



ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
УПРАВЛЕНИЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ И ПОВЫШЕНИЯ
КВАЛИФИКАЦИИ

Кафедра «Программное обеспечение вычислительной тех-
ники и автоматизированных систем»

Учебно-методическое пособие по дисциплине

«РАСПРЕДЕЛЕННЫЕ ПРИЛОЖЕНИЯ И ПРОТОКОЛЫ»

Автор
Долгов В.В.

Ростов-на-Дону, 2018

Аннотация

Учебно-методическое пособие предназначено для студентов очной формы обучения направления 09.04.04 «Программная инженерия».

Авторы

доцент, к.т.н.,
зав.каф. ПОВТиАС
Долгов В.В.



Оглавление

1. Общие указания к выполнению лабораторных работ	4
1.1. Требование к лабораторному оборудованию	4
1.2. Требования, предъявляемые при сдаче лабораторных работ	4
2. Лабораторная работа №1: Распределение вычислительной задачи в системах с централизованным управлением	5
2.1. Цель работы	5
2.2. Задание к лабораторной работе	5
2.3. Контрольные вопросы	6
3. Лабораторная работа №2: Распределение вычислительной задачи в децентрализованных системах	6
3.1. Цель работы	6
3.2. Задание к лабораторной работе	7
3.3. Контрольные вопросы	7
4. Лабораторная работа №3: Разработка централизованной распределенной системы хранения информации	7
4.1. Цель работы	7
4.2. Задание к лабораторной работе	8
4.3. Дополнительное (необязательное) задание	8
4.4. Контрольные вопросы	9
5. Лабораторная работа №4: Разработка децентрализованной распределенной системы хранения информации	9
5.1. Цель работы	9
5.2. Задание к лабораторной работе	9
5.3. Контрольные вопросы	10
Список литературы	11

1. Общие указания к выполнению лабораторных работ

1.1. Требование к лабораторному оборудованию

Аудитория для проведения лабораторных занятий должна быть укомплектована специализированной мебелью и индивидуальными компьютерами следующей минимальной комплектации:

- Процессор: не менее двух исполнительных ядер, совместимый с системой команд x86 и x64, с поддержкой аппаратной виртуализации.

- Оперативная память: не менее 4 Гб.

- Монитор: не менее 22" (дюймов) по диагонали.

- Наличие локальной сети со скоростью обмена не менее 1 Гб/сек.

- Наличие доступа в сеть Интернет со скоростью не менее 1 Мбит/сек.

- Наличие клавиатуры и манипулятора «мышь».

На компьютерах должно быть установлено следующее программное обеспечение:

- Операционная система: Microsoft Windows 7 (или выше) и Ubuntu Desktop 14.04 (или выше). Допускается конфигурация, когда одна из операционных систем установлена внутри виртуальной машины (гипервизора).

- Среды разработки программ: Microsoft Visual Studio 2015 (или выше), IntelliJ IDEA Community Edition, Eclipse.

1.2. Требования, предъявляемые при сдаче лабораторных работ

При сдаче студентом лабораторной работы отчетом выступает исходный код созданной в процессе выполнения работы программы.

Исходный код должен быть отформатирован согласно принятым для языка F# стандартам. Допускается наличие в исходном коде комментариев, описывающих основные части программы и особенности их функционирования. В то же время студент должен быть готов объяснить работу программы в целом и каждую отдельную ее часть при полном отсутствии комментариев (например, они могут быть удалены или сделаны нечитаемыми перед сдачей).

Обязательным условием сдачи является умение студента восстановить любой участок исходного кода программы после его удаления (при этом этот участок должен быть набран студентом

заново, запрещается использование операции Undo текстовых редакторов).

Исходный код сдаваемой программы должен быть представлен в электронном виде.

Сдача исходного кода ранее сдававшихся программ или программ, код которых выложен в сети Интернет, не допускается.

2. ЛАБОРАТОРНАЯ РАБОТА №1: РАСПРЕДЕЛЕНИЕ ВЫЧИСЛИТЕЛЬНОЙ ЗАДАЧИ В СИСТЕМАХ С ЦЕНТРАЛИЗОВАННЫМ УПРАВЛЕНИЕМ

2.1. Цель работы

Получить навыки решения сложно-вычислительных задач в распределенных системах с центральным управлением. Получить навыки создания и/или использования систем организации распределенных вычислений в кластерных системах.

