

Документы
Информация о владельце: "Самарский государственный экономический университет"
ФИО: Кандрашина Елена Александровна
Должность: И.о. ректора ФГАОУ ВО «Самарский государственный экономический университет»
Дата подписания: 04.08.2026 16:19:30
Уникальный программный ключ:
2db64eb9605ce27edd3b8e8fdd32c70e0674ddd2

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
«СИСТЕМЫ ОБРАБОТКИ И АНАЛИЗА БОЛЬШИХ МАССИВОВ ДАННЫХ»**

Уровень высшего образования: магистратура

Направление подготовки: 09.04.03 Прикладная информатика

Направленность (профиль) подготовки: Искусственный интеллект и большие данные

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Год набора (приема на обучение): 2026

Срок получения образования: 2 года

Объем: в зачетных единицах: 6 з.е.
в академических часах: 216 ак.ч.

г. Самара, 2026

Разработчики:

Не имеет Колотилина М. А.

Рабочая программа дисциплины (модуля) составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.04.03 Прикладная информатика, утвержденного приказом Минобрнауки от 19.09.2017 № 916, с учетом трудовых функций профессиональных стандартов: "Руководитель проектов в области информационных технологий", утвержден приказом Минтруда России от 27.04.2023 № 369н.

Согласование и утверждение

№	Подразделение или коллегиальный орган	Ответственное лицо	ФИО	Виза	Дата, протокол (при наличии)
1	Кафедра прикладной информатики	Заведующий кафедрой, руководитель подразделения, реализующего ОП	Ермолина Л. В.	Рассмотрено	21.05.2026, № 9

1. Цель и задачи освоения дисциплины (модуля)

Цель освоения дисциплины - Сформировать у магистров теоретические знания и практические навыки применения современных методов, моделей и распределённых систем для эффективного хранения, обработки и анализа больших массивов данных в различных предметных областях

Задачи изучения дисциплины:

- Изучить архитектуры и технологии распределённой обработки данных (Hadoop, Spark) и модели хранения NoSQL для работы с большими объёмами данных;
- Научиться применять инженерные практики: строить пайплайны обработки (ETL/ELT), использовать потоковую обработку (Streaming) и оркестрацию процессов (Airflow);
- Освоить навыки масштабируемой аналитики, включая построение аналитических витрин (OLAP, Data Warehousing) и применение распределённых библиотек машинного обучения (MLlib) .

2. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы

Компетенции, индикаторы и результаты обучения

ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

ОПК-3.1 Анализирует профессиональную информацию, выделяет в ней главное и определяет ее структуру

Знать:

ОПК-3.1/Зн1 Методы и приёмы анализа профессиональной информации, критерии выделения главного в больших массивах данных и текстовых источниках; принципы структурирования информации для последующего представления в виде аналитических обзоров; методики систематизации данных и построения логических схем для формирования обоснованных выводов.

ОПК-3.1/Зн2 Методологию структурирования профессиональной информации и её оформления в виде аналитических обзоров, научных отчетов и служебных записок; требования к визуализации данных и подготовке презентационных материалов; техники аргументации и формулирования обоснованных выводов и рекомендаций; принципы адаптации стиля и содержания обзора под уровень подготовки и запросы целевой группы; основы публичной защиты аналитических материалов.

Уметь:

ОПК-3.1/Ум1 Проводить анализ профессиональной информации из различных источников (научные статьи, техническая документация, отчёты); выделять главное, отсеивая второстепенное, и определять структуру информации для дальнейшего использования; систематизировать и группировать данные по ключевым тематическим блокам; формулировать предварительные выводы на основе структурированной информации.

ОПК-3.1/Ум2 Структурировать профессиональную информацию и оформлять её в формате аналитических обзоров, отчётов и служебных записок; применять методы визуализации данных (графики, диаграммы, инфографика) и подготавливать презентационные материалы для различных аудиторий; формулировать обоснованные выводы и рекомендации, аргументируя их фактами и расчётами; адаптировать стиль и содержание обзора под уровень подготовки и запросы целевой группы; публично защищать аналитические материалы перед аудиторией, отвечая на вопросы и корректируя аргументацию по запросу слушателей.

Владеть:

ОПК-3.1/Нв1 Навыками анализа профессиональной информации и выделения главного при работе с техническими и научными текстами; методами структурирования информации (построение ментальных карт, таблиц, логических схем, аннотирование); техниками подготовки материалов для последующего оформления в виде аналитических обзоров, отчетов и презентаций; способностью обосновывать значимость выделенной информации для решения профессиональных задач.

ОПК-3.1/Нв2 Навыками структурирования профессиональной информации и её оформления в виде аналитических обзоров, научных отчетов и служебных записок; методами визуализации данных и подготовки презентационных материалов; техниками представления результатов анализа с аргументацией выводов и обоснованием рекомендаций; навыками публичных выступлений и защиты своих аналитических материалов перед аудиторией; способностью адаптировать стиль и содержание обзора под уровень подготовки и запросы целевой группы.

ОПК-3.2 Структурирует, оформляет и представляет информацию в виде аналитических обзоров с обоснованными выводами и рекомендациями

Знать:

ОПК-3.2/Зн1 Требования к структурированию, оформлению и представлению профессиональной информации в виде аналитических обзоров; стандарты и форматы оформления аналитических документов, отчетов и презентаций (ГОСТ, корпоративные стандарты); методологию формирования обоснованных выводов и рекомендаций на основе анализа данных; способы визуализации и аргументации информации для различных целевых аудиторий (заказчики, руководство, коллеги).

Уметь:

ОПК-3.2/Ум1 Структурировать информацию (выделять разделы, подразделы, логические блоки) для построения четкой и аргументированной структуры аналитического обзора; оформлять аналитические обзоры, отчеты и презентации в соответствии с требованиями и стандартами; представлять результаты анализа в наглядной форме (графики, таблицы, диаграммы, инфографика) для различных стейкхолдеров; формулировать обоснованные выводы и рекомендации на основе систематизации данных, подтверждая их фактами и расчетами.

Владеть:

ОПК-3.2/Нв1 Навыками структурирования профессиональной информации и её оформления в виде аналитических обзоров, научных отчетов и служебных записок; методами визуализации данных и подготовки презентационных материалов; техниками представления результатов анализа с аргументацией выводов и обоснованием рекомендаций; навыками публичных выступлений и защиты своих аналитических материалов перед аудиторией; способностью адаптировать стиль и содержание обзора под уровень подготовки и запросы целевой группы.

ОПК-5 Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем

ОПК-5.1 Изучает и модернизирует программное и аппаратное обеспечение информационных и автоматизированных систем для решения профессиональных задач

Знать:

ОПК-5.1/Зн1 Принципы организации, архитектуру и компоненты программного и аппаратного обеспечения информационных и автоматизированных систем; методы и подходы к изучению существующих систем для выявления узких мест и потенциала модернизации; современные технологии и платформы для разработки и модернизации ПО и аппаратных средств; требования к производительности, надёжности, масштабируемости и безопасности при модернизации систем для решения профессиональных задач.

Уметь:

ОПК-5.1/Ум1 Проводить изучение текущего состояния программного и аппаратного обеспечения информационных и автоматизированных систем; выявлять недостатки, устаревшие компоненты и зоны для улучшения; осуществлять модернизацию программных модулей, библиотек и аппаратных компонентов с учётом поставленных профессиональных задач; оценивать влияние изменений на общую производительность и совместимость системы; разрабатывать техническую документацию для обоснования модернизации.

