

---

**УТВЕРЖДЕНА**

Решением Ученого совета  
АНО ВО «Центральный университет»  
«25» мая 2026 г.  
Протокол №2

**Рабочая программа дисциплины (модуля)  
«Системы обработки больших данных»**

**Направление подготовки:** 02.03.01 Математика и компьютерные науки

**Направленность (профиль) подготовки:** Прикладной искусственный интеллект

**Квалификация (степень) выпускника:** бакалавр

**Форма обучения:** очная

**Срок освоения программы:** 4 года

**Год набора:** 2026

**Москва  
2026**

## Содержание

|                                                            |           |
|------------------------------------------------------------|-----------|
| <b>1. Краткая характеристика дисциплины (модуля) .....</b> | <b>3</b>  |
| <b>2. Перечень планируемых результатов обучения .....</b>  | <b>5</b>  |
| <b>3. Тематический план .....</b>                          | <b>8</b>  |
| <b>4. Содержание дисциплины (модуля) .....</b>             | <b>8</b>  |
| <b>5. Учебно-методическое обеспечение .....</b>            | <b>9</b>  |
| <b>6. Материально-техническое обеспечение .....</b>        | <b>9</b>  |
| <b>7. Методические и оценочные материалы .....</b>         | <b>11</b> |

## **1. Краткая характеристика дисциплины (модуля)**

Рабочая программа дисциплины (модуля) «Системы обработки больших данных» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению 02.03.01 «Математика и компьютерные науки», профиль «Прикладной искусственный интеллект», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Системы обработки больших данных» позволяет освоить современные методы и технологии эффективного хранения, обработки и анализа огромных объёмов информации, что критично для принятия обоснованных решений в бизнесе и науке. Эти знания обеспечивают конкурентоспособность специалистов в условиях стремительного роста данных и востребованности Big Data-решений в различных отраслях.

### **Место дисциплины (модуля) в структуре образовательной программы**

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 «Математика и компьютерные науки», профиль «Прикладной искусственный интеллект», формируемую участниками образовательных отношений.

Дисциплина (модуль) является выборной и доступна для изучения на 3 курсе в 6 семестре после успешного освоения дисциплин «Разработка на Rust» или «Разработка на C++» или «Распределенные системы», «Базы данных».

**Цель изучения дисциплины (модуля):** формирование навыков проектирования, внедрения и эксплуатации масштабируемых систем для сбора, хранения и анализа больших объёмов данных с использованием современных технологий и инструментов.

#### **Задачи изучения дисциплины (модуля):**

- изучить основные технологии и архитектуры систем обработки больших данных, включая Hadoop, Spark и NoSQL базы данных;
- освоить методы анализа и визуализации больших данных с применением алгоритмов машинного обучения и статистических моделей;
- освоить навыки проектирования и реализации масштабируемых решений для обработки больших данных в облачных и распределенных средах.

#### **В результате освоения дисциплины (модуля) обучающийся должен:**

##### **знать:**

- основные понятия и архитектуру распределенных вычислительных систем;
- принципы работы MapReduce алгоритмов и их применения для обработки больших данных;
- особенности и возможности SQL-подобного интерфейса Hive для анализа данных в Hadoop;

##### **уметь:**

- развертывание и управление Hadoop кластером;
- работать с системой очередей Kafka и распределенной базой данных Cassandra;
- написание и оптимизация MapReduce задач на Hadoop для эффективной обработки данных;
- создание и выполнение запросов в Hive, анализ и устранение узких мест в запросах;

##### **владеть:**

- инструментами экосистемы Hadoop для создания и оптимизации систем обработки больших данных;

— практикой применения алгоритмов на основе MapReduce для решения практических задач в области больших данных.

