

Программа учебной дисциплины «Распределенные информационные системы»

Утверждена

Академическим советом ООП

Протокол № 10 от «26» 06 2018 г.

Автор	Самодуров В.А., к.ф.-м.н., vsamodurov@hse.ru
Число кредитов	4
Контактная работа (час.)	64
Самостоятельная работа (час.)	88
Курс	3
Формат изучения дисциплины	Без использования онлайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целью изучения дисциплины «Распределенные информационные системы» является подготовка специалистов, обладающих фундаментальными знаниями и практически-ми навыками в области построения распределенных информационных систем и сетей, программной инженерии, общей теории построения математических моделей и их реализации, глубоким знанием основ информатики, теории и практики руководства проектами по созданию распределенных информационных систем.

Настоящая дисциплина относится к циклу дисциплин программы «Бизнес-информатика», являясь дисциплиной по выбору. Курс рассчитан на студентов, имеющих подготовку по предшествующим курсам, касающихся основ программирования с использованием алгоритмических языков, вычислительным методам. В течение преподавания курса предполагается, что студенты знакомы с основными понятиями алгебры, комбинаторики, логики, информатики, которые читаются на факультете перед изучением данной дисциплины.

Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

• Знать:

- базовые принципы построения распределенных систем: согласованное время, распределенные транзакции;
- механизмы и средства современных операционных систем, используемые для построения распределенных приложений;
- владеть широкой общей подготовкой (базовыми знаниями) для решения практических задач в области информационных систем и технологий;
- проектирование базовых и прикладных информационных технологий;
- инструментальные программные средства, применяемые для построения распределенных систем.

• Уметь:

- выбрать набор средств операционной системы и инструментальных программных средств для разработки распределенного приложения;
- проектировать и разрабатывать распределенные информационные систем.
- заниматься разработкой и исследованием методик анализа, синтеза, оптимизации и прогнозирования качества процессов функционирования информационных систем и технологий
- разрабатывать средства реализации информационных технологий (методические, информационные, математические, алгоритмические, технические и программные);

- осуществлять инсталляции, отладки программных и настройки технических средств для ввода информационных систем в промышленную эксплуатацию;
- поддерживать работоспособность информационных систем и технологий в заданных функциональных характеристиках и соответствии критериям качества;

• Иметь навыки:

- описания задачи обработки данных;
- обосновать метод решения выбранной задачи;
- реализовать поставленную задачу на языке программирования
- создания приложений распределенных систем на основе систем баз данных.

Основные положения дисциплины должны быть использованы в дальнейшем при написании выпускной квалификационной работ, подготовке научных статей, докладов, презентаций исследовательских работ, в практической и исследовательской деятельности.

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Тематический план учебной дисциплины

№	Название тем	Всего часов			
			Лекции	Семинары	Самостоятельная работа
1	Понятие распределенной системы. Преимущества и недостатки распределенных систем. Масштабируемость. Прозрачность. Аппаратные и программные средства построения распределенных систем.	12	2	2	8
2	Связь в распределенных системах. Удаленный вызов процедур. Сохранность. Типы связей.	8	2	2	4
3	Средства современных ОС. Многозадачность. Многопоточность. Планировщик ОС. Изоляция приложений. Механизмы синхронизации процессов.	28	4	4	20
4	Синхронизация времени в распределенных системах.	14	4	4	6
5	Основные понятия теории реляционных СУБД. Структурированный язык запросов. Понятие транзакции.	27	6	6	15
6	Базы данных NoSQL	27	6	6	15
7	Распределенная система объектов CORBA.	9	2	2	5
8	Технология DCOM. Развитие модели COM. Управление жизненным циклом объекта.	9	2	2	5
9	Распределенные файловые системы. Файловая система NFS. Семантика совместного использования файлов. Проблема отказов.	9	2	2	5
10	Тенденции в области распределенных систем.	9	2	2	5
	Итого:	152	32	32	88

Тема 1. Понятие распределенной системы. Преимущества и недостатки распределенных систем. Масштабируемость. Прозрачность. Аппаратные и программные средства построения распределенных систем.