Владеть:

ОПК-5.1/Нв1 Навыками изучения и анализа программного и аппаратного обеспечения информационных и автоматизированных систем; методами и инструментарием модернизации ПО (рефакторинг, обновление версий, интеграция новых библиотек) и аппаратных средств (апгрейд серверов, сетевое оборудование); навыками тестирования и валидации модернизированных решений; способностью документировать изменения и обосновывать их необходимость для решения профессиональных задач; техниками оценки эффективности модернизации по критериям производительности, стоимости и надёжности.

ОПК-5.2 Разрабатывает и выбирает программное и аппаратное обеспечение информационных и автоматизированных систем для эффективного решения профессиональных задач

Знать:

ОПК-5.2/Зн1 Современные подходы к разработке и выбору программного и аппаратного обеспечения информационных и автоматизированных систем; критерии оценки эффективности ПО и аппаратных средств для решения профессиональных задач (производительность, масштабируемость, стоимость владения, совместимость, безопасность); архитектурные паттерны и платформы для построения распределённых систем обработки данных; методы сравнительного анализа альтернативных решений при выборе технологического стека.

Уметь:

ОПК-5.2/Ум1 Разрабатывать и обосновывать выбор программного и аппаратного обеспечения под конкретные профессиональные задачи (обработка больших данных, визуализация, машинное обучение); проводить сравнительный анализ и оценку эффективности различных программных продуктов и аппаратных платформ; адаптировать выбранные решения под ограничения проекта (бюджет, сроки, компетенции команды); проектировать архитектуру систем с учётом возможности масштабирования и модернизации; аргументировать выбор технологий перед заказчиком и руководством.

Владеть:

ОПК-5.2/Нв1 Навыками разработки и выбора программного и аппаратного обеспечения для информационных и автоматизированных систем; методами оценки и сравнения альтернативных решений (TCO, ROI, бенчмарки); инструментами моделирования и прототипирования архитектурных решений; навыками обоснования выбора технологического стека с учётом требований к производительности, надёжности и стоимости; способностью адаптировать выбранные решения под изменения требований и условий проекта для эффективного решения профессиональных задач.

ОПК-7 Способен использовать методы научных исследований и математического моделирования в области проектирования и управления информационными системами

ОПК-7.2 Применяет в практике создания информационных систем современные методы научных исследований и математического моделирования

Знать:

ОПК-7.2/Зн1 Современные методы научных исследований и подходы к математическому моделированию, применяемые при проектировании и управлении информационными системами; классификацию моделей (аналитические, имитационные, статистические, оптимизационные) и их применимость для решения задач анализа данных и управления; методологию проведения экспериментов, сбора и обработки данных для верификации моделей; требования к точности, адекватности и интерпретируемости математических моделей в практике создания ИС.

Уметь:

ОПК-7.2/Ум1 Применять в практике создания информационных систем современные методы научных исследований (сравнительный анализ, эксперимент, проверка гипотез) для обоснования проектных решений; использовать математическое моделирование для прогнозирования поведения системы, оценки нагрузки, оптимизации ресурсов и управления рисками; выбирать адекватные модели и методы анализа данных для конкретных профессиональных задач (кластеризация, регрессия, оптимизация); интерпретировать результаты моделирования и формулировать обоснованные рекомендации для проектирования и управления ИС.

Владеть:

ОПК-7.2/Нв1 Навыками применения современных методов научных исследований и математического моделирования в практике создания информационных систем; инструментами и библиотеками для моделирования и анализа данных (Python, R, MATLAB, Simulink, GPSS); методами верификации и валидации моделей; навыками документирования и представления результатов моделирования для обоснования управленческих и проектных решений; способностью адаптировать методы моделирования под ограничения проектов малого и среднего уровня сложности.

3. Место дисциплины в структуре ОП

Дисциплина (модуль) «Системы обработки и анализа больших массивов данных» относится к обязательной части образовательной программы и изучается в семестре(ах): 1.

В процессе изучения дисциплины студент готовится к решению типов задач профессиональной деятельности, предусмотренных ФГОС ВО и образовательной программой.

Компетенция	Предшествующие дисциплины	Последующие дисциплины
ОПК-3 - Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями		

ОПК-3.1 Анализирует профессиональную информацию, выделяет в ней главное и определяет ее структуру	Иностранный язык в профессиональной деятельности (английский), Учебная практика: технологическая (проектно-технологическая) практика	Выполнение и защита выпускной квалификационной работы, Иностранный язык в профессиональной деятельности (английский), Производственная практика: научно-исследовательская работа, Учебная практика: технологическая (проектно-технологическая) практика
ОПК-3.2 Структурирует, оформляет и представляет информацию в виде аналитических обзоров с обоснованными выводами и рекомендациями	Иностранный язык в профессиональной деятельности (английский), Учебная практика: технологическая (проектно-технологическая) практика	Выполнение и защита выпускной квалификационной работы, Иностранный язык в профессиональной деятельности (английский), Производственная практика: научно-исследовательская работа, Учебная практика: технологическая (проектно-технологическая) практика
ОПК-5 - Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем		
ОПК-5.1 Изучает и модернизирует программное и аппаратное обеспечение информационных и автоматизированных систем для решения профессиональных задач		Выполнение и защита выпускной квалификационной работы, Прикладные программные продукты в профессиональной деятельности
ОПК-5.2 Разрабатывает и выбирает программное и аппаратное обеспечение информационных и автоматизированных систем для эффективного решения профессиональных задач		Выполнение и защита выпускной квалификационной работы, Прикладные программные продукты в профессиональной деятельности
ОПК-7 - Способен использовать методы научных исследований и математического моделирования в области проектирования и управления информационными системами		
ОПК-7.2 Применяет в практике создания информационных систем современные методы научных исследований и математического моделирования		Выполнение и защита выпускной квалификационной работы, Управление проектной деятельностью в профессиональной сфере

4. Объем дисциплины (модуля) и виды учебной работы

Период обучения	Общая трудоемкость (часы)	Общая трудоемкость (ЗЕТ)	Контактная работа (часы, всего)	Лекционные занятия (часы)	Практические занятия (часы)	Групповая контактная работа (часы)	Индивидуальная контактная работа (часы)	Самостоятельная работа (часы)	Промежуточная аттестация
Первый семестр	216	6	12	4	8	2	0,3	167,7	Экзамен
Всего	216	6	12	4	8	2	0,3	167,7	34

5. Содержание дисциплины (модуля)

5.1. Разделы, темы дисциплины и виды занятий

(часы промежуточной аттестации не указываются)

Наименование раздела, темы	Всего	Лекционные занятия	Практические занятия	Самостоятельная работа
Раздел 1. Системы обработки и анализа больших массивов данных	182	4	8	167,7
Тема 1.1. Введение в Big Data: вызовы и парадигмы Определение Big Data и модель 5V (Volume, Velocity, Variety, Veracity, Value). Сравнение традиционных (OLTP) и распределённых (OLAP) систем. Обзор типов данных (структурированные, полуструктурированные, неструктурированные) и форматов хранения	42		2	40
Тема 1.2. Распределённые файловые системы: HDFS Архитектура HDFS (NameNode, DataNode). Схема репликации, отказоустойчивость. Операции чтения и записи. Интерфейсы доступа (Web, shell, Java API). Сравнение с другими DFS	46,3	2	2	40

