

## АННОТАЦИЯ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ «Технологии хранения и обработки данных в распределенных системах»

по основной профессиональной образовательной программе по направлению подготовки  
09.04.04 «Программная инженерия» (уровень магистратуры)

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

**Общий объем дисциплины** – 4 з.е. (144 часов)

**Форма промежуточной аттестации** – Экзамен.

**В результате освоения дисциплины у обучающихся должны быть сформированы компетенции с соответствующими индикаторами их достижения:**

- ПК-1.1: Осуществляет выбор методов программной реализации распределенных информационных систем;
- ПК-1.2: Создает программное обеспечение распределенных информационных систем;
- 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-10.1: Анализирует и выбирает методы тестирования создаваемого программного обеспечения;
- ПК-10.2: Осуществляет тестирование создаваемого программного обеспечения;

**Содержание дисциплины:**

Дисциплина «Технологии хранения и обработки данных в распределенных системах» включает в себя следующие разделы:

**Форма обучения очная. Семестр 2.**

**1. Введение в дисциплину. Методы программной реализации распределенных информационных систем..** 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.** 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..**

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.** 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.** 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.** Текущее состояние в области распределенной обработки данных, примеры технологий, задач и объемов обрабатываемых данных. Обзор актуальных задач в области распределенной обработки данных..

Разработал:  
доцент  
кафедры ПМ

В.С. Троицкий

Проверил:  
Декан ФИТ

А.С. Авдеев