2.2. Задание к лабораторной работе

Разработать информационную систему (программу), решающую в параллельном режиме одну из задач (согласно варианту задания) из таблицы 2.1.

При решении задачи должны быть задействованы не менее трех вычислительных узлов локальной сети (отдельные компьютеры, виртуальные машины или docker-контейнеры), один из которых является центральным (управляющим).

При решении допускается как использование уже готовых систем распределения заданий (OpenMP, Parallel Virtual Machine) так и создание своей собственной распределяющей системы.

Таблица 2.1 – Варианты заданий к лабораторной работе №1

№ варианта	Задание
1	Оптимизация целевой функции генетическим алгоритмом (островная модель). Каждый вычислительный узел сети представляет изолированный «остров», производящий локальную оптимизацию. Центральный узел выбирает лучшее решение из локальных результатов, полученных на «островах».

2	Поиск всех простых чисел на заданном интервале (значения чисел не вмещаются в стандартные элементарные типы данных). Проверку на простоту чисел можно производить на основе вероятностных алгоритмов проверки ().
3	Распараллелить решение большой системы линейных уравнений. Исходные матрицы системы задаются через файлы в формате CSV. Метод распределения по узлам – произвольный.

2.3. Контрольные вопросы

1. В чем преимущество использования централизованных распределенных вычислительных систем?

2. Каким условиям должна удовлетворять централизованная вычислительная задача, чтобы ее расчет был выгоден на распределенной системе?

3. Какие основные библиотеки используются для организации централизованных распределенных вычислений? Опишите кратко каждую из них?

4. Какие основные языки программирования используются при организации распределенных вычислений?

5. Надо ли учитывать задержки при передачи данных в сети при распределении вычислений?

6. Как влияет количество вычислительных узлов на коэффициент ускорения решаемой задачи? Зависит ли это от самой задачи?

3. ЛАБОРАТОРНАЯ РАБОТА №2: РАСПРЕДЕЛЕНИЕ ВЫЧИСЛИТЕЛЬНОЙ ЗАДАЧИ В ДЕЦЕНТРАЛИЗОВАННЫХ СИСТЕМАХ

3.1. Цель работы

Получить навыки решения вычислительных задач в распределенных децентрализованных системах. Получить навыки создания и/или использования систем организации распределенных децентрализованных вычислений.

3.2. Задание к лабораторной работе

Разработать информационную систему (программу), решающую в параллельном режиме одну из задач (согласно варианту задания) из таблицы 2.1.

При решении задачи должны быть задействованы не менее трех вычислительных узлов локальной сети (отдельные компьютеры, виртуальные машины или docker-контейнеры) ни один из которых не должен быть выделен в качестве центрального. Отказ любого из работающих узлов не должен приводить к сбою в работе системы.

Отправка задачи на решение, взаимодействие узлов между собой, а также получение окончательного ответа должно производиться с использованием механизма сообщений, передаваемых между узлами сети.

При решении допускается как использование уже готовых систем взаимодействия узлов (очереди сообщений, локальные облачные системы, системы независимых агентов) так и создание своей собственной системы.

3.3. Контрольные вопросы

1. В чем отличие децентрализованных вычислительных сетей от сетей с централизованным управлением?
2. С какими трудностями сталкиваются разработчики при организации децентрализованных вычислительных сетей? Какие преимущества они получают?
3. Что такое очереди сообщений? В каких сценариях организации вычислений они могут использоваться?
4. Какие различают режимы работы у очередей сообщений? Какой из этих режимов наиболее подходит для организации распределенных вычислений?
5. Каким образом в децентрализованных системах решается проблема неожиданного отказа вычислительного узла?

4. ЛАБОРАТОРНАЯ РАБОТА №3: РАЗРАБОТКА ЦЕНТРАЛИЗОВАННОЙ РАСПРЕДЕЛЕННОЙ СИСТЕМЫ ХРАНЕНИЯ ИНФОРМАЦИИ

4.1. Цель работы

Получить навыки проектирования и разработки распределенной системы хранения файлов с централизованным управле-

нием.