Тема 1.3. Парадигма MapReduce и экосистема Hadoop Модель MapReduce: этапы Map, Shuffle, Reduce. Реализация в Hadoop (JobTracker, TaskTracker / YARN). Типы Join'ов в MR (Map-side, Reduce-side) и паттерны проектирования (pairs, stripes, составные ключи). Оптимизация задач и управление ресурсами	44	2	2	40
Тема 1.4. SQL над большими данными: Hive Трансляция SQL-запросов в MapReduce-задачи. Виды таблиц в Hive (Managed, External), типы данных. Аналитические функции и оптимизация (партиционирование, бакетирование). Пользовательские функции (UDF)	49,7		2	47,7

5.2. Контрольные мероприятия по дисциплине

Вид контроля	Форма контроля/Оценочное средство
Текущий контроль	Тестовое задание
Промежуточная аттестация	Экзамен

№ п/п	Наименование раздела	Вид контроля/ используемые оценочные материалы	
		Текущий	Промежут. аттестация
1	Системы обработки и анализа больших массивов данных	Тестовое задание	Экзамен

6. Оценочные материалы текущего контроля

1. Системы обработки и анализа больших массивов данных Тестовое задание

№ п/п	Содержание вопроса		Компетенция
		Правильный ответ (ключ ответа)	
1	В рамках анализа профессиональной информации о системах хранения больших данных выделите главное – какую модель хранения целесообразно применять для данных с высокой вариативностью структуры (схема не фиксирована)? а) Реляционные СУБД (OLTP) б) Документно-ориентированные NoSQL (MongoDB) в) Колоночные СУБД (OLAP) г) Файловые системы		ОПК-3
	Ответ:	б	
2	При структурировании информации о подходах к потоковой обработке данных выделите главное – какой фреймворк обеспечивает обработку событий в реальном времени с гарантией exactly-once? а) Apache Spark Streaming б) Apache Flink в) Apache Hadoop MapReduce г) Apache Hive		ОПК-3
	Ответ:	б	

3	В рамках анализа профессиональной информации о распределённых вычислительных моделях определите главное – какая модель позволяет выполнять итеративные алгоритмы машинного обучения наиболее эффективно за счёт кэширования данных в памяти? a) MapReduce b) Dask (Distributed) c) Apache Spark d) Hadoop Streaming	ОПК-3								
	<table><tr><td>Ответ:</td><td>c</td></tr></table>	Ответ:	c							
Ответ:	c									
4	При оформлении аналитического обзора по технологиям ETL выделите главное – какой инструмент оркестрации используется для построения сложных пайплайнов с зависимостями между задачами? a) Apache Airflow b) Apache Kafka c) Apache ZooKeeper d) PostgreSQL	ОПК-3								
	<table><tr><td>Ответ:</td><td>a</td></tr></table>	Ответ:	a							
Ответ:	a									
5	В рамках структурирования информации о методах обработки больших данных выберите главное – какая система используется для построения аналитических витрин (OLAP-кубов) и выполнения сложных агрегационных запросов на больших объёмах? a) ClickHouse b) Redis c) Elasticsearch d) Cassandra	ОПК-3								
	<table><tr><td>Ответ:</td><td>a</td></tr></table>	Ответ:	a							
Ответ:	a									
6	Соотнесите тип системы хранения с её характерным применением при анализе профессиональной информации о Big Data: <table><tr><th>Тип системы</th><th>Применение</th></tr><tr><td>1. Реляционная СУБД (PostgreSQL)</td><td>А. Хранение сырых событий и логов с высокой пропускной способностью записи</td></tr><tr><td>2. Документная NoSQL (MongoDB)</td><td>Б. Транзакционные системы с высокой целостностью данных (OLTP)</td></tr><tr><td>3. Колоночная СУБД (ClickHouse)</td><td>В. Аналитические запросы с агрегацией по нескольким полям (OLAP)</td></tr></table>	Тип системы	Применение	1. Реляционная СУБД (PostgreSQL)	А. Хранение сырых событий и логов с высокой пропускной способностью записи	2. Документная NoSQL (MongoDB)	Б. Транзакционные системы с высокой целостностью данных (OLTP)	3. Колоночная СУБД (ClickHouse)	В. Аналитические запросы с агрегацией по нескольким полям (OLAP)	ОПК-3
Тип системы	Применение									
1. Реляционная СУБД (PostgreSQL)	А. Хранение сырых событий и логов с высокой пропускной способностью записи									
2. Документная NoSQL (MongoDB)	Б. Транзакционные системы с высокой целостностью данных (OLTP)									
3. Колоночная СУБД (ClickHouse)	В. Аналитические запросы с агрегацией по нескольким полям (OLAP)									
	<table><tr><td>Ответ:</td><td>1–Б, 2–А, 3–В</td></tr></table>	Ответ:	1–Б, 2–А, 3–В							
Ответ:	1–Б, 2–А, 3–В									
7	Соотнесите этап пайплайна обработки данных с его описанием для структурирования информации о Big Data: <table><tr><th>Этап пайплайна</th><th>Описание</th></tr><tr><td>1. Ingestion (Сбор)</td><td>А. Преобразование данных в формат для аналитики (очистка, обогащение)</td></tr><tr><td>2. Processing (Обработка)</td><td>Б. Выгрузка подготовленных данных в хранилище или витрину</td></tr><tr><td>3. Serving (Представление)</td><td>В. Приём данных из внешних источников (Kafka, API, файлы)</td></tr></table>	Этап пайплайна	Описание	1. Ingestion (Сбор)	А. Преобразование данных в формат для аналитики (очистка, обогащение)	2. Processing (Обработка)	Б. Выгрузка подготовленных данных в хранилище или витрину	3. Serving (Представление)	В. Приём данных из внешних источников (Kafka , API, файлы)	ОПК-3
Этап пайплайна	Описание									
1. Ingestion (Сбор)	А. Преобразование данных в формат для аналитики (очистка, обогащение)									
2. Processing (Обработка)	Б. Выгрузка подготовленных данных в хранилище или витрину									
3. Serving (Представление)	В. Приём данных из внешних источников (Kafka , API, файлы)									
	<table><tr><td>Ответ:</td><td>1–В, 2–А, 3–Б</td></tr></table>	Ответ:	1–В, 2–А, 3–Б							
Ответ:	1–В, 2–А, 3–Б									
8	Соотнесите инструмент с его ролью в экосистеме Big Data для оформления аналитического обзора: <table><tr><th>Этап пайплайна</th><th>Описание</th></tr><tr><td>1. Ingestion (Сбор)</td><td>А. Преобразование данных в формат для аналитики (очистка, обогащение)</td></tr><tr><td>2. Processing (Обработка)</td><td>Б. Выгрузка подготовленных данных в хранилище или витрину</td></tr><tr><td>3. Serving (Представление)</td><td>В. Приём данных из внешних источников (Kafka, API, файлы)</td></tr></table>	Этап пайплайна	Описание	1. Ingestion (Сбор)	А. Преобразование данных в формат для аналитики (очистка, обогащение)	2. Processing (Обработка)	Б. Выгрузка подготовленных данных в хранилище или витрину	3. Serving (Представление)	В. Приём данных из внешних источников (Kafka , API, файлы)	ОПК-3
Этап пайплайна	Описание									
1. Ingestion (Сбор)	А. Преобразование данных в формат для аналитики (очистка, обогащение)									
2. Processing (Обработка)	Б. Выгрузка подготовленных данных в хранилище или витрину									
3. Serving (Представление)	В. Приём данных из внешних источников (Kafka , API, файлы)									
	<table><tr><td>Ответ:</td><td>1–Б, 2–В, 3–А</td></tr></table>	Ответ:	1–Б, 2–В, 3–А							
Ответ:	1–Б, 2–В, 3–А									