## 2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

| Код и содержание компетенции                                                                                                         | Индикатор компетенции и перечень планируемых результатов обучения по дисциплине (модулю)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач | УК-1.1. Знает методы поиска и анализа информации в области разработки, основные принципы критической оценки источников информации и их релевантности.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|                                                                                                                                      | УК-1.2. Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|                                                                                                                                      | УК-1.3. Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| УК-3. Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде<br>(соответствует УНК-02)                  | <p>УК-3.1. (УНК-02.01.) Принимает участие в групповом взаимодействии в ходе профессиональной деятельности (базовый)</p> <p>УНК-02.01.01. У-1. Проявляет понимание своей роли в команде</p> <p>УНК-02.01.01. У-2. Умеет ясно выражать свои мысли в группе</p> <p>УНК-02.01.01. У-3. Способен слушать и учитывать мнения других участников команды.</p> <p>УНК-02.01.01. У-4. Умеет давать конструктивную обратную связь</p> <p>УНК-02.01.01. У-5. Умеет работать в команде, проявлять инициативу и поддерживать коллег</p> <p>УНК-02.01.01. У-6. Обладает навыками разрешения конфликтных ситуаций и поиска компромиссов</p> <p>УНК-02.01.01. З-1. Знает основы командной работы, роли и ответственности каждого участника.</p>                                                                                                          |
|                                                                                                                                      | <p>УК-3.2. (УНК-02.02.) Проявляет лидерство и осуществляет наставничество (базовый)</p> <p>УНК-02.02.01. У-1. Умеет мотивировать и вдохновлять команду для достижения общих целей</p> <p>УНК-02.02.01. У-2. Способен организовывать работу группы, распределять обязанности и контролировать выполнение задач</p> <p>УНК-02.02.01. У-3. Умеет делиться знаниями и опытом, оказывать поддержку и осуществлять наставничество коллегам</p> <p>УНК-02.02.01. У-4. Способен принимать ответственные решения и демонстрировать пример профессионализма</p> <p>УНК-02.02.01. З-1. Знает основы лидерства, мотивации и командообразования</p> <p>УНК-02.02.01. З-2. Знает принципы эффективного наставничества и развития персонала</p> <p>УНК-02.02.01. З-3. Знает методы оценки эффективности работы команды и индивидуальных участников</p> |
| ОПК-6. Способен разрабатывать алгоритмы и компьютерные программы,                                                                    | ОПК-6.1. Знает принципы алгоритмизации, основные структуры данных, парадигмы программирования, методы разработки, тестирования и отладки компьютерных программ.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |

|                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| пригодные для практического применения                                                            | ОПК-6.2. Умеет разрабатывать, документировать, тестировать и отлаживать алгоритмы и компьютерные программы, пригодные для практического применения, с использованием современных языков и сред разработки.                                                                                                                                                                                                                                                                                     |
|                                                                                                   | ОПК-6.3. Имеет практический опыт разработки и сопровождения компьютерных программ и программных комплексов для решения прикладных задач.                                                                                                                                                                                                                                                                                                                                                       |
| ППК-ДА3. Способен проводить аналитические исследования с применением технологий больших данных    | ППК-ДА3.1. – Применяет методы машинного обучения и статистического анализа (базовый)<br>ППК-ДА3.1. 3-1. Знает основные алгоритмы машинного обучения<br>ППК-ДА3.1. 3-2. Знает методы статистического анализа данных<br>ППК-ДА3.1. 3-3. Знает критерии выбора алгоритмов для различных задач<br>ППК-ДА3.1. У-1. Умеет реализовывать алгоритмы машинного обучения<br>ППК-ДА3.1. У-2. Умеет интерпретировать результаты статистического анализа                                                    |
|                                                                                                   | ППК-ДА3.2. – Обеспечивает соответствие результатов анализа бизнес-задачам заказчика (средний)<br>ППК-ДА3.2. 3-2. Знает ключевые бизнес-метрики в предметной области                                                                                                                                                                                                                                                                                                                            |
| ППК-Р5. Способен разрабатывать требования и проектировать программное обеспечение                 | ППК-Р5.1. Анализирует возможности реализации требований к компьютерному программному обеспечению (средний)<br>ППК-Р5.1. 3-1. Знает возможности существующей программно-технической архитектуры.<br>ППК-Р5.1. 3-2. Знает возможности современных и перспективных средств разработки программных продуктов, технических средств.                                                                                                                                                                 |
| ППК-Р6. Способен участвовать в промышленной разработке программного обеспечения                   | ППК-Р6.1. Работает в соответствии с промышленными методологиями разработки (средний)<br>ППК-Р6.1. 3-1. Знает принципы Agile и их применение в промышленных проектах.<br>ППК-Р6.1. 3-2. Знает процессы code review, принципы коллективного владения кодом (collective code ownership).<br>ППК-Р6.1. У-1. Умеет оценивать объем задачи и срок ее выполнения, участвовать в планировании спринтов.<br>ППК-Р6.1. У-2. Умеет работать в команде с использованием инструментов управления проектами. |
| ППК-Р7. Способен применять искусственный интеллект (ИИ) для генерации и отладки программного кода | ППК-Р7.1. Применяет ИИ-инструменты для генерации программного кода (базовый)<br>ППК-Р7.1. 3-1. Знает принципы работы современных генеративных ИИ-моделей для генерации кода.<br>ППК-Р7.1. У-1. Умеет формулировать корректные текстовые запросы (промты) для генерации кода.<br>ППК-Р7.1. У-2. Умеет интегрировать ИИ-инструменты в среду разработки.                                                                                                                                          |
|                                                                                                   | ППК-Р7.2. Использует ИИ для анализа и отладки кода (средний)<br>ППК-Р7.2. 3-1. Знает методы ИИ-анализа кода.<br>ППК-Р7.2. 3-2. Знает форматы и инструменты для автоматизированного тестирования с ИИ.                                                                                                                                                                                                                                                                                          |