4.2. Задание к лабораторной работе

Разработать простую файловую систему (ФС), осуществляющую распределенное хранение информации (файлов) в узлах сети. Большие файлы должны быть разбиты на отдельные блоки фиксированного размера каждый из которых может быть сохранен отдельно от других блоков того же файла.

Система должна поддерживать следующий набор обязательных функций:

- сохранение файла в распределенной системе под выбранным именем;
- получение целого файла по имени или любой его последовательности состоящей из подряд идущих блоков;
- замена любого целого блока файла на новый;
- проверка наличия файла с заданным именем в ФС;
- удаление файла по имени.

При решении задачи должны быть задействованы не менее трех подчиненных узлов, обеспечивающих хранение информации (отдельные компьютеры, виртуальные машины или docker-контейнеры), и один центральный (управляющий) узел. Управляющий узел может быть пространственно совмещен с одним из подчиненных узлов.

4.3. Дополнительное (необязательное) задание

Реализовать систему дедубликации данных на уровне блоков файлов. То есть в случае, если два разных блока данных (неважно одного файла или различных) совпадают друг с другом, хранить в системе только один из них.

Обратите внимание, что при реализации такой системы необходимо учитывать количество использований каждого хранимого блока в файлах, так как удаление блока из системы возможно только когда он не используется ни в одном из хранимых файлов.

4.4. Контрольные вопросы

1. Назовите и опишите основные характеристики современных распределенных файловых систем.
2. В чем причина появления кластерных ФС? Какие основные проблемы они решают?
3. Что такое репликация данных? Какой коэффициент репликации на сегодняшний день считается приемлемым?
4. Что такое дедубликация данных? Какие технологии позволяют реализовать дедубликацию?
5. В чем сильные и слабые стороны централизованных распределенных ФС?
6. С какими проблемами приходится сталкиваться разработчикам централизованных распределенных ФС?

5. ЛАБОРАТОРНАЯ РАБОТА №4: РАЗРАБОТКА ДЕЦЕНТРАЛИЗОВАННОЙ РАСПРЕДЕЛЕННОЙ СИСТЕМЫ ХРАНЕНИЯ ИНФОРМАЦИИ

5.1. Цель работы

Закрепить теоретические знания алгоритмов поиска информации в децентрализованных (пиринговых) системах. Получить навыки проектирования и разработки распределенной децентрализованной системы хранения информации

5.2. Задание к лабораторной работе

Разработать программную систему для распределенного децентрализованного хранения данных вида «Ключ-Значение» (Key-Value).

Система должна поддерживать следующий набор обязательных функций:

- сохранение (замена) блока данных ассоциированного с указанным ключом;
- получение блока данных по ключу;
- проверка наличия сохраненного ключа в системе;
- удаление ключа из системы (вместе с ассоциированными данными).

Распределенные приложения и протоколы

При решении задачи должны быть задействованы не менее четырех узлов, обеспечивающих хранение информации (отдельные компьютеры, виртуальные машины или docker-контейнеры).

При решении задачи допускается как использование готового алгоритма поиска информации в децентрализованных сетях (Pastry, Tapestry, Chord, Kademia, др.), так и разработка своего собственного алгоритма

5.3. Контрольные вопросы

1. В чем преимущество распределенных децентрализованных ФС перед централизованными? В чем недостатки?
2. Какой алгоритмической сложностью должны обладать алгоритмы поиска в децентрализованных сетях?
3. Что такое CAP-теорема? Какие ограничения она накладывает на распределенные децентрализованные системы?
4. Объясните работу одного из алгоритмов поиска данных в децентрализованных сетях?
5. В чем заключается проблема публикации данных в децентрализованных сетях?
6. Какие общие принципы построения логических часов в распределенных системах?

СПИСОК ЛИТЕРАТУРЫ

1. Э. Таненбаум Компьютерные сети. – СПб: Питер, 2012.
2. Федотов И.Е. Модели параллельного программирования. – СОЛОН-Пресс, 2012. – URL: <https://e.lanbook.com/book/13807>
3. Болодурина И.П., Волкова Т.В. Проектирование компонентов распределенных информационных систем. – Оренбургский государственный университет, ЭБС АСВ, 2012. – URL: <http://www.iprbookshop.ru/30122.html>
4. Э. Таненбаум, М. ван Стеен Распределенные системы. Принципы и парадигмы. – СПб.: Питер, 2003.