

УТВЕРЖДЕНА

Приказом Ректора АНО ВО
«Центральный университет»
Ивашкевич Е.В.
от «19» января 2024 г. № 0119.37

**Рабочая программа дисциплины (модуля)
«Распределенные системы»
дополнительной профессиональной программы – программы
профессиональной переподготовки «Академия data science»**

Траектория: Машинное обучение

**Москва
2024**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Тематический план	4
3. Содержание дисциплины (модуля)	4
4. Учебно-методическое обеспечение	5
5. Материально-техническое обеспечение	5
6. Методические и оценочные материалы	7

1. Краткая характеристика дисциплины (модуля)

Изучение дисциплины (модуля) «Распределенные системы» позволяет слушателям освоить современные технологии, обеспечивающие высокую доступность и отказоустойчивость приложений, что критически важно в эпоху больших данных и облачных вычислений. Кроме того, знание принципов работы распределенных систем открывает возможности для создания инновационных решений в различных областях, от финансов до здравоохранения.

Цель изучения дисциплины (модуля): формирование у слушателей понимания принципов проектирования, разработки и управления распределенными вычислительными системами для эффективного решения сложных задач.

Задачи изучения дисциплины (модуля):

- формирование знаний основных понятий и проблем теории распределенных систем, модели согласованности;
- формирование знаний алгоритмов репликации и консенсуса в распределенных системах;
- формирование знаний понятия транзакции, уровни изоляции транзакций и способы их реализации;
- формирование знаний о видах отказов узлов, понятия византийских отказов;
- формирование умения реализовывать на любом языке программирования классические алгоритмы репликации и консенсуса в распределенных системах;
- формирование умения выбирать готовые реализации распределенных систем в зависимости от необходимых требований к согласованности и устойчивости к отказам;
- формирование умения использовать MapReduce и распределенные файловые системы для работы с большими данными;
- формирование навыка проектировать архитектуру распределенной системы с учетом требований к согласованности данных и устойчивости к отказам;
- формирование навыка отличать достижимые требования от недостижимых при проектировании распределенных систем.

2. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		<i>Очная форма</i>				
		Аудиторная работа		Контроль	Самостоя тельная работа	
Лекции	Семинары (практичес кие занятия)					
1	Введение. Основные понятия и проблемы теории распределенных систем	5	5		15	Домашнее задание Подготовка к семинару
2	Репликация и секционирование	5	5		15	Домашнее задание Подготовка к семинару
3	Консенсус в распределенных системах	5	6		16	Домашнее задание Тест
4	Транзакции	6	6		16	Домашнее задание Подготовка к семинару
5	Потоковая и пакетная обработка данных	6	6		16	Домашнее задание Тест
6	Византийские отказы	6	6		16	Домашнее задание Подготовка к семинару
	<i>Зачет</i>			4		
	Итого:	33	34	4	94	
Объем дисциплины (модуля) (в ак. ч.)		165				

3. Содержание дисциплины (модуля)

№ п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Введение. Основные понятия и проблемы теории распределенных систем	Модели согласованности. Синхронные и асинхронные модели. Отказы узлов, виды отказов. Проблема синхронизации часов. Алгоритмы синхронизации часов
2	Репликация и секционирование	Алгоритмы репликации: с одним ведущим узлом, с несколькими ведущими узлами, без ведущих узлов. Проблемы задержки репликации. Кворумы, записи и чтения по кворуму. Применение алгоритмов репликации в реальных системах
3	Консенсус в распределенных системах	Линеаризуемые системы. Линеаризуемость и причинная согласованность. Atomic Broadcast. Понятие консенсуса. Теорема FLP о невозможности консенсуса. Алгоритмы Paxos и RAFT. Достижение консенсуса в реальных системах
4	Транзакции	Понятие транзакции. Уровни изоляции транзакций. Two-phase locking и Snapshot Isolation. Распределенные транзакции. Релизация транзакций в реальных распределенных системах
5	Потоковая и пакетная обработка данных	Идея MapReduce. Распределенные вычисления в MapReduce. Реализация и применение MapReduce-систем. Распределенные файловые системы. Система Hadoop. Очереди сообщений. Брокеры сообщений.
6	Византийские отказы	Задача византийских генералов. Алгоритмы работы с византийскими отказами. Примеры реальных систем, переживающих византийские отказы

4. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый слушатель в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Слушателям обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Бабичев, С. Л. Распределенные системы : учебник для вузов / С. Л. Бабичев, К. А. Коньков. — Москва : Издательство Юрайт, 2025. — 507 с. — (Высшее образование). — ISBN 978-5-534-11380-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/566315>.