9	Расположите в правильной последовательности этапы построения аналитического обзора на основе профессиональной информации о системах Big Data: 1 Сбор и систематизация источников данных (статьи, документация, отчёты) 2 Формулировка выводов и рекомендаций по выбору технологий 3 Выделение главного и структурирование информации по тематическим блокам 4 Оформление обзора в виде текста с таблицами, графиками и выводами 5 Определение цели обзора и целевой аудитории	ОПК-3
	Ответ: 5 → 1 → 3 → 4 → 2	
10	Расположите в правильной последовательности шаги анализа профессиональной информации при выборе системы хранения больших данных: 1 Сравнение альтернатив по критериям (стоимость, скорость, масштабируемость) 2 Формулировка требований к системе хранения (объём, скорость записи/чтения, тип данных) 3 Выделение главного – выбор наиболее критичных критериев для проекта 4 Оформление аналитического заключения с рекомендацией по выбору 5 Сбор информации о доступных системах (ClickHouse, MongoDB, HDFS)	ОПК-3
	Ответ: 2 → 5 → 1 → 3 → 4	
11	Дайте развернутый письменный ответ (3-5 предложений). В рамках анализа профессиональной информации о системах обработки больших данных выделите главное и структурируйте ответ: перечислите ключевые критерии выбора между Apache Spark и Hadoop MapReduce для проекта по обработке данных объёмом 5 ТБ с итеративными алгоритмами. Сделайте вывод и рекомендацию.	ОПК-3
	Ответ: Spark предпочтителен из-за in-мемори вычислений, скорости и поддержки итеративных алгоритмов; MapReduce – для пакетной обработки с низкими требованиями к памяти. Рекомендация: Spark.	
12	Дайте развернутый письменный ответ (3-5 предложений). Проанализируйте профессиональную информацию о моделях хранения NoSQL и выделите главное: в каких случаях следует применять колоночные СУБД (ClickHouse) вместо документо-ориентированных (MongoDB)? Сформулируйте вывод и рекомендацию.	ОПК-3
	Ответ: ClickHouse для агрегаций и аналитики на больших объёмах, MongoDB для гибкой схемы и операций CRUD. Рекомендация: ClickHouse для аналитических витрин.	
13	Дайте развернутый письменный ответ (3-5 предложений). Структурируйте информацию об инструментах потоковой обработки (Apache Flink, Spark Streaming, Kafka Streams). Выделите главное – по каким критериям выбирать инструмент для проекта с требованиями latency < 100 мс и exactly-once семантикой. Оформите краткий вывод.	ОПК-3
	Ответ: Apache Flink (низкая задержка, exactly-once, управление состоянием); Kafka Streams – лёгкий, но ограничен; Spark Streaming – микробатчи, задержка выше. Вывод: Flink.	
14	Дайте развернутый письменный ответ (3-5 предложений). В рамках анализа профессиональной информации об объёмах данных рассчитайте требуемый размер кластера для хранения 200 ТБ данных в HDFS с коэффициентом репликации 3 и средним размером блока 128 МБ. Оцените количество узлов, если каждый узел имеет 8 ТБ дискового пространства. Ответ дайте числом (округлить вверх).	ОПК-3
	Ответ: $200 * 3 = 600$ ТБ (с учётом репликации); $600 / 8 = 75$ узлов	
15	Дайте развернутый письменный ответ (3-5 предложений) Проанализируйте подходы к построению ETL-пайплайнов (ELT vs ETL) для больших данных. Выделите главное и сформулируйте рекомендацию – какой подход предпочтительнее для облачных решений (Snowflake, BigQuery) и почему. Оформите в виде краткого аналитического обзора (2–3 предложения).	ОПК-3
	Ответ: ELT предпочтительнее: сначала загрузка, затем трансформация в облачных хранилищах с мощным вычислительным движком; ETL – для локальных систем; рекомендация: ELT для масштабируемости и производительности.	
16	В рамках изучения текущего состояния программного обеспечения кластера обработки больших данных выявлено, что задачи выполняются медленно из-за высокой нагрузки на дисковую подсистему. Какой компонент следует модернизировать для повышения производительности без замены всего ПО? а) Заменить HDD на SSD в узлах хранения данных б) Обновить версию Apache Hadoop до более новой в) Увеличить количество узлов в кластере г) Перейти с Java на Python для написания задач	ОПК-5
	Ответ: а	