Тема 2. Связь в распределенных системах. Удаленный вызов процедур. Сохранность. Типы связей.

Тема 3. Средства современных ОС. Многозадачность. Многопоточность. Планировщик ОС. Изоляция приложений. Механизмы синхронизации процессов.

Тема 4. Синхронизация времени в распределенных системах. Необходимость. Алгоритм Кристиана. Алгоритм Беркли. Децентрализованный алгоритм. Логическое время. Алгоритмы голосования. Алгоритм забияки и кольцевой алгоритм. Алгоритмы взаимного исключения. Централизованный и распределенный алгоритмы, алгоритм маркерного кольца.

Тема 5. Основные понятия теории реляционных СУБД. Структурированный язык запросов. Реляционная модель данных. Основные понятия теории реляционных БД. Целостность базы данных. Структурированный язык запросов. Операции реляционной алгебры. Программирование приложений для СУБД. Понятие транзакции. Распределенные транзакции. Принцип ACID. Вложенные транзакции.

Тема 6. NoSQL – основные принципы в сравнении с реляционными базами данных. Теорема CAP и ее влияние на технологии NoSQL. Понятие функционального программирования. Технология MapReduce и Hadoop.

Тема 7. Распределенная система объектов CORBA. Архитектура CORBA. Объектный адаптер. IDL-стабы.

Тема 8. Технология DCOM. Развитие модели COM. Управление жизненным циклом объекта.

Тема 9. Распределенные файловые системы. Файловая система NFS. Семантика совместного использования файлов. Проблема отказов.

Тема 10. Тенденции в области распределенных систем.

Нерешенные и перспективные проблемы теории и практики распределенных систем. Направления исследований. Обработка информации в суперсетях (Грид). Архитектура Грид. Мобильный компьютеринг. Тотальный (pervasive) компьютеринг. Глобальное «умное» пространство.

III. ОЦЕНИВАНИЕ

В рамках дисциплины проводится изучение основных понятий облачных вычислений, необходимых для последующего изучения моделей, методов и технологий облачных вычислений.

Дисциплина изучается на лекциях и семинарских, практических занятиях.

На лекциях рассматриваются основополагающие теоретические вопросы и способы решения задач прикладного и системного характера. Лекции строятся на последовательном систематическом устном изложении преподавателем учебного материала, представляющего логически законченное целое. Каждая лекция сопровождается презентацией, содержащей краткий теоретический материал и иллюстративный материал.

Каждая презентация построена по следующему шаблону: название лекционного занятия, цель и задачи лекции, краткое содержание предыдущей лекции (при необходимости), теоретический материал, в конце приведены итоги лекционного занятия, обозначена тема следующей лекции, а также вопросы и задания для самостоятельного изучения.

Презентации по лекционному курсу разбиты по темам, по отдельно взятой теме может быть несколько лекций.

Семинарские занятия проводятся с целью освоения методов распределенных приложений. Студенты привлекаются к разбору и сравнительному анализу предлагаемых вариантов программных реализаций решаемых задач.

Успешное освоение дисциплины основывается на систематической повседневной работе студентов. В процессе самостоятельной работы студенты в течение одного – двух дней прорабатывают материалы лекционных и практических занятий по конспектам и рекомендованной основной литературе.

Написание реферата является одной из форм обучения студентов. Данная форма обучения направлена на организацию и повышение уровня самостоятельной работы студентов.

Реферат, как форма обучения студентов - это краткий обзор максимального количества доступных публикаций по заданной теме, подготовка самого реферативного обзора и презентации по нему. При проведении обзора должна проводиться и исследовательская работа, но объем ее ограничен, так как анализируются уже сделанные выводы и в связи с небольшим объемом данной формы работы. Преподавателю предоставляется сам реферат и презентация к нему. Сдача реферата происходит в форме защиты-доклада с использованием подготовленной презентации.