Дополнительная литература:

1. Петров Алекс Ю. Сергиенко К. Тульцева А. Коцюба А. Руденко В. Мостипан Н. Петрова, М. Одинокова Е. Неволainen Распределенные данные. Алгоритмы работы современных систем хранения информации. — СПб.: Питер, 2021. — 336 с.: ил. — (Серия «Бестселлеры O'Reilly»). ISBN 978-5-4461-1640-9.

5. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Распределенные системы:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое

Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

6. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Распределенные системы» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, тесты, домашние задания, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Участие в семинаре (аудиторная работа) – активная работа слушателя на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре слушателям рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Тест – особая форма проверки знаний. Проводится после освоения одной или нескольких тем и свидетельствует о качестве понимания основных понятий изучаемого материала. Тестовые задания составлены к ключевым понятиям, основным разделам, важным терминологическим категориям изучаемой дисциплины (модуля).

Для подготовки к тесту необходимо знать терминологический аппарат дисциплины (модуля), понимать смысл научных категорий и уметь их использовать в профессиональной лексике. Владение понятийным аппаратом, включённым в тестовые задания, позволяет преподавателю быстро проверить уровень понимания слушателями важных методологических категорий.

Самостоятельная работа – работа слушателей, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы слушатели взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи слушателя включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *зачета*.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	Слушатель полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Слушатель хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	Зачтено	
8	Отлично	Зачтено	
7	Хорошо	Зачтено	Слушатель обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать
6	Хорошо	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Слушатель хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Зачтено	Слушатель обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Слушатель способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	
3	Не сдан	Не зачтено	Слушатель не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Распределенные системы» оценивается следующим образом:

Активность	Вес	Описание
Домашние задания	50%	За каждое из заданий можно набрать 10 баллов
Аудиторная работа	10%	На каждом семинаре слушатель может заработать баллы за интересные вопросы, работу на семинаре и выполнение заданий
Тесты	10%	Ответы на вопросы по изученным темам
Зачет	30%	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Распределенные системы»: « $0,5 \times$ среднее за домашние задания + $0,1 \times$ аудиторная работа + $0,1 \times$ среднее за тесты + $0,3 \times$ зачет».

Текущий контроль успеваемости обучающихся по дисциплине (модулю) Примерные вопросы для подготовки к семинарам

Введение. Основные понятия и проблемы теории распределенных систем

1. Основные понятия распределенных систем: определения и характеристики.

2. Проблемы теории распределенных систем: надежность, доступность и производительность.
3. Модели согласованности: строгая, слабая и причинная согласованность.
4. Синхронные и асинхронные модели взаимодействия в распределенных системах.
5. Виды отказов узлов: временные, постоянные и сетевые.
6. Проблема синхронизации часов в распределенных системах.
7. Алгоритмы синхронизации часов: NTP и другие подходы.
8. Примеры реальных проблем, возникающих из-за недостаточной синхронизации.

Репликация и секционирование

9. Алгоритмы репликации с одним ведущим узлом: принципы и применение.
10. Алгоритмы репликации с несколькими ведущими узлами: преимущества и недостатки.
11. Репликация без ведущих узлов: концепции и реализация.
12. Проблемы задержки репликации: причины и решения.
13. Кворумы в распределенных системах: определение и применение.
14. Записи и чтения по кворуму: алгоритмы и примеры.
15. Применение алгоритмов репликации в реальных системах: кейсы и анализ.

Консенсус в распределенных системах

16. Линеаризуемые системы: понятие и важность.
17. Линеаризуемость и причинная согласованность: различия и примеры.
18. Atomic Broadcast: определение и алгоритмы.
19. Понятие консенсуса: значение в распределенных системах.
20. Теорема FLP о невозможности консенсуса: объяснение и последствия.
21. Алгоритм Paxos: принципы работы и реализация.
22. Алгоритм RAFT: сравнение с Paxos и примеры использования.
23. Достижение консенсуса в реальных системах: практические примеры.

Транзакции

24. Понятие транзакции: основные характеристики и свойства.
25. Уровни изоляции транзакций: определение и примеры.
26. Two-phase locking: принципы и применение.
27. Snapshot Isolation: концепция и преимущества.
28. Распределенные транзакции: особенности и сложности.
29. Реализация транзакций в реальных распределенных системах: примеры и анализ.

Потоковая и пакетная обработка данных

30. Идея MapReduce: основные принципы и применение.
31. Распределенные вычисления в MapReduce: алгоритмы и реализация.
32. Распределенные файловые системы: концепции и примеры.
33. Система Hadoop: архитектура и применение.
34. Очереди сообщений: принципы работы и применение в распределенных системах.
35. Брокеры сообщений: реализация и использование в реальных системах.