17	При выборе программного обеспечения для системы потоковой обработки данных с низкой задержкой (latency < 100 мс) и гарантией exactly-once семантики, какое решение следует выбрать для эффективного решения профессиональной задачи? a) Apache Hadoop MapReduce b) Apache Flink c) Apache Hive d) PostgreSQL <table><tr><td>Ответ:</td><td>b</td></tr></table>	Ответ:	b	ОПК-3								
Ответ:	b											
18	В ходе изучения аппаратной инфраструктуры проекта по анализу больших данных было установлено, что текущий объём оперативной памяти недостаточен для кэширования промежуточных данных Spark. Какое действие по модернизации аппаратного обеспечения следует выполнить в первую очередь? a) Добавить оперативную память на узлы кластера b) Заменить процессоры на более мощные c) Увеличить сетевую пропускную способность d) Перейти на более новую версию Spark <table><tr><td>Ответ:</td><td>a</td></tr></table>	Ответ:	a	ОПК-5								
Ответ:	a											
19	При разработке программного обеспечения для ETL-пайплайна обработки больших данных требуется выбрать систему оркестрации для управления зависимостями между задачами. Какое ПО следует выбрать для эффективного решения профессиональной задачи? a) Apache Airflow b) Apache Kafka c) Apache ZooKeeper d) Redis <table><tr><td>Ответ:</td><td>a</td></tr></table>	Ответ:	a	ОПК-5								
Ответ:	a											
20	В рамках выбора программного обеспечения для построения аналитической витрины на больших объёмах данных (OLAP) с поддержкой SQL и высокой скоростью агрегаций, какую систему следует выбрать? a) ClickHouse b) MongoDB c) Redis d) Elasticsearch <table><tr><td>Ответ:</td><td>a</td></tr></table>	Ответ:	a	ОПК-5								
Ответ:	a											
21	Соотнесите компонент программного или аппаратного обеспечения с его ролью в экосистеме Big Data при разработке и модернизации систем: <table><tr><th>Компонент</th><th>Роль в экосистеме Big Data</th></tr><tr><td>1. HDFS (NameNode, DataNode)</td><td>А. Управление ресурсами и планирование задач</td></tr><tr><td>2. Apache Spark (Executors, Driver)</td><td>Б. Распределённое хранение с отказоустойчивостью</td></tr><tr><td>3. YARN (ResourceManager, NodeManager)</td><td>В. Вычисления in-memory, обработка потоков, MLlib</td></tr></table> <table><tr><td>Ответ:</td><td>1–Б, 2–В, 3–А</td></tr></table>	Компонент	Роль в экосистеме Big Data	1. HDFS (NameNode, DataNode)	А. Управление ресурсами и планирование задач	2. Apache Spark (Executors, Driver)	Б. Распределённое хранение с отказоустойчивостью	3. YARN (ResourceManager, NodeManager)	В. Вычисления in-memory, обработка потоков, MLlib	Ответ:	1–Б, 2–В, 3–А	ОПК-5
Компонент	Роль в экосистеме Big Data											
1. HDFS (NameNode, DataNode)	А. Управление ресурсами и планирование задач											
2. Apache Spark (Executors, Driver)	Б. Распределённое хранение с отказоустойчивостью											
3. YARN (ResourceManager, NodeManager)	В. Вычисления in-memory, обработка потоков, MLlib											
Ответ:	1–Б, 2–В, 3–А											
22	Соотнесите тип задачи с программным обеспечением, которое следует выбрать для её эффективного решения: <table><tr><th>Тип задачи</th><th>Рекомендуемое ПО</th></tr><tr><td>1. Потоковая обработка событий (логи, IoT)</td><td>А. Apache Hive</td></tr><tr><td>2. Пакетная обработка с итеративными алгоритмами (ML)</td><td>Б. Apache Flink</td></tr><tr><td>3. SQL-запросы к данным в HDFS</td><td>В. Apache Spark (MLlib)</td></tr></table> <table><tr><td>Ответ:</td><td>1–Б, 2–В, 3–А</td></tr></table>	Тип задачи	Рекомендуемое ПО	1. Потоковая обработка событий (логи, IoT)	А. Apache Hive	2. Пакетная обработка с итеративными алгоритмами (ML)	Б. Apache Flink	3. SQL-запросы к данным в HDFS	В. Apache Spark (MLlib)	Ответ:	1–Б, 2–В, 3–А	ОПК-5
Тип задачи	Рекомендуемое ПО											
1. Потоковая обработка событий (логи, IoT)	А. Apache Hive											
2. Пакетная обработка с итеративными алгоритмами (ML)	Б. Apache Flink											
3. SQL-запросы к данным в HDFS	В. Apache Spark (MLlib)											
Ответ:	1–Б, 2–В, 3–А											
23	Соотнесите действие по модернизации с его влиянием на производительность системы при изучении и модернизации программно-аппаратного комплекса: <table><tr><th>Действие</th><th>Влияние на производительность</th></tr><tr><td>1. Замена HDD на SSD в узлах хранения</td><td>А. Снижение времени выполнения задач за счёт уменьшения накладных расходов на планирование</td></tr><tr><td>2. Увеличение количества ядер CPU на узлах</td><td>Б. Ускорение операций ввода-вывода и уменьшение задержек чтения/записи</td></tr><tr><td>3. Переход на более быстрый сетевой интерфейс (10GbE)</td><td>В. Ускорение параллельных вычислений за счёт обработки большего числа задач одновременно</td></tr></table>	Действие	Влияние на производительность	1. Замена HDD на SSD в узлах хранения	А. Снижение времени выполнения задач за счёт уменьшения накладных расходов на планирование	2. Увеличение количества ядер CPU на узлах	Б. Ускорение операций ввода-вывода и уменьшение задержек чтения/записи	3. Переход на более быстрый сетевой интерфейс (10GbE)	В. Ускорение параллельных вычислений за счёт обработки большего числа задач одновременно	ОПК-5		
Действие	Влияние на производительность											
1. Замена HDD на SSD в узлах хранения	А. Снижение времени выполнения задач за счёт уменьшения накладных расходов на планирование											
2. Увеличение количества ядер CPU на узлах	Б. Ускорение операций ввода-вывода и уменьшение задержек чтения/записи											
3. Переход на более быстрый сетевой интерфейс (10GbE)	В. Ускорение параллельных вычислений за счёт обработки большего числа задач одновременно											

	Ответ: 1–Б, 2–В, 3–А	
24	Расположите в правильной последовательности этапы изучения и модернизации программно-аппаратного обеспечения кластера обработки больших данных: 1 Выявление узких мест (мониторинг, профилирование нагрузки) 2 Внедрение обновлений и настройка новых компонентов 3 Оценка текущего состояния системы и сбор показателей производительности 4 Разработка плана модернизации (выбор компонентов, бюджет, сроки) 5 Тестирование модернизированной системы и сравнение с предыдущими показателями Ответ: 3 → 1 → 4 → 2 → 5	ОПК-5
25	Расположите в правильной последовательности шаги выбора программного и аппаратного обеспечения для нового проекта по обработке больших данных: 1 Определение критериев выбора (производительность, стоимость, масштабируемость) 2 Проведение тестирования (PoC) на пилотном объёме данных 3 Анализ требований проекта (объём данных, скорость, тип нагрузки) 4 Сравнение альтернатив (бенчмарки, анализ документации) 5 Принятие решения о закупке и развёртывании Ответ: → 1 → 4 → 2 → 5	ОПК-5
26	Дайте развернутый ответ В рамках изучения текущего программного обеспечения проекта по анализу больших данных выявлено, что используется устаревшая версия Spark. Предложите план модернизации с учётом рисков и опишите ожидаемый эффект. Ответ: Аудит текущих приложений на совместимость с новой версией; 2. Развёртывание тестового кластера; 3. Миграция задач; 4. Поэтапный переход; эффект: ускорение до 30%, новые оптимизации (AQE, dynamic partitioning).	ОПК-5
27	Дайте развернутый ответ Выберите программное обеспечение для системы мониторинга кластера Big Data, позволяющее собирать метрики (CPU, память, диск, сеть) и визуализировать их в реальном времени. Обоснуйте выбор для эффективного решения профессиональной задачи. Ответ: Prometheus + Grafana – сбор метрик через экспортёры, мощная визуализация, оповещения, горизонтальное масштабирование; интеграция с Hadoop/Spark.	ОПК-5
28	Дайте развернутый ответ В рамках разработки архитектуры системы хранения больших данных выберите и обоснуйте тип хранилища (объектное S3 или HDFS) для проекта с частыми запросами на чтение малых объёмов данных (случайный доступ) и требованием низкой стоимости хранения. Ответ: S3-совместимое хранилище (MinIO / AWS S3) – дешевле, лучше для случайного доступа, поддерживает S3 API; HDFS – для потокового чтения больших файлов и высокой пропускной способности.	ОПК-5
29	Дайте развернутый ответ При выборе аппаратного обеспечения для кластера Spark рассчитайте необходимое количество узлов, если требуется обработать 10 ТБ данных in-memory, а каждый узел имеет 256 ГБ ОЗУ. Учитывайте коэффициент кэширования 0.7 и резерв 30% для системных процессов. Ответ дайте числом (округлить вверх). Ответ: $10\,000\text{ ГБ} * 0.7 = 7\,000\text{ ГБ}; 256 * 0.7 = 179\text{ ГБ на узел}; 7000 / 179 \approx 39.1 \rightarrow 40\text{ узлов}$	ОПК-5
30	Дайте развернутый ответ Опишите, как вы будете модернизировать аппаратную инфраструктуру кластера для перехода с пакетной обработки (Hadoop MapReduce) на потоковую (Apache Flink), не меняя полностью оборудование. Назовите ключевые изменения. Ответ: Увеличение оперативной памяти для кэширования состояний; настройка сетевых интерфейсов для уменьшения задержек; использование SSD вместо HDD; увеличение количества ядер CPU; балансировка нагрузки и конфигурирование Flink checkpointing (distributed snapshots).	ОПК-5
31	В рамках применения методов математического моделирования для оценки производительности кластера Big Data выберите модель, которая позволяет описать зависимость времени выполнения задачи от объёма данных и количества узлов: а) Линейная регрессия б) Имитационная модель (система массового обслуживания) с) Модель на основе цепей Маркова д) Нейросетевая модель Ответ: а	ОПК-7