Критерии оценки знаний, навыков

Итоговая оценка по учебной дисциплине складывается из следующих элементов:

- тесты по материалам лекций (теория)
- работа на семинарских и практических занятиях (решение лабораторных работ с использованием компьютеров, доклады, обсуждения);
- экзамен

Структура экзаменационной оценки по учебной дисциплине:

Форма работы	Вклад в итоговую оценку (%)
Тесты по материалам лекций (теория) ТР	15
Задания на семинарах и практических занятиях ПР	25
Реферат РФ	20
Экзамен ЭКЗ	40

$$O(\text{общая}) = 0,15 * O(\text{ТР}) + 0,25 * O(\text{ПР}) + 0,20 * O(\text{РФ}) + 0,40 * O(\text{ЭКЗ}).$$

В конце 2-го модуля на зачетной неделе состоится итоговый контроль в форме зачета.

Используется арифметический способ округления оценок. В ведомость ставится результирующая итоговая оценка.

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

4.1 Примеры тем рефератов

"Железо", сетевые технологии, программирование, история технологий:

Динамика развития процессорной архитектуры

История развития UNIX систем.

Суперкомпьютеры

Параллельные вычисления при помощи Nvidia
Технология AJAX
Node.js
Потоковая обработка данных в реальном времени
Распределенные вычисления на платформе .NET

Сети:

Информационная безопасность распределённых информационных систем
Виртуальные частные сети
Архитектура серверов ММО (массовых многопользовательских онлайн-игр).
"Микросервисы" (Микросервисная архитектура приложений)
Распределенная платформа для параллельного обучения искусственных нейронных сетей DisANN
Сеть Tor
Механизмы функционирования воспроизведения видео и алгоритмов поиска системы Youtube

Олап-технологии:

Олап-технологии
Противостояние BI-системы QlikView и OLAP"

NoSQL технологии

Концепция и технологии NoSQL и их применение в распределенных информационных системах.
Системы управления базами данных (СУБД) Tarantool
Hadoop как платформа для построения распределенных приложений
Обзор NoSQL СУБД Redis
Apache Spark как эффективный инструмент анализа данных
Модель распределённых вычислений MapReduce
Язык программирования Erlang для создания распределенных вычислительных систем
Плюсы и минусы использования системы Hadoop
Документ-ориентированные хранилища
Teradata
NoSQL графовой базе данных Neo4j
NoSQL-технологии и распределенные информационные системы: система Couchbase Server

Облачные вычисления и интернет вещей:

Технологии облачных вычислений"
Высокопроизводительные распределенные вычисления на базе платформы BOINC
Интернет вещей

Технологии блокчейн:

Обзор по технологиям блокчейн
Смарт контракты
Применение блокчейн технологии в различных сферах, не связанных напрямую с финансовым сектором
Биткоин-деньги будущего?
Применение технологии блокчейн в различных сферах