Примерные домашние задания

Домашнее задание: Консенсус в распределенных системах

1. **Определите линеаризуемость.** Приведите примеры, где линеаризуемость критически важна для работы системы.
2. **Сравните причинную согласованность и линеаризуемость.** Какие преимущества и недостатки имеет каждая из моделей?

3. **Объясните принцип работы Atomic Broadcast.** Как он помогает достичь согласованности в распределенных системах?

4. **Изучите теорему FLP о невозможности консенсуса.** Каковы ее основные выводы и как они влияют на проектирование распределенных систем?

1. **Опишите алгоритмы Paxos и RAFT.** В чем основные различия между ними и в каких случаях каждый из них предпочтителен?

Домашнее задание: Транзакции

1. **Определите понятие транзакции.** Какие основные свойства транзакций (ACID) существуют и почему они важны?

2. **Объясните уровни изоляции транзакций.** Как они влияют на производительность и согласованность данных?

3. **Сравните Two-phase locking и Snapshot Isolation.** В каких ситуациях предпочтительнее использовать каждый из подходов?

4. **Рассмотрите особенности распределенных транзакций.** Какие проблемы могут возникнуть при их реализации и как их можно решить?

5. **Проанализируйте реальную распределенную систему,** использующую транзакции. Опишите, как она реализует управление транзакциями и какие алгоритмы использует.

Домашнее задание: Поточковая и пакетная обработка данных

1. **Объясните идею MapReduce.** Как она позволяет обрабатывать большие объемы данных?

2. **Опишите распределенные вычисления в MapReduce.** Как происходит распределение задач и сбор результатов?

3. **Изучите реализацию и применение MapReduce-систем.** Приведите примеры реальных систем, использующих эту технологию.

4. **Опишите архитектуру распределенных файловых систем.** Какие ключевые компоненты обеспечивают их функциональность?

5. **Рассмотрите систему Hadoop.** Какие ее основные компоненты и как они взаимодействуют для обработки данных?

Примерные задания для тестов

Тест 1: Консенсус в распределенных системах

1. **Что такое линейризуемость?**

- A) Модель, при которой все операции выполняются одновременно
- B) Модель, при которой операции могут выполняться в произвольном порядке
- C) Модель, при которой операции выполняются последовательно в некотором порядке
- D) Модель, при которой операции могут быть отменены

2. **Какова основная идея причинной согласованности?**

- A) Все операции должны быть выполнены в одном порядке
- B) Операции, которые зависят друг от друга, должны быть выполнены в порядке их возникновения
- C) Все операции могут быть выполнены одновременно
- D) Операции могут быть выполнены в произвольном порядке

3. **Что такое Atomic Broadcast?**

- A) Механизм, который гарантирует, что сообщения будут доставлены в том порядке, в котором они были отправлены
- B) Механизм, который гарантирует, что сообщения будут доставлены всем узлам
- C) Механизм, который гарантирует, что сообщения не будут потеряны

- D) Все вышеперечисленное
4. **Какое из следующих утверждений о консенсусе является верным?**
- A) Консенсус всегда достигается в одно и то же время
 - B) Консенсус может быть достигнут только с помощью одного узла
 - C) Консенсус — это процесс, в котором все участники должны согласовать одно значение
 - D) Консенсус не имеет значения в распределенных системах
5. **Что утверждает теорема FLP?**
- A) Консенсус возможен в любой распределенной системе
 - B) Консенсус невозможен в системах с ненадежными узлами
 - C) Консенсус всегда достигается в конечное время
 - D) Консенсус возможен только при наличии одного узла
6. **Какой алгоритм используется для достижения консенсуса в распределенных системах?**
- A) MapReduce
 - B) Paxos
 - C) Hadoop
 - D) FIFO
7. **Какой из следующих алгоритмов является более простым для понимания и реализации?**
- A) Paxos
 - B) RAFT
 - C) FLP
 - D) Atomic Broadcast
8. **Какой из следующих факторов влияет на достижение консенсуса в реальных системах?**
- A) Количество узлов
 - B) Сетевые задержки
 - C) Отказы узлов
 - D) Все вышеперечисленное
9. **Что определяет линеаризуемость в системе?**
- A) Согласованность данных
 - B) Порядок операций
 - C) Доступность системы
 - D) Производительность системы
10. **Какой из следующих алгоритмов более устойчив к сбоям?**
- A) Paxos
 - B) RAFT
 - C) Оба алгоритма одинаково устойчивы
 - D) Ни один из алгоритмов не устойчив к сбоям
11. **Какова основная цель алгоритма Paxos?**
- A) Обеспечить высокую доступность
 - B) Достичь консенсуса среди узлов
 - C) Увеличить производительность
 - D) Минимизировать задержки
12. **Что такое "кворум" в контексте консенсуса?**
- A) Минимальное количество узлов, необходимое для принятия решения
 - B) Максимальное количество узлов, которые могут участвовать в голосовании
 - C) Система, которая не требует согласия
 - D) Узел, который принимает все решения
13. **Какой из следующих факторов не влияет на консенсус?**
- A) Степень надежности сети
 - B) Алгоритм, используемый для достижения консенсуса