32	При использовании методов научных исследований для оптимизации ресурсов кластера распределённой обработки данных выберите метод, который позволяет оценить влияние различных конфигураций (количество узлов, объём памяти) на время выполнения задачи без проведения натурного эксперимента: a) Экспертный опрос b) Имитационное моделирование (Discrete-Event Simulation) c) Корреляционный анализ d) Кластерный анализ	ОПК-7								
	<table><tr><td>Ответ:</td><td>b</td></tr></table>	Ответ:	b							
Ответ:	b									
33	В рамках математического моделирования системы хранения больших данных выберите показатель, который используется для оценки надёжности хранения с учётом коэффициента репликации и частоты отказов узлов: a) MTBF (среднее время между отказами) b) Доступность (Availability) c) Коэффициент готовности (A) d) Вероятность потери данных (Data Loss Probability)	ОПК-7								
	<table><tr><td>Ответ:</td><td>d</td></tr></table>	Ответ:	d							
Ответ:	d									
34	При применении методов научных исследований для прогнозирования роста объёма данных в проекте (Big Data) выберите метод, который позволяет построить трендовую модель на основе исторических данных: a) Регрессионный анализ b) Кластеризация c) Анализ главных компонент (PCA) d) Метод опорных векторов	ОПК-7								
	<table><tr><td>Ответ:</td><td>a</td></tr></table>	Ответ:	a							
Ответ:	a									
35	В рамках математического моделирования распределённой системы обработки потоковых данных выберите параметр, который моделируется с помощью распределения Пуассона для оценки нагрузки: a) Время выполнения задачи (Service Time) b) Интенсивность поступления событий (Arrival Rate) c) Объём данных в буфере d) Коэффициент загрузки системы	ОПК-7								
	<table><tr><td>Ответ:</td><td>b</td></tr></table>	Ответ:	b							
Ответ:	b									
36	Соотнесите метод математического моделирования с его применением в проектировании информационных систем: <table><tr><th>Метод моделирования</th><th>Применение в проектировании ИС</th></tr><tr><td>1. Система массового обслуживания (СМО)</td><td>А. Оценка производительности кластера с учётом интенсивности запросов и времени обработки</td></tr><tr><td>2. Регрессионная модель</td><td>Б. Прогнозирование объёма данных и времени выполнения задач на основе исторических показателей</td></tr><tr><td>3. Имитационное моделирование (Agent-based)</td><td>В. Оценка эффективности распределения нагрузки между узлами и сценариев отказоустойчивости</td></tr></table>	Метод моделирования	Применение в проектировании ИС	1. Система массового обслуживания (СМО)	А. Оценка производительности кластера с учётом интенсивности запросов и времени обработки	2. Регрессионная модель	Б. Прогнозирование объёма данных и времени выполнения задач на основе исторических показателей	3. Имитационное моделирование (Agent-based)	В. Оценка эффективности распределения нагрузки между узлами и сценариев отказоустойчивости	ОПК-7
Метод моделирования	Применение в проектировании ИС									
1. Система массового обслуживания (СМО)	А. Оценка производительности кластера с учётом интенсивности запросов и времени обработки									
2. Регрессионная модель	Б. Прогнозирование объёма данных и времени выполнения задач на основе исторических показателей									
3. Имитационное моделирование (Agent-based)	В. Оценка эффективности распределения нагрузки между узлами и сценариев отказоустойчивости									
	<table><tr><td>Ответ:</td><td>1–А, 2–Б, 3–В</td></tr></table>	Ответ:	1–А, 2–Б, 3–В							
Ответ:	1–А, 2–Б, 3–В									
37	Соотнесите этап научного исследования с его содержанием при применении методов научных исследований в практике создания информационных систем: <table><tr><th>Этап</th><th>Содержание</th></tr><tr><td>1. Постановка задачи и сбор данных</td><td>А. Проверка адекватности модели на контрольной выборке</td></tr><tr><td>2. Построение математической модели</td><td>Б. Формулировка цели, сбор исторических данных (нагрузка, метрики)</td></tr><tr><td>3. Валидация модели</td><td>В. Выбор математического аппарата (регрессия, СМО, имитация)</td></tr></table>	Этап	Содержание	1. Постановка задачи и сбор данных	А. Проверка адекватности модели на контрольной выборке	2. Построение математической модели	Б. Формулировка цели, сбор исторических данных (нагрузка, метрики)	3. Валидация модели	В. Выбор математического аппарата (регрессия, СМО, имитация)	ОПК-7
Этап	Содержание									
1. Постановка задачи и сбор данных	А. Проверка адекватности модели на контрольной выборке									
2. Построение математической модели	Б. Формулировка цели, сбор исторических данных (нагрузка, метрики)									
3. Валидация модели	В. Выбор математического аппарата (регрессия, СМО, имитация)									
	<table><tr><td>Ответ:</td><td>1–Б, 2–В, 3–А</td></tr></table>	Ответ:	1–Б, 2–В, 3–А							
Ответ:	1–Б, 2–В, 3–А									
38	Соотнесите показатель производительности с его математической интерпретацией для моделирования систем обработки данных: <table><tr><th>Показатель</th><th>Математическая интерпретация</th></tr><tr><td>1. Пропускная способность (Throughput)</td><td>А. Среднее время между поступлением запроса и получением результата</td></tr><tr><td>2. Время отклика (Response Time)</td><td>Б. Количество обработанных задач в единицу времени</td></tr><tr><td>3. Коэффициент загрузки (Utilization)</td><td>В. Отношение времени занятости ресурса к общему времени наблюдения</td></tr></table>	Показатель	Математическая интерпретация	1. Пропускная способность (Throughput)	А. Среднее время между поступлением запроса и получением результата	2. Время отклика (Response Time)	Б. Количество обработанных задач в единицу времени	3. Коэффициент загрузки (Utilization)	В. Отношение времени занятости ресурса к общему времени наблюдения	ОПК-7
Показатель	Математическая интерпретация									
1. Пропускная способность (Throughput)	А. Среднее время между поступлением запроса и получением результата									
2. Время отклика (Response Time)	Б. Количество обработанных задач в единицу времени									
3. Коэффициент загрузки (Utilization)	В. Отношение времени занятости ресурса к общему времени наблюдения									