|                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ППК-ДА4. Способен выявлять бизнес-проблемы или бизнес-возможности                      | ППК-ДА4.2. Выявление истинных бизнес-проблем или бизнес-возможностей (продвинутый)<br>ППК-ДА4.2. 3-1. Знает предметную область и специфику деятельности организации в объеме, достаточном для решения задач бизнес-анализа<br>ППК-ДА4.2. 3-2. Знает инструменты, техники анализа бизнес-ситуации и предметной области, включая анализ данных.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| РГ-1. Способен руководить процессами разработки компьютерного программного обеспечения | РГ-1.2. Руководит проверкой работоспособности компьютерного программного обеспечения (средний)<br>РГ-1.2. 3-1. Знает основные методы измерения и оценки характеристик компьютерного программного обеспечения<br>РГ-1.2. 3-2. Знает методы и средства проверки работоспособности компьютерного программного обеспечения<br>РГ-1.2. У-1. Умеет распределять задачи на проверку работоспособности компьютерного программного обеспечения между исполнителями.<br>РГ-1.2. У-2. Умеет оценивать результаты проверки работоспособности компьютерного программного обеспечения<br>РГ-1.3. Руководит интеграцией программных модулей и компонентов компьютерного программного обеспечения (продвинутый)<br>РГ-1.3. 3-1. Знает методы и средства сборки модулей и компонентов компьютерного программного обеспечения<br>РГ-1.3. 3-2. Знает методы и программные интерфейсы взаимодействия компьютерного программного обеспечения с внешними программными компонентами |
| РГ-4. Способен организовать промышленную разработку программного обеспечения           | РГ-4.1. Внедряет промышленные методологии разработки (средний)<br>РГ-4.1. 3-1. Знает принципы Agile и их применение в промышленных проектах.<br>РГ-4.1. 3-2. Знает процессы code review, принципы коллективного владения кодом (collective code ownership).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

### 3. Тематический план

| №п/<br>п                                  | Наименование<br>раздела дисциплины<br>(модуля)          | Трудоемкость, академические часы |                      |          |                        | ТКУ<br>(текущий<br>контроль<br>успеваемости) |
|-------------------------------------------|---------------------------------------------------------|----------------------------------|----------------------|----------|------------------------|----------------------------------------------|
|                                           |                                                         | Очная форма                      |                      |          |                        |                                              |
|                                           |                                                         | Аудиторная работа                |                      | Контроль | Самостоятельная работа |                                              |
|                                           |                                                         | Лекции                           | Практические занятия |          |                        |                                              |
| 1                                         | Основы системы Hadoop и MapReduce                       | 8                                | 8                    |          | 30                     | Кейс                                         |
| 2                                         | Инструменты анализа данных в экосистеме Hadoop          | 8                                | 8                    | 2        | 30                     | Кейс                                         |
| 3                                         | Потоковая обработка данных и инструменты взаимодействия | 6                                | 6                    | 2        | 30                     | Кейс                                         |
| 4                                         | NoSQL хранилища и архитектуры данных                    | 8                                | 8                    | 2        | 32                     | Кейс<br>Проект                               |
|                                           | Зачет с оценкой                                         |                                  |                      | 2        |                        |                                              |
| Итого:                                    |                                                         | 30                               | 30                   | 8        | 122                    |                                              |
| Объем дисциплины (модуля)<br>(в ак. ч.)   |                                                         | 190                              |                      |          |                        |                                              |
| Объем дисциплины (модуля)<br>(в зач. ед.) |                                                         | 5                                |                      |          |                        |                                              |

