОС, которые хранятся на кафедре-разработчике РПД в печатном виде и в ЭИОС.

Фонд оценочных материалов (ФОМ) по дисциплине представлен в приложении А.

9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем

Для успешного освоения дисциплины используются ресурсы электронной информационно-образовательной среды, образовательные интернет-порталы, глобальная компьютерная сеть Интернет. В процессе изучения дисциплины происходит интерактивное взаимодействие обучающегося с преподавателем через личный кабинет студента.

№пп	Используемое программное обеспечение
1	Git
2	LibreOffice
3	Mozilla Firefox
4	Visual Studio
5	Windows
6	Антивирус Kaspersky

№пп	Используемые профессиональные базы данных и информационные справочные системы
1	Бесплатная электронная библиотека онлайн "Единое окно к образовательным ресурсам" для студентов и преподавателей; каталог ссылок на образовательные интернет-ресурсы (http://Window.edu.ru)
2	Крупнейший веб-сервис IT-проектов и их совместной разработки (https://github.com/)
3	Национальная электронная библиотека (НЭБ) — свободный доступ читателей к фондам российских библиотек. Содержит коллекции оцифрованных документов (как открытого доступа, так и ограниченных авторским правом), а также каталог изданий, хранящихся в библиотеках России. (http://нэб.рф/)
4	Программа Microsoft и интернет-ресурс, содержащий техническую информацию, новости и предстоящие события для профессионалов в сфере информационных технологий. На данный момент представляет собой сборник технической информации на русском языке для IT-специалистов (https://technet.microsoft.com/ru-ru/)

№пп	Используемые профессиональные базы данных и информационные справочные системы
	https://docs.microsoft.com/ru-ru/welcome-to-docs)

10. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Наименование специальных помещений и помещений для самостоятельной работы
учебные аудитории для проведения учебных занятий
помещения для самостоятельной работы

Материально-техническое обеспечение и организация образовательного процесса по дисциплине для инвалидов и лиц с ограниченными возможностями здоровья осуществляется в соответствии с «Положением об обучении инвалидов и лиц с ограниченными возможностями здоровья».