	Ответ:	1–Б, 2–А, 3–В	
39	Расположите в правильной последовательности этапы применения методов научных исследований при проектировании системы обработки больших данных: 1. Проверка адекватности модели и корректировка параметров 2 Сбор исходных данных (метрики текущей системы, требования к производительности) 3 Формулировка гипотез о влиянии параметров (количество узлов, объём памяти) на производительность 4 Построение математической модели (СМО, регрессия, имитация) 5 Проведение экспериментов и интерпретация результатов 6 Внедрение рекомендаций в практику проектирования		ОПК-7
	Ответ:	2 → 3 → 4 → 1 → 5 → 6	
40	Расположите в правильной последовательности шаги математического моделирования для оценки времени выполнения задачи в кластере Apache Spark с учётом объёма данных и числа узлов: 1 Сбор данных о времени выполнения задач на разных конфигурациях 2 Формулировка гипотезы о линейной зависимости времени выполнения от объёма данных и обратной зависимости от числа узлов 3 Внедрение модели в процесс планирования ресурсов кластера 4 Построение регрессионной модели (например, $T = a \cdot V + b/N + c$) 5 Проверка модели на контрольной выборке и оценка точности (R^2, RMSE) 6 Интерпретация коэффициентов модели для принятия решений о масштабировании		ОПК-7
	Ответ:	1 → 2 → 4 → 5 → 6 → 3	
41	Дайте развернутый ответ Опишите, как вы примените методы научных исследований (эксперимент, регрессионный анализ) для оценки влияния количества узлов и объёма памяти на время выполнения ETL-задачи в Spark. Сформулируйте гипотезу и план эксперимента.		ОПК-7
	Ответ:	Гипотеза: время выполнения обратно пропорционально числу узлов и уменьшается с увеличением памяти. Эксперимент: варьирование узлов (4, 8, 16), памяти (64, 128, 256 ГБ), замер времени на одном наборе данных, регрессионная модель.	
42	Дайте развернутый ответ В рамках математического моделирования системы хранения больших данных предложите модель для оценки вероятности потери данных в HDFS с учётом коэффициента репликации (R) и вероятности отказа узла (p). Опишите ограничения модели.		ОПК-7
	Ответ:	Модель: $P_loss = p^R$ (для независимых отказов), но при отказе узлов одной стойки – зависимость; ограничения: не учитывает сетевые задержки, восстановление реплик.	
43	Дайте развернутый ответ В рамках математического моделирования кластера Big Data известно, что среднее время выполнения задачи на 10 узлах составляет 120 секунд, а на 20 узлах – 70 секунд. Используя степенную модель зависимости времени от числа узлов ($T = a \cdot N^b$), найдите показатель b (масштабируемость). Ответ дайте с точностью до двух знаков.		ОПК-7
	Ответ:	$70 / 120 = (10/20)^b \rightarrow 0.583 = 0.5^b \rightarrow b = \log(0.583)/\log(0.5) \approx 0.78$	
44	Дайте развернутый ответ Опишите, как вы примените имитационное моделирование (система массового обслуживания) для оценки пропускной способности кластера Kafka при разных сценариях нагрузки. Какие параметры являются входными и выходными?		ОПК-7
	Ответ:	Входные параметры: интенсивность поступления сообщений (λ), время обработки (μ), число партиций; выходные: средняя задержка, длина очереди, коэффициент загрузки. Модель M/M/k для оценки.	
45	Дайте развернутый ответ В рамках применения методов научных исследований для оптимизации кластера Big Data предложите план экспериментов по выбору оптимального размера блока HDFS и количества реплик для задачи потокового чтения. Обоснуйте выбор метрик.		ОПК-7
	Ответ:	План экспериментов: варьирование размера блока (64, 128, 256 МБ) и реплик (2, 3, 4); метрики: пропускная способность чтения, задержка, использование дискового пространства; оптимальный по соотношению скорость/накладные расходы.	

7. Оценочные материалы промежуточной аттестации

Экзамен первый семестр

№ п/п	Содержание вопроса		Компетенция
		Правильный ответ (ключ ответа)	

1	Дайте развернутый ответ Проведите анализ профессиональной информации о системах обработки больших данных (Hadoop, Spark, Flink), выделите главное и структурируйте критерии выбора между ними для проекта с требованиями: объём данных – 10 ТБ, тип нагрузки – пакетная и потоковая обработка. Сформулируйте рекомендацию.	ОПК-7
	Ответ: Spark (пакетная + потоковая, in-memory) предпочтительнее Hadoop (только пакетная, дисковая) и Flink (потоковая, сложнее в настройке).	
2	Дайте развернутый ответ Проанализируйте модели хранения NoSQL (документо-ориентированные, колоночные, ключ-значение) и выделите главное – для какой задачи какая модель подходит. Оформите краткий аналитический обзор (2–3 предложения) с выводами.	ОПК-7
	Ответ: Документо-ориентированные – для гибкой схемы; колоночные – для аналитики; ключ-значение – для кэширования. Вывод: выбор зависит от типа запросов и структуры данных.	
3	Дайте развернутый ответ В рамках структурирования информации о ETL-пайплайнах опишите последовательность этапов обработки данных (Ingestion → Processing → Serving) для системы рекомендаций на основе логов пользователей. Выделите ключевые инструменты для каждого этапа.	ОПК-7
	Ответ: Ingestion: Kafka; Processing: Spark Structured Streaming; Serving: Cassandra / ClickHouse для хранения признаков и результатов.	
4	Дайте развернутый ответ Проанализируйте подходы к обеспечению отказоустойчивости в распределённых системах (репликация, контрольные точки, журналирование). Выделите главное – какой подход критичен для потоковой обработки (Flink) и почему. Сформулируйте вывод.	ОПК-7
	Ответ: Контрольные точки (checkpointing) – критичны для Flink, так как обеспечивают exactly-once и восстановление после сбоев.	
5	Дайте развернутый ответ Структурируйте информацию о метриках производительности кластера Big Data (пропускная способность, задержка, использование ресурсов). Оформите краткий обзор с рекомендациями по выбору ключевых метрик для мониторинга.	ОПК-7
	Ответ: Ключевые метрики: Throughput (задачи/сек), Latency (p95), CPU и Memory Utilization, Disk I/O, Network I/O; рекомендовано отслеживать через Prometheus + Grafana.	
6	Дайте развернутый ответ В рамках модернизации кластера Big Data предложите план замены HDD на SSD в узлах хранения. Опишите ожидаемый эффект на производительность и риски.	ОПК-7
	Ответ: Эффект: ускорение чтения/записи в 5–10 раз, снижение задержек; риски: стоимость, несовместимость с контроллерами, необходимость миграции данных.	
7	Дайте развернутый ответ Выберите программное обеспечение для построения аналитической витрины данных (OLAP) с поддержкой SQL, высокой скоростью агрегаций и масштабированием до 100 ТБ. Обоснуйте выбор.	ОПК-7
	Ответ: ClickHouse – колоночная СУБД, поддержка SQL, высокая скорость агрегаций, масштабируемость, низкая стоимость хранения.	
8	Дайте развернутый ответ Опишите, как модернизировать аппаратную инфраструктуру для перехода с пакетной обработки (Hadoop) на потоковую (Flink). Назовите ключевые изменения и обоснуйте их.	ОПК-7
	Ответ: Увеличение ОЗУ (для состояний), настройка сети (10GbE), использование SSD; обоснование: Flink требует низких задержек и высокой пропускной способности.	
9	Дайте развернутый ответ В рамках математического моделирования кластера Big Data постройте линейную регрессию времени выполнения задачи (T) от объёма данных (V): $T = a \cdot V + b$. Известно: при $V=10$ ГБ $T=30$ с, при $V=100$ ГБ $T=180$ с. Найдите a и b. Ответ дайте в виде «a=..., b=...».	ОПК-7
	Ответ: $a = (180-30)/(100-10) = 150/90 = 1.667$; $b = 30 - 1.667 \cdot 10 = 13.33$	
10	Дайте развернутый ответ В рамках применения методов научных исследований опишите план эксперимента по оценке влияния размера блока HDFS (64, 128, 256 МБ) на пропускную способность чтения. Укажите гипотезу, переменные и ожидаемый результат.	ОПК-7
	Ответ: Гипотеза: увеличение блока до 256 МБ повышает пропускную способность за счёт уменьшения накладных расходов; независимая переменная – размер блока, зависимая – Throughput (МБ/с); ожидается рост до 128 МБ, затем насыщение.	