### 4. Содержание дисциплины (модуля)

| №п/п | Наименование раздела дисциплины (модуля)                | Содержание дисциплины (модуля) по темам                                                                                                                                                                                      |
|------|---------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | Основы системы Hadoop и MapReduce                       | Основы Hadoop и MapReduce. Архитектура распределённых вычислительных систем. Паттерны и оптимизация MapReduce задач. Hive и SQL-on-Hadoop для аналитики                                                                      |
| 2    | Инструменты анализа данных в экосистеме Hadoop          | Оптимизация запросов Hive и MPP-решения (ClickHouse, Greenplum). Введение в Apache Spark: RDD, DataFrame, Dataset. Продвинутый Spark: Catalyst, Tungsten, MLlib. Потоковая обработка: Spark Streaming и Structured Streaming |
| 3    | Потоковая обработка данных и инструменты взаимодействия | Основы Apache Kafka и интеграция потоков данных. NoSQL хранилища: Cassandra модель данных и запросы. Оптимизация Cassandra и масштабирование кластера                                                                        |
| 4    | NoSQL хранилища и архитектуры данных                    | Архитектуры Data Lake: Lambda, Delta Lake, Iceberg. Построение ETL/ELT пайплайнов в экосистеме Hadoop. Подготовка и реализация проектных решений. Защита проектов и разбор кейсов                                            |

## 5. Учебно-методическое обеспечение

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

Каждый студент 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.

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

### *Основная литература:*

1. Волкова, П. А. Статистическая обработка данных в учебно-исследовательских работах : учебное пособие / П.А. Волкова, А.Б. Шипунов. — Москва : ФОРУМ : ИНФРА-М, 2022. — 96 с. — (Высшее образование: Бакалавриат). - ISBN 978-5-00091-710-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1862854>

2. Мартишин, С. А. Базы данных. Практическое применение СУБД SQL и NoSQL-типа для проектирования информационных систем : учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — Москва : ФОРУМ : ИНФРА-М, 2024. — 368 с. — (Высшее образование). - ISBN 978-5-8199-0946-1. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2096940>

### *Дополнительная литература:*

1. Мартишин, С. А. Базы данных: Работа с распределенными базами данных и файловыми системами на примере MongoDB и HDFS с использованием Node.js, Express.js, Apache Spark и Scala : учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — Москва : ИНФРА-М, 2023. — 235 с. + Доп. материалы [Электронный ресурс]. — (Среднее профессиональное образование). - ISBN 978-5-16-015643-9. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2111334>

## 6. 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 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<br>(издания, курса, документа)                            | Ссылка                                                                          |
|----|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| 1. | Научная электронная библиотека elibrary.ru библиотека                          | <a href="https://elibrary.ru/defaultx.asp">https://elibrary.ru/defaultx.asp</a> |
| 2. | База данных для IT-специалистов                                                | <a href="https://habr.com">https://habr.com</a>                                 |
| 3. | База данных ScienceDirect                                                      | <a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>       |
| 4. | Официальный сайт Министерства науки и высшего образования Российской Федерации | <a href="https://minobrnauki.gov.ru/">https://minobrnauki.gov.ru/</a>           |
| 5. | Федеральный портал «Российское образование»                                    | <a href="https://www.edu.ru/">https://www.edu.ru/</a>                           |
| 6. | Информационная система "Единое окно доступа к образовательным ресурсам"        | <a href="http://window.edu.ru/">http://window.edu.ru/</a>                       |
| 7. | Единая коллекция цифровых образовательных ресурсов                             | <a href="http://school-collection.edu.ru/">http://school-collection.edu.ru/</a> |
| 8. | Федеральный центр информационно - образовательных ресурсов                     | <a href="http://fcior.edu.ru/">http://fcior.edu.ru/</a>                         |