- C) Количество операций, выполняемых системой
 - D) Количество узлов в системе
14. **Что происходит, если узел выходит из строя во время достижения консенсуса?**
- A) Система немедленно прекращает работу
 - B) Процесс консенсуса может продолжаться, если другие узлы могут достичь соглашения
 - C) Все узлы должны быть перезапущены
 - D) Система автоматически выбирает новый узел
15. **Какой из следующих примеров является реализацией консенсуса?**
- A) Система управления базами данных
 - B) Распределенная файловая система
 - C) Система блокчейн
 - D) Все вышеперечисленное

Тест 2: Поточковая и пакетная обработка данных

1. **В чем заключается основная идея MapReduce?**
 - A) Параллельная обработка данных
 - B) Обработка данных в реальном времени
 - C) Синхронная обработка данных
 - D) Обработка данных на одном узле
2. **Что такое "Map" в контексте MapReduce?**
 - A) Процесс сбора результатов
 - B) Процесс распределения задач
 - C) Процесс преобразования входных данных в промежуточные данные
 - D) Процесс хранения данных
3. **Что делает функция "Reduce" в MapReduce?**
 - A) Объединяет промежуточные данные
 - B) Делает выборку данных
 - C) Удаляет дубликаты
 - D) Выполняет сортировку данных
4. **Какой из следующих компонентов не является частью архитектуры MapReduce?**
 - A) Master node
 - B) Worker node
 - C) Client node
 - D) Network node
5. **Каковы основные преимущества использования распределенных вычислений в MapReduce?**
 - A) Увеличение производительности и масштабируемости
 - B) Упрощение кода
 - C) Снижение затрат на оборудование
 - D) Все вышеперечисленное
6. **Что такое распределенные файловые системы?**
 - A) Системы, которые хранят данные на одном узле
 - B) Системы, которые хранят данные на нескольких узлах
 - C) Системы, которые используют облачное хранилище
 - D) Системы, которые не требуют резервного копирования
7. **Какой из следующих компонентов является частью системы Hadoop?**
 - A) Hadoop Distributed File System (HDFS)
 - B) Hadoop MapReduce
 - C) Hadoop YARN
 - D) Все вышеперечисленное

8. **Какова основная функция HDFS?**
- A) Обработка данных
 - B) Хранение больших объемов данных
 - C) Управление вычислительными ресурсами
 - D) Обеспечение безопасности данных
9. **Что такое очереди сообщений?**
- A) Механизм для хранения данных
 - B) Механизм для асинхронной передачи сообщений между компонентами
 - C) Механизм для синхронной передачи данных
 - D) Механизм для обработки данных в реальном времени
10. **Какой из следующих брокеров сообщений является наиболее популярным?**
- A) RabbitMQ
 - B) HDFS
 - C) MapReduce
 - D) Apache Spark
11. **Какой из следующих аспектов не является преимуществом использования брокеров сообщений?**
- A) Упрощение интеграции между системами
 - B) Увеличение задержек передачи данных
 - C) Обеспечение надежности передачи сообщений
 - D) Поддержка асинхронной обработки
12. **Какой из следующих методов используется для реализации очередей сообщений?**
- A) FIFO (First In, First Out)
 - B) LIFO (Last In, First Out)
 - C) Random Access
 - D) Circular Buffer
13. **Какой из следующих сценариев является примером использования очередей сообщений?**
- A) Синхронное взаимодействие между клиентом и сервером
 - B) Асинхронная обработка платежей в интернет-магазине
 - C) Хранение данных в базе данных
 - D) Обработка данных в реальном времени
14. **Какой из следующих компонентов системы Hadoop отвечает за управление ресурсами?**
- A) HDFS
 - B) MapReduce
 - C) YARN
 - D) Hive
15. **Какую роль играют брокеры сообщений в распределенных системах?**
- A) Обеспечивают синхронность
 - B) Упрощают взаимодействие между компонентами
 - C) Увеличивают сложность архитектуры
 - D) Уменьшают производительность