7.1. Уровни овладения

Компетенция: ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.

Индикатор достижения компетенции: ОПК-3.1 Анализирует профессиональную информацию, выделяет в ней главное и определяет ее структуру.

Уровень	Характеристика	Оценка в баллах
---------	----------------	-----------------

Повышенный	Достигнуто полное овладение знаниями, умениями и навыками. Студент свободно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	81-100
Базовый	Достигнуто достаточное овладение знаниями, умениями и навыками. Студент уверенно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	61-80
Пороговый	Достигнуто овладение минимально необходимыми знаниями, умениями и навыками. Студент владеет основной терминологией, умеет применять теоретические знания для решения поставленных задач в стандартных ситуациях.	41-60
Ниже порогового	Компетенция не освоена	0-40

Индикатор достижения компетенции: ОПК-3.2 Структурирует, оформляет и представляет информацию в виде аналитических обзоров с обоснованными выводами и рекомендациями.

Уровень	Характеристика	Оценка в баллах
Повышенный	Достигнуто полное овладение знаниями, умениями и навыками. Студент свободно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	81-100
Базовый	Достигнуто достаточное овладение знаниями, умениями и навыками. Студент уверенно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	61-80
Пороговый	Достигнуто овладение минимально необходимыми знаниями, умениями и навыками. Студент владеет основной терминологией, умеет применять теоретические знания для решения поставленных задач в стандартных ситуациях.	41-60
Ниже порогового	Компетенция не освоена	0-40

Компетенция: ОПК-5 Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем.

Индикатор достижения компетенции: ОПК-5.1 Изучает и модернизирует программное и аппаратное обеспечение информационных и автоматизированных систем для решения профессиональных задач.

Уровень	Характеристика	Оценка в баллах
Повышенный	Достигнуто полное овладение знаниями, умениями и навыками. Студент свободно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	81-100
Базовый	Достигнуто достаточное овладение знаниями, умениями и навыками. Студент уверенно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	61-80

Пороговый	Достигнуто овладение минимально необходимыми знаниями, умениями и навыками. Студент владеет основной терминологией, умеет применять теоретические знания для решения поставленных задач в стандартных ситуациях.	41-60
Ниже порогового	Компетенция не освоена	0-40

Индикатор достижения компетенции: ОПК-5.2 Разрабатывает и выбирает программное и аппаратное обеспечение информационных и автоматизированных систем для эффективного решения профессиональных задач.

Уровень	Характеристика	Оценка в баллах
Повышенный	Достигнуто полное овладение знаниями, умениями и навыками. Студент свободно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	81-100
Базовый	Достигнуто достаточное овладение знаниями, умениями и навыками. Студент уверенно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	61-80
Пороговый	Достигнуто овладение минимально необходимыми знаниями, умениями и навыками. Студент владеет основной терминологией, умеет применять теоретические знания для решения поставленных задач в стандартных ситуациях.	41-60
Ниже порогового	Компетенция не освоена	0-40

Компетенция: ОПК-7 Способен использовать методы научных исследований и математического моделирования в области проектирования и управления информационными системами.

Индикатор достижения компетенции: ОПК-7.2 Применяет в практике создания информационных систем современные методы научных исследований и математического моделирования.

Уровень	Характеристика	Оценка в баллах
Повышенный	Достигнуто полное овладение знаниями, умениями и навыками. Студент свободно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	81-100
Базовый	Достигнуто достаточное овладение знаниями, умениями и навыками. Студент уверенно владеет терминологией, умеет применять теоретические знания в различных ситуациях для решения поставленных задач.	61-80
Пороговый	Достигнуто овладение минимально необходимыми знаниями, умениями и навыками. Студент владеет основной терминологией, умеет применять теоретические знания для решения поставленных задач в стандартных ситуациях.	41-60
Ниже порогового	Компетенция не освоена	0-40

8. Материально-техническое и учебно-методическое обеспечение дисциплины

8.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. Терехов, А. М. Методы статистического анализа в индустрии туризма. Практический курс: учебник для вузов / А. М. Терехов, П. Г. Николенко, М. В. Ефремова. - Москва: Юрайт, 2026. - 154 с - 978-5-534-18626-0. - Текст: электронный // ИКО Юрайт: [сайт]. - URL: <https://urait.ru/bcode/589618> (дата обращения: 21.05.2026). - Режим доступа: по подписке

Дополнительная литература

1. Маркин, А. В. Программирование на SQL: учебник для спо / А. В. Маркин. - Москва: Юрайт, 2026. - 435 с - 978-5-534-11093-7. - Текст: электронный // ИКО Юрайт: [сайт]. - URL: <https://urait.ru/bcode/587541> (дата обращения: 21.05.2026). - Режим доступа: по подписке

2. Платонов, А. В. Машинное обучение: учебное пособие для вузов / А. В. Платонов. - 2-е изд. - Москва: Юрайт, 2026. - 89 с - 978-5-534-20732-3. - Текст: электронный // ИКО Юрайт: [сайт]. - URL: <https://urait.ru/bcode/589132> (дата обращения: 21.05.2026). - Режим доступа: по подписке

3. Федоров, Д. Ю. Программирование на Python: учебное пособие для вузов / Д. Ю. Федоров. - 6-е изд. - Москва: Юрайт, 2026. - 187 с - 978-5-534-19666-5. - Текст: электронный // ИКО Юрайт: [сайт]. - URL: <https://urait.ru/bcode/585806> (дата обращения: 21.05.2026). - Режим доступа: по подписке

8.2. Профессиональные базы данных и ресурсы «Интернет», к которым обеспечивается доступ обучающихся

Профессиональные базы данных

Не используются.

Ресурсы «Интернет»

1. <https://edu.lcfresh.com/> - 1С:Предприятие 8 через Интернет для учебных заведений
2. <https://stepik.org> - Платформа с онлайн-курсами от авторов-практиков
3. <https://www.rusprofile.ru/> - Система комплексного анализа рынков и компаний (СКАР) / Rusprofile

8.3. Программное обеспечение и информационно-справочные системы, используемые при осуществлении образовательного процесса по дисциплине

Перечень программного обеспечения

(обновление производится по мере появления новых версий программы)

1. MySQL Server 8.0.45;
2. MySQL Workbench 8.0.45;
3. Python версии 3.14.4 ;
4. VpWin и ErWin Обновление;
5. Консультант Плюс;
6. Мой офис;

Перечень информационно-справочных систем

(обновление выполняется еженедельно)

Не используется.

8.4. Специальные помещения, лаборатории и лабораторное оборудование

Учебные аудитории для проведения занятий лекционного типа	Комплекты ученической мебели Мультимедийный проектор Доска Экран
---	---

Учебные аудитории для проведения практических занятий (занятий семинарского типа)	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СПб
Учебные аудитории для групповых и индивидуальных консультаций	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СПб
Учебные аудитории для текущего контроля и промежуточной аттестации	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СПб
Помещения для самостоятельной работы	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СПб
Помещения для хранения и профилактического обслуживания оборудования	Комплекты специализированной мебели для хранения