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 |
|---------------------------------------------------------------------|---------------|------------------------------------------|
| <b>Операционные системы:</b>                                        |               |                                          |
| Microsoft Imagine (Windows Client, Server)                          | зарубежное    | лицензионное                             |
| <b>Браузеры:</b>                                                    |               |                                          |
| Яндекс.Браузер                                                      | отечественное | свободно распространяемое                |
| Google Chrome                                                       | зарубежное    | свободно распространяемое                |
| <b>Офисные приложения:</b>                                          |               |                                          |
| Microsoft Imagine (Visio, OneNote)                                  | зарубежное    | лицензионное                             |
| TeXstudio                                                           | зарубежное    | свободно распространяемое                |
| Adobe Acrobat Reader                                                | зарубежное    | свободно распространяемое                |
| <b>Программное обеспечение для планирования и учета времени:</b>    |               |                                          |
| Toggle app                                                          | зарубежное    | свободно распространяемое                |
| <b>Системы управления проектами:</b>                                |               |                                          |
| Microsoft Imagine (Project)                                         | зарубежное    | лицензионное                             |
| <b>Системы управления базами данных:</b>                            |               |                                          |
| Microsoft Imagine (SQL Server)                                      | зарубежное    | лицензионное                             |
| <b>Системы резервного копирования (backup):</b>                     |               |                                          |
| Acronis Backup Advanced for HyperV                                  | зарубежное    | лицензионное                             |
| <b>Справочно-правовые системы:</b>                                  |               |                                          |
| КонсультантПлюс: справочно-правовая система                         | отечественное | лицензионное                             |
| <b>Средства антивирусной защиты:</b>                                |               |                                          |
| Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition | отечественное | лицензионное                             |

|                                                          |            |                           |
|----------------------------------------------------------|------------|---------------------------|
| <b>Среды разработки:</b>                                 |            |                           |
| Visual Studio Code                                       | зарубежное | свободно распространяемое |
| Bash (Unix shell)                                        | зарубежное | свободно распространяемое |
| Anaconda                                                 | зарубежное | свободно распространяемое |
| Robot Operating System                                   | зарубежное | свободно распространяемое |
| CopelliaSim                                              | зарубежное | свободно распространяемое |
| Google Colaboratory                                      | зарубежное | свободно распространяемое |
| <b>Пакеты программных средств и библиотек:</b>           |            |                           |
| AutoPsy                                                  | зарубежное | свободно распространяемое |
| Interactive Disassembler (IDA)                           | зарубежное | свободно распространяемое |
| <b>Системы управления библиографической информацией:</b> |            |                           |
| Zotero                                                   | зарубежное | свободно распространяемое |
| <b>Сервисы и службы:</b>                                 |            |                           |
| Bind                                                     | зарубежное | свободно распространяемое |
| Docker                                                   | зарубежное | свободно распространяемое |

## 7. Методические и оценочные материалы

### Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Системы обработки больших данных» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, кейсы, проект, 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.

*Семинар* — это форма учебного занятия, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

*Кейс* – практическая работа студентов над реальными или смоделированными задачами, что позволяет студенту применять теоретические знания на практике.

Студент самостоятельно разрабатывает стратегию решения поставленной задачи, что способствует развитию навыков критического мышления и самостоятельного принятия решений. Такой подход помогает подготовить будущих специалистов к реальным вызовам в их профессиональной деятельности.

*Проект* – исследовательская работа по курсу и презентация результатов.

Для успешной подготовки к проекту: четко определите цели и задачи проекта, распределите роли и обязанности между участниками, а также установите сроки

выполнения каждой части работы. Регулярно проводите встречи для обсуждения прогресса и решения возникающих вопросов.

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

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента 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)**

#### **Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Системы обработки больших данных»**

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 с оценкой*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

| <b>Десятибалльная оценка</b> | <b>Пятибалльная оценка</b> | <b>Общая характеристика результата обучения по дисциплине (модулю)</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 10                           | Отлично                    | Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. 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. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами. |
| 9                            | Отлично                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 8                            | Отлично                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 7                            | Хорошо                     | Студент 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                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 6                            | Хорошо                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |

| Десятибалльная оценка | Пятибалльная оценка | Общая характеристика результата обучения по дисциплине (модулю)                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-----------------------|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                       |                     | анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать 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. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами. |
| 5                     | Удовлетворительно   | Студент 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. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.                                                                                                                                                                          |
| 4                     | Удовлетворительно   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 3                     | Не сдан             | Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.                                                                                                                                                                                                                                                                                                                                          |
| 2                     | Не сдан             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 1                     | Не сдан             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

Дисциплина (модуль) «Системы обработки больших данных» оценивается следующим образом:

| Активность      | Вес | Количество | Описание                                                                                                                                     |
|-----------------|-----|------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| Кейс            | 40% | 4          | Практическая работа студентов над реальными или смоделированными задачами, что позволяет студенту применять теоретические знания на практике |
| Проект          | 30% | 1          | Исследовательская работа по дисциплине (модулю) и презентация результатов                                                                    |
| Зачет с оценкой | 30% | 1          | Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)                       |

**Формула расчёта итоговой оценки по дисциплине (модулю) «Системы обработки больших данных»:**  $\langle 0,4 \times \text{среднее за кейсы} + 0,3 \times \text{проекты} + 0,3 \times \text{зачёт с оценкой} \rangle$ .