4.2 Вопросы для оценки качества освоения дисциплины

Примерный перечень вопросов к экзамену

1. В чем состоит отличие между параллельной и распределенной системами?
2. Концепция параллельного программирования.
3. Какие мотивации привели к созданию распределенных систем?
4. Что характеризует масштабируемое приложение и способы достижения масштабируемости?
5. Что такое прозрачность, формы прозрачности?
6. Что такое открытая система, ее преимущества?
7. Какие концепции аппаратных решений существуют для построения распределенных систем, их особенности?
8. Какие концепции программных решений существуют для построения распределенных систем, их особенности?
9. Какие преимущества и недостатки распределенных систем?
10. Почему необходимо синхронизировать время в распределенной системе? Приведите пример.
11. Алгоритм Кристиана.
12. Алгоритм Беркли.
13. Децентрализованный алгоритм.
14. Понятие логического времени.
15. Отметки времени Лампорта.
16. Что такое глобальное состояние и алгоритм получения распределенного снимка состояния?
17. Алгоритмы голосования: алгоритм забияки и кольцевой алгоритм.
18. Алгоритмы взаимного исключения: централизованный и распределенный алгоритмы, алгоритм маркерного кольца.
19. Какие требования предъявляются программистом к современным ОС?
20. Какие стандартные API имеются в современных ОС?
21. Что такое многозадачность и какие имеются разновидности.
22. Какие механизмы существуют для обмена данными между процессами?
23. Что такое многопоточность?
24. Что такое планировщик ОС и какие имеются алгоритмы планирования? Как реализован планировщик в Windows и UNIX-системах?
25. Что такое изоляция приложений и методы ее обеспечения?
26. Что такое взаимная блокировка (dead-lock) и как ее избежать?
27. Что такое инверсия приоритетов и как ее предотвратить,
28. Какие API синхронизации имеются в Windows?
29. Какие API синхронизации имеются в UNIX?
30. Архитектура ядер ОС.
31. Для чего необходимо управление правами доступа? Какие основные цели и средства описаны в «Критериях определения безопасности компьютерных систем»?
32. В чем стоит принцип мандатного управления доступом?
33. В чем стоит принцип избирательного (дискреционного) управления доступом?
34. Какие средства сетевого взаимодействия существуют в современных ОС?
35. Перечислите этапы развития реляционных СУБД и дайте определение основным понятиям теории реляционных БД.
36. В чем заключается целостность базы данных, перечислите операции реляционной алгебры?
37. Опишите модель сервера БД (DBS).
38. Опишите модель сервера приложений (AS).
39. Опишите эволюцию серверов БД.
40. Перечислите состав задач активного сервера.

41. Приведите аспекты сетевого взаимодействия в распределенных системах.
42. Сформулируйте принципы взаимодействия «клиент-сервер».
43. Перечислите основные классы СУБД в соответствии с моделями данных.
44. Расскажите о принципах сравнения различных СУБД
45. Что такое транзакция и в чем состоит принцип ACID? Какие примитивы транзакций вы знаете? Что такое вложенные транзакции и их особенность?
46. Как реализуются распределенные транзакции? Менеджеры транзакций.
47. Для чего используется журнал транзакций. Опишите механизм отката транзакций.
48. Опишите механизм распределенных транзакций.
49. Как организован одновременный доступ к данным. Опишите механизм блокировок.
50. В чем стоит принцип двухфазной блокировки? В чем отличие реализации централизованной и распределенной двухфазной блокировки?
51. В чем отличие оптимистичной и пессимистичной блокировок?
52. Как работают мультиверсионные транзакции?
53. Особенности Больших Данных и их влияние на изменение парадигмы бизнеса
54. NoSQL – основные принципы в сравнении с реляционными базами данных.
55. Теорема CAP и ее влияние на технологии NoSQL.
56. Понятие функционального программирования.
57. Технология MapReduce
58. Принципы работы Hadoop
59. Какие компоненты составляют архитектуру CORBA?
60. Что обусловило «закат» технологии CORBA?
61. На какой технологии базируется DCOM и какие новшества она привнесла?
62. В чем похожи и чем отличаются технологии CORBA и DCOM?
63. Опишите, какие модели доступа существуют в распределенной файловой системе?
64. Какие задачи решает файловая система компьютера?
65. Опишите принципы работы сетевых файловых систем.
66. Каковы основные проблемы теории и практики распределенных систем?
67. Расскажите о основных принципах построения архитектуры Грид.
68. Что такое мобильный компьютеринг?
69. Что называют глобальным «умным» пространством?
70. Основные принципы облачных вычислений.
71. Тенденции развития современных распределенных систем: интернет вещей, block chain-технологии, smart-контракты, ...

