

**Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение высшего  
образования  
«ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ»**

Структурное подразделение «Институт информационных технологий и анализа данных»

**УТВЕРЖДЕНА:**

на заседании Совета института ИТиАД им. Е.И.Попова

Протокол №8 от 24 февраля 2025 г.

**Рабочая программа дисциплины**

**«ОСНОВЫ BIG DATA»**

---

Направление: 09.03.02 Информационные системы и технологии

---

Информационные системы и технологии в административном управлении

---

Квалификация: Бакалавр

---

Форма обучения: очная

---

Документ подписан простой  
электронной подписью  
Составитель программы:  
Харахинов Владимир  
Александрович  
Дата подписания: 22.06.2025

Документ подписан простой  
электронной подписью  
Утвердил: Говорков Алексей  
Сергеевич  
Дата подписания: 23.06.2025

Документ подписан простой  
электронной подписью  
Согласовал: Аршинский  
Вадим Леонидович  
Дата подписания: 22.06.2025

Год набора – 2025

Иркутск, 2025 г.

# 1 Перечень планируемых результатов обучения по дисциплине, соотнесённых с планируемыми результатами освоения образовательной программы

## 1.1 Дисциплина «Основы Big Data» обеспечивает формирование следующих компетенций с учётом индикаторов их достижения

Код, наименование компетенции	Код индикатора компетенции
ПКС-3 Способен обеспечивать эффективную работу баз данных, являющихся частью различных информационных систем, включая развертывание, сопровождение, оптимизацию их функционирования	ПКС-3.1

## 1.2 В результате освоения дисциплины у обучающихся должны быть сформированы

Код индикатора	Содержание индикатора	Результат обучения
ПКС-3.1	Способен настраивать и поддерживать инфраструктуру для систем сбора, обработки, аналитики и хранения информационных потоков и массивов	<b>Знать</b> особенности классификации систем баз данных и назначение каждой из них. <b>Уметь</b> проектировать модели данных для различных типов хранилищ, разрабатывать и оптимизировать спроектированные модели. <b>Владеть</b> методами и средствами представления данных и знаний о предметной области; инструментами для работы с хранилищами данных.

## 2 Место дисциплины в структуре ООП

Изучение дисциплины «Основы Big Data» базируется на результатах освоения следующих дисциплин/практик: «Базы данных», «Объектно-ориентированное программирование», «Методы анализа данных»

Дисциплина является предшествующей для дисциплин/практик: «Основы DevOps», «Производственная практика: преддипломная практика», «Проектирование информационных систем»

## 3 Объем дисциплины

Объем дисциплины составляет – 3 ЗЕТ

Вид учебной работы	Трудоемкость в академических часах (Один академический час соответствует 45 минутам астрономического часа)	
	Всего	Семестр № 6
Общая трудоемкость дисциплины	108	108
Аудиторные занятия, в том числе:	48	48
лекции	16	16
лабораторные работы	32	32
практические/семинарские занятия	0	0

Самостоятельная работа (в т.ч. курсовое проектирование)	60	60
Трудоемкость промежуточной аттестации	0	0
Вид промежуточной аттестации (итогового контроля по дисциплине)	Зачет	Зачет

#### 4 Структура и содержание дисциплины

##### 4.1 Сводные данные по содержанию дисциплины

###### Семестр № 6

№ п/п	Наименование раздела и темы дисциплины	Виды контактной работы						СРС		Форма текущего контроля
		Лекции		ЛР		ПЗ(СЕМ)		№	Кол. Час.	
		№	Кол. Час.	№	Кол. Час.	№	Кол. Час.			
1	2	3	4	5	6	7	8	9	10	11
1	Хранилища данных и их проектирование	1	2	1	4			1, 3, 4	5	Отчет по лабораторной работе
2	Оптимизация запросов. Низкоуровневая архитектура баз данных. Мониторинг производительности сервера	2	6	2, 3	14			1, 3, 4	16	Отчет по лабораторной работе
3	NoSQL базы данных	3	4	4	8			1, 3, 4	14	Отчет по лабораторной работе
4	Инструменты для организации работы с Big Data	4	4	5	6			1, 2, 3, 4	25	Отчет по лабораторной работе
	Промежуточная аттестация									Зачет
	Всего		16		32				60	

##### 4.2 Краткое содержание разделов и тем занятий

###### Семестр № 6

№	Тема	Краткое содержание
1	Хранилища данных и их проектирование	Рассматриваются отличия в подходах к проектированию хранилищ данных и транзакционных баз
2	Оптимизация запросов. Низкоуровневая архитектура баз данных. Мониторинг производительности сервера	Рассматриваются: структуры данных, используемые для организации доступа к блокам памяти внутри базы данных; процесс сбора статистики сервером; факторы, влияющие на снижение производительности сервера баз данных, путем его мониторинга
3	NoSQL базы данных	Рассматриваются основные виды нереляционных баз данных, их место в современных