**Текущий контроль успеваемости обучающихся по дисциплине (модулю)**

**Примерные задания для кейсов**

## Основы системы Hadoop и MapReduce

1. **Кейс: Анализ логов веб-сервера.** Компания обрабатывает миллионы записей логов с веб-сервера. Используя MapReduce, разработайте задачу для подсчета количества запросов по IP-адресам и выявления наиболее активных пользователей, оптимизировав ее для снижения сетевых затрат.

2. **Кейс: Обработка текстовых данных.** Библиотека оцифровывает книги и нуждается в индексации слов. Создайте MapReduce-задачу для подсчета частоты слов в корпусе текстов, используя паттерны вроде WordCount, и интегрируйте ее с архитектурой Hadoop для распределенной обработки.

3. **Кейс: Агрегация финансовых транзакций.** Банк анализирует ежедневные транзакции клиентов. Реализуйте MapReduce-программу для агрегации сумм по категориям расходов, оптимизировав shuffle-фаза для минимизации времени выполнения на кластере.

4. **Кейс: Аналитика социальных сетей.** Платформа социальных медиа собирает данные о взаимодействиях пользователей. Используя Hive, напишите SQL-запрос для анализа паттернов связей между пользователями на основе MapReduce-вычислений, с оптимизацией для больших объемов данных.

5. **Кейс: Мониторинг IoT-устройств.** Компания с IoT-датчиками собирает сенсорные данные. Разработайте MapReduce-задачу для выявления аномалий в потоках данных (например, пиковые значения), интегрируя ее с архитектурой распределенных систем для реального времени.

## Инструменты анализа данных в экосистеме Hadoop

1. **Кейс: Оптимизация запросов к данным продаж.** Розничная сеть анализирует продажи в Hive. Оптимизируйте запрос для расчета средних продаж по регионам, используя партиционирование и сравнив производительность с MPP-решением вроде Greenplum.

2. **Кейс: Анализ данных с использованием Spark RDD.** Телеком-компания обрабатывает CDR-данные звонков. Создайте Spark-приложение на RDD для фильтрации и агрегации данных по длительности звонков, переходя к DataFrame для оптимизации производительности.

3. **Кейс: Машинное обучение на Spark MLlib.** Интернет-магазин прогнозирует спрос на товары. Используйте MLlib для обучения модели кластеризации на данных о покупках, интегрируя Catalyst для оптимизации запросов и Tungsten для ускорения.

4. **Кейс: Поточковая обработка событий.** Сервис стриминга анализирует просмотры в реальном времени. Реализуйте Spark Streaming для подсчета популярных видео, переходя к Structured Streaming для обработки микробатчей и визуализации трендов.

5. **Кейс: Сравнение MPP и Spark.** Финансовая компания сравнивает ClickHouse и Spark для аналитики транзакций. Напишите запросы в обоих инструментах для расчета волатильности цен, оптимизируя Hive-запросы и оценивая производительность на больших данных.

## NoSQL хранилища и архитектуры данных

1. **Кейс: Построение Data Lake с Delta Lake.** Компания строит Data Lake для хранения исторических и потоковых данных. Спроектируйте архитектуру с использованием Delta Lake для ACID-транзакций, интегрируя ETL-пайплайн для загрузки данных из различных источников.

2. **Кейс: Миграция на Iceberg.** Стартап мигрирует от традиционных хранилищ к Iceberg. Реализуйте ETL-процесс для партиционирования и версионирования данных о пользователях, обеспечивая совместимость с Hadoop-экосистемой.

3. **Кейс: Lambda-архитектура для аналитики.** Платформа e-commerce анализирует поведение покупателей. Постройте Lambda-архитектуру с NoSQL (HBase) для batch-обработки и потоковой аналитики, включая ETL для очистки данных.

**4. Кейс: Реализация проектного решения.** Команда разрабатывает систему рекомендаций. Подготовьте проектное решение с использованием ETL-пайплайнов в Hadoop для обработки данных отзывов, включая защиту проекта с демонстрацией кейсов масштабирования.

**5. Кейс: Разбор кейсов и защита.** Студенческая группа анализирует реальный кейс обработки больших данных в здравоохранении. Разработайте архитектуру Data Lake с Iceberg, реализуйте ETL/ELT для анонимизации данных и подготовьте презентацию с разбором преимуществ и вызовов.