V. РЕСУРСЫ

5.1. Основная литература

1. Крюков В.А. Курс лекций по дисциплине «Распределенные операционные системы» <http://parallel.ru/krukov/>
2. Современные операционные системы / Э. Таненбаум, Х. Бос; Пер. с англ. А. Леонтьевой. – 4-е изд. – СПб.: Питер, 2018. – 1119 с. – (Сер. "Классика computer science") . - ISBN 9785496013956: 1404.00.
3. Архитектура компьютера / Э. Таненбаум, Т. Остин; Пер. с англ. А. Матвеева. – 6-е изд. – СПб.: Питер, 2014. – 811 с. – (Сер. "Классика computer science") . - ISBN 978-5-496-00337-7.
4. Компьютерные сети / Э. Таненбаум; Пер. с англ. В. Шраги. – 4-е изд. – СПб.: Питер, 2010. – 991 с. – (Сер. "Классика computer science") . - ISBN 978-5-318-00492-6.

5. Т.1 Операционные системы / Х. М. Дейтел, П. Д. Дейтел, Д. Р. Чофнес; Пер. с англ. А. В. Бутко, и др.; Пер. с англ. под ред. С. М. Молявко. – 3-е изд. – М.: БИНОМ, 2007.
6. Т.2 : Распределенные системы, сети, безопасность / Х. М. Дейтел, П. Д. Дейтел, Д. Р. Чофнес. – 2007. – 704 с. - ISBN 5-9518015-5-9.
7. Операционные системы: параллельные и распределенные системы / Д. Бэкон, Т. Харрис; Пер. с англ. О. Здир. – СПб.: Питер; Киев: Изд. группа BHV, 2004. – 799 с. - ISBN 978-5-947239-69-0.
8. Распределенные вычислительные системы / Г. Лорин; Пер. с англ. В. А. Антропова; Под ред. Э. В. Евреинова. – М.: Радио и связь, 1984. – 294 с.
9. Компьютерные сети: принципы, технологии, протоколы: учеб. пособие для вузов / В. Г. Олифер, Н. А. Олифер. – Изд. 4-е. – СПб.: Питер, 2015. – 943 с. – (Сер. "Учебник для вузов") . - ISBN 978-5-496-00004-8.
10. Сетевые операционные системы / В. Г. Олифер, Н. А. Олифер. – 2-е изд. – СПб.: Питер, 2009. – 668 с. – (Сер. "Учебник для вузов") . - Ц. - ISBN
11. Безопасность компьютерных сетей / В. Г. Олифер, Н. А. Олифер. – М.: Горячая линия-Телеком, 2016. – 643 с. - ISBN 978-5-9912042-0-0.
12. Распределенные алгоритмы: интуитивный подход / У. Фоккинк; Пер. с англ. К. Русецкого. – СПб.: Питер, 2017. – 269 с. – (Сер. "Для профессионалов") . - ISBN 9785496021630: 1040.40.

5.2 Дополнительная литература

1. NoSQL: новая методология разработки нереляционных баз данных / П. Дж. Садаладж, М. Фаулер; Пер. с англ. и ред. Д. А. Ключина. – М.; СПб.; Киев: Вильямс, 2016. – 183 с. - ISBN 978-5-84591-920-5.
2. Storing and managing big data NoSQL, Hadoop and more: high-impact strategies- what you need to know: definitions, adoptions, impact, benefits, maturity, vendors / К. Roebuck. – Samford: Tebbo, 2011. – 228 с. – На англ. яз. - ISBN 978-1-7430-4574-9.
3. NoSQL: database for storage and retrieval of data in cloud / Ed. by G. C. DeKa. – Boca Raton [etc.]: CRC Press: Taylor & Francis Group, 2017. – 455 с. – На англ. яз. - ISBN 9781498784368.
4. Технологии программирования на Java 2: распределенные приложения / Х. М. Дейтел, П. Д. Дейтел, С. И. Сантри; Пер. с англ. под ред. А. И. Тихонова. – М.: Бином пресс, 2009. – 464 с. - Ц. - ISBN 978-5-9518027-8-9.

5.3 Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.

Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены персональными компьютерами с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.