		информационных системах, а также недостатки и преимущества использования по сравнению с реляционными базами
4	Инструменты для организации работы с Big Data	Рассматриваются: концепции распределённой файловой системы; парадигмы MapReduce для обработки больших данных

#### 4.3 Перечень лабораторных работ

##### Семестр № 6

№	Наименование лабораторной работы	Кол-во академических часов
1	Проектирование специализированных хранилищ данных	4
2	Генератор данных и моделирование нагрузки на сервер	6
3	Оптимизация запросов. Оценка производительности посредством мониторинга	8
4	Проектирование, разработка и наполнение NoSQL базы данных	8
5	Создание распределенной файловой системы и применение парадигмы MapReduce	6

#### 4.4 Перечень практических занятий

Практических занятий не предусмотрено

#### 4.5 Самостоятельная работа

##### Семестр № 6

№	Вид СРС	Кол-во академических часов
1	Оформление отчетов по лабораторным и практическим работам	8
2	Подготовка к зачёту	8
3	Подготовка к практическим занятиям (лабораторным работам)	30
4	Подготовка к сдаче и защите отчетов	14

В ходе проведения занятий по дисциплине используются следующие интерактивные методы обучения: Интерактивная лекция, публичная презентация, интервью

#### 5 Перечень учебно-методического обеспечения дисциплины

##### 5.1 Методические указания для обучающихся по освоению дисциплины

##### 5.1.1 Методические указания для обучающихся по лабораторным работам:

Приведены в курсе «Основы Big Data», опубликованном в системе электронного обучения ИРНИТУ.

## 5.1.2 Методические указания для обучающихся по самостоятельной работе:

Приведены в курсе «Основы Big Data», опубликованном в системе электронного обучения ИРНИТУ.

## 6 Фонд оценочных средств для контроля текущей успеваемости и проведения промежуточной аттестации по дисциплине

### 6.1 Оценочные средства для проведения текущего контроля

#### 6.1.1 семестр 6 | Отчет по лабораторной работе

##### Описание процедуры.

По завершении лабораторной работы студент оформляет отчет, который ему необходимо защитить. В ходе защиты отчета преподаватель задаёт один теоретический вопрос по теме изучаемого раздела из списка ниже, а также один практический вопрос по программному коду, полученному в результате выполнения студентом лабораторной работы. В том случае, если студент дает нечеткий или ошибочный ответ на любой из этих вопросов, преподаватель задает один или два дополнительных вопроса.

Цели, задачи, содержание задания и порядок его выполнения, а также требования к содержанию отчета, указаны в соответствующем разделе методических указаний для каждой лабораторной работы в электронном курсе.

Вопросы для контроля.

К лабораторной работе №1:

1. Особенности проектирования Data Warehouse
2. Особенности проектирования OLTP
3. OLAP. Назначение
4. Каково влияние некорректно выбранного типа данных? Как это влияет на процесс хранения?
5. Преимущества и недостатки Data Warehouse
6. Преимущества и недостатки OLTP
7. Какие запросы в большинстве случаев обрабатываются в Data Warehouse и в OLTP? Какие реже?
8. Какие данные виды данных (исторические, актуальные) хранятся в Data Warehouse и OLTP?

К лабораторной работе №2:

1. Основные библиотеки для генерации данных
2. В чем практическая ценность наполнения базы данных до ввода ее в стадию эксплуатации?
3. В чем минус использования равномерного закона распределения для генерации данных?
4. Какие подходы могут быть использованы для реализации генератора данных?
5. Место искусственного интеллекта в генерации данных
6. Какие готовые решения для генерации данных есть на рынке? В чем их недостатки?
7. Библиотека Faker. Основные классы и методы для генерации данных.
8. Роль транзакций в процессе генерации данных

К лабораторной работе №3:

1. Структура данных B-tree. Ее преимущество при хранении большого объема данных
2. Добавление нового уровня в B-tree. Что происходит с уже имеющимися данными?

3. Кластеризованный индекс. Его назначение. Какие есть ограничения?
4. Некластеризованный индекс. Его назначение. Какие есть ограничения? В чем минусы построения таких индексов?
5. Способы проведения мониторинга сервера. Общие сведения
6. Назначение статистики для сервера
7. Как оценить производительность выполнения запроса?
8. Почему важно самому корректно описывать запрос? Как это влияет на процесс его выполнения?

К лабораторной работе №4:

1. Архитектура хранения данных в MongoDB
2. Основные виды нереляционных баз данных
3. В чем преимущества и недостатки документо-ориентированных баз данных?
4. Язык запросов в MongoDB
5. План выполнения запроса. В чем его необходимость при анализе запросов?
6. Подходы к оптимизации запросов в документо-ориентированных базах данных
7. Колоночные базы данных. Их виды и особенности хранения данных
8. В чем суть горизонтального масштабирования баз данных? Какие существуют подходы к данному масштабированию?

К лабораторной работе №5:

1. Отличительные черты BigData
2. Парадигма MapReduce
3. Datanode и namenode. Их отличия
4. Архитектура HDFS
5. В чем цель репликации данных? Какой коэффициент репликации является наиболее широкоприменяемым?
6. Какие преимущества дает распределенная модель хранения данных?
7. Место базы HBase в системе Apache Hadoop
8. Какие задачи могут быть решены с использованием парадигмы MapReduce?

### **Критерии оценивания.**

Зачтено: демонстрирует понимание основ проектирования, разработки и администрирования хранилищ данных. Владеет инструментальными средствами для достижения этих целей.

Не зачтено: наличие грубых ошибок в ответах на вопросы, непонимание сущности излагаемого вопроса

## **6.2 Оценочные средства для проведения промежуточной аттестации**

### **6.2.1 Критерии и средства (методы) оценивания индикаторов достижения компетенции в рамках промежуточной аттестации**

<b>Индикатор достижения компетенции</b>	<b>Критерии оценивания</b>	<b>Средства (методы) оценивания промежуточной аттестации</b>
ПКС-3.1	Демонстрирует способность к проектированию специализированных хранилищ данных	Выполнение практических заданий и устное собеседование

## 6.2.2 Типовые оценочные средства промежуточной аттестации

### 6.2.2.1 Семестр 6, Типовые оценочные средства для проведения зачета по дисциплине

#### 6.2.2.1.1 Описание процедуры

Зачет ставится по результатам выполнения лабораторных работ, посещения лекций и ответов на вопросы в ходе зачета. Для допуска к зачету должны быть выполнены и защищены все лабораторные работы. На зачете студент должен устно ответить на вопросы билета, а также быть способен продемонстрировать владение практическими навыками на тему вопроса из билета. Студент вправе не отвечать на вопросы билета и получить зачет по дисциплине, если он не пропустил ни одной лекции по дисциплине в течение семестра и активно работал в ходе лабораторного практикума. Студент вправе отвечать только на один из вопросов билета (по своему выбору) и получить зачет по дисциплине, если он пропустил не более 2 лекций по дисциплине. Допускается письменный ответ на вопросы билета на зачете (по решению преподавателя).

Вопросы к зачету:

1. Структура B-tree, значимость данной структуры в организации данных
2. Отличия хранилища данных от транзакционной базы
3. BigData и ее место в современных информационных системах
4. Индексы в реляционной базе данных
5. Распределенная модель файловой системы
6. Архитектура HDFS
7. Компоненты HDFS
8. Подходы к мониторингу сервера
9. Что дает анализ статистики серверу баз данных?
10. Парадигма MapReduce
11. Язык запросов в MongoDB
12. План выполнения запроса. Чтение плана. Его компоненты и назначение.
13. Колоночные базы данных. Их виды и особенности хранения данных
14. Горизонтальное масштабирование. Основные подходы
15. Особенности проектирования OLTP и Data Warehouse
16. Статистика сервера. Ее влияние на производительность выполнения запросов
17. Особенности проектирования документо-ориентированных баз данных
18. Библиотека Faker. Основные классы и методы для генерации данных.
19. Основные виды нереляционных баз данных
20. Компоненты Apache Hadoop
21. Подходы к оптимизации запросов в документо-ориентированных базах данных
22. Особенности OLAP
23. Репликация данных. Ее назначение. Реализация в распределённых файловых системах.
24. Особенности проектирования колоночных баз данных
25. Основные подходы к оптимизации запросов в нереляционных базах данных

#### 6.2.2.1.2 Критерии оценивания

<b>Зачтено</b>	<b>Не зачтено</b>
демонстрирует понимание основ проектирования, разработки и администрирования хранилищ данных.	наличие грубых ошибок в ответах на вопросы, непонимание сущности излагаемого вопроса.

Владеет инструментальными средствами для достижения этих целей.	
---	--

## **7 Основная учебная литература**

1. Макшанов А. В. Большие данные. Big Data : учебник для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь, 2023. - 188.
2. Hari Seetha. Modern Technologies for Big Data Classification and Clustering / Hari Seetha, M. Narasimha Murty, B. K. Tripathy, 2018. - 382.
3. Силен Д. Основы Data Science и Big Data. Python и наука о данных / Д. Силен, А. Мейсман, М. Али; пер. с англ. Е. Матвеев, 2018. - 334.

## **8 Дополнительная учебная литература и справочная**

1. Стружкин Н. П. Базы данных: проектирование : учебник для академического бакалавриата вузов по инженерно-техническим направлениям и специальностям / Н. П. Стружкин, В. В. Годин, 2016. - 475.

## **9 Ресурсы сети Интернет**

1. <http://library.istu.edu/>
2. <https://e.lanbook.com/>

## **10 Профессиональные базы данных**

1. <http://new.fips.ru/>
2. <http://www1.fips.ru/>

## **11 Перечень информационных технологий, лицензионных и свободно распространяемых специализированных программных средств, информационных справочных систем**

1. Лицензионное программное обеспечение Системное программное обеспечение
2. Лицензионное программное обеспечение