### **Примерное описание и критерии оценивания к итоговому проекту**

#### **Описание проекта:**

В рамках итогового проекта необходимо разработать комплексное решение для обработки и анализа больших данных, демонстрирующее глубокое понимание и практические навыки работы с основными технологиями и инструментами, изученными в ходе курса. Проект должен включать развертывание и управление Hadoop кластером, реализацию MapReduce алгоритмов, использование Hive для анализа данных, применение MPP-решений (ClickHouse или GreenPlum) для эффективной аналитики, а также внедрение Spark для пакетной и потоковой обработки данных. Кроме того, проект должен демонстрировать интеграцию с Kafka для организации потоковой передачи данных и использование Cassandra в качестве NoSQL хранилища. В итоговом решении необходимо продемонстрировать современные архитектурные подходы к построению систем обработки больших данных.

#### **Требования к проекту:**

- Развертывание и настройка Hadoop кластера с демонстрацией его работоспособности.
- Разработка и запуск MapReduce задач для решения прикладных задач обработки данных.
- Создание и оптимизация Hive-запросов для анализа данных, включая работу с партиционированием и индексами.
- Использование MPP-решений (ClickHouse или GreenPlum) для выполнения сложных аналитических запросов.
- Реализация пакетной обработки данных с помощью Apache Spark, включая оптимизацию производительности.
- Разработка потокового приложения на Spark Streaming с интеграцией Kafka для обработки данных в реальном времени.
- Использование Cassandra для хранения и управления ключ-значение данными с учётом требований к репликации и согласованности.
- Описание архитектуры решения с обоснованием выбранных технологий и подходов.
- Документирование проекта с подробным описанием этапов разработки, конфигураций и результатов.

#### **Критерии оценивания:**

1. Техническая корректность и полнота решения.
2. Архитектурное решение и интеграция компонентов.
3. Оптимизация и масштабируемость.
4. Документирование и презентация проекта.
5. Практическая ценность и инновационность.

### Задания для промежуточной аттестации по дисциплине (модулю)

| № п/п | Задание                                                                                                                                   | Ответ                                                                                                                                        | Компетенция   | Индикатор         |
|-------|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|---------------|-------------------|
| 1     | Опишите роли участников команды при разработке ETL-пайплайнов и распределении задач по обработке данных.                                  | Разработчик пишет код обработки, инженер данных настраивает пайплайны, архитектор проектирует архитектуру, все несут ответственность за этап | УК-3, УК-1    | УК-3.1, УК-1.2    |
| 2     | Как мотивировать команду и распределить ресурсы при оптимизации производительности Spark-задач с ограниченным временем?                   | Четкое распределение задач, контроль сроков, пример профессионализма, оценка трудоемкости этапов проекта                                     | УК-3, РГ-1    | УК-3.2, РГ-1.2    |
| 3     | Подготовьте презентацию результатов оптимизации обработки данных для стейкхолдеров и руководства компании.                                | Структура: цель, методы, результаты, выводы, рекомендации с учетом целевых установок заинтересованных сторон                                 | УК-1, ППК-ДА3 | УК-1.3, ППК-ДА3.2 |
| 4     | Оформите документацию по проекту обработки больших данных согласно профессиональным стандартам и требованиям к техническим спецификациям. | Титульный лист, содержание, введение, основная часть, выводы, список источников, приложения по ГОСТ                                          | УК-1, ППК-Р5  | УК-1.3, ППК-Р5.1  |
| 5     | Дайте определение архитектуре Lambda и запишите её смысл в контексте обработки потоковых данных.                                          | Архитектура Lambda это подход сочетающий пакетную и потоковую обработку данных для обеспечения актуальности и надёжности результатов         | УК-1, ОПК-6   | УК-1.2, ОПК-6.1   |
| 6     | Проанализируйте экономическую информацию для выбора между облачными решениями для развёртывания Hadoop-кластера.                          | Сравнение затрат, производительности, рисков, учет специфики задач, влияние на эффективность бизнес-решений                                  | УК-1, ОПК-6   | УК-1.2, ОПК-6.2   |

|    |                                                                                                                       |                                                                                                                                      |                |                    |
|----|-----------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|----------------|--------------------|
| 7  | Обоснуйте экономическое решение о выборе стратегии обработки данных на основе анализа затрат и выгод.                 | Решение выбирается если учитывает интересы всех сторон, повышает защиту инвестора от неблагоприятных изменений                       | УК-1, ОПК-6    | УК-1.3, ОПК-6.3    |
| 8  | Запишите формулу для расчета производительности Spark-задачи и выберите метод анализа данных для оценки.              | Формула производительности и зависит от количества партиций и узлов, метод статистического анализа для оценки значимости оптимизаций | ОПК-6, ППК-ДА3 | ОПК-6.2, ППК-ДА3.1 |
| 9  | Оптимизируйте выбор формата хранения данных и обоснуйте влияние на бизнес показатели компании.                        | Выбор формата с максимальной эффективностью, оценка воздействия на ключевые показатели компании, скорость обработки запросов         | ОПК-6, ППК-P5  | ОПК-6.3, ППК-P5.1  |
| 10 | Обоснуйте экономическое решение в учебном проекте с учетом системного мышления при проектировании архитектуры данных. | Использование модели с учетом взаимосвязей между компонентами, данными и бизнес-целями компании                                      | ОПК-6, РГ-4    | ОПК-6.3, РГ-4.1    |
| 11 | Назовите правовые нормы регулирующие обработку персональных данных в распределённых системах и их смысл для проекта.  | ФЗ-152 О персональных данных, требования к локализации данных, ответственность за нарушение требований при обработке больших данных  | ППК-P5, УК-1   | ППК-P5.1, УК-1.1   |
| 12 | Примените правовые нормы при оценке рисков нарушения требований при работе с данными в Hadoop-кластере.               | Анализ рисков несоблюдения законодательных требований, оценка вероятности и пороговых значений, меры минимизации штрафов             | ППК-P5, ППК-P6 | ППК-P5.1, ППК-P6.1 |
| 13 | Дайте правовую оценку ситуации с нарушением требований по защите                                                      | Определение нарушений законодательных актов, применение                                                                              | ППК-P5, УК-1   | ППК-P5.1, УК-1.3   |

|    |                                                                                                                             |                                                                                                                               |                  |                      |
|----|-----------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|------------------|----------------------|
|    | данных в проекте обработки больших данных.                                                                                  | санкций, влияние на репутацию организации                                                                                     |                  |                      |
| 14 | Выберите метод статистического анализа для проверки гипотез о влиянии оптимизаций на скорость обработки данных.             | А/В тест, проверка гипотез о значимости различий, расчет доверительных интервалов для оценок времени выполнения               | ППК-ДА3, ППК-ДА4 | ППК-ДА3.1, ППК-ДА4.2 |
| 15 | Адаптируйте аналитическую модель оценки пайплайнов данных под бизнес потребности и оцените влияние решений.                 | Перевод технических метрик в бизнес показатели, оценка воздействия на ключевые показатели компании, оптимизация процессов     | ППК-ДА3, ППК-Р6  | ППК-ДА3.2, ППК-Р6.1  |
| 16 | Выберите тип визуализации для представления данных о результатах обработки и коммуникации с руководством.                   | График времени выполнения и доверительных интервалов, интерактивный дашборд для публичной дискуссии о результатах анализа     | ППК-ДА3, УК-1    | ППК-ДА3.1, УК-1.3    |
| 17 | Выявите бизнес проблему на основе анализа отклонений в производительности пайплайнов и примите решение.                     | Проблема низкой скорости обработки данных, выбор стратегии реагирования, решение о корректировке архитектуры или конфигурации | ППК-ДА4, ППК-Р7  | ППК-ДА4.2, ППК-Р7.2  |
| 18 | Классифицируйте бизнес возможности при масштабировании инфраструктуры больших данных с учетом стратегического планирования. | Оптимизация вычислительных ресурсов, разработка плана действий с учетом целей компании, выявление резервов роста              | ППК-ДА4, РГ-4    | ППК-ДА4.2, РГ-4.1    |
| 19 | Сформулируйте рекомендации на основе статистических выводов о результатах тестирования для управления продуктом.            | При значимом улучшении внедрение изменения, при отсутствии эффекта отказ от изменения для эффективности системы               | ППК-Р6, ППК-Р7   | ППК-Р6.1, ППК-Р7.2   |

|    |                                                                                           |                                                                                                            |            |                |
|----|-------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|------------|----------------|
| 20 | Оцените риски реализации проекта обработки больших данных и разработайте меры управления. | Анализ вероятности срывов сроков, мониторинг индикаторов успешности, пересмотр плана при выявлении проблем | РГ-1, УК-1 | РГ-1.3, УК-1.3 |
|----|-------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|------------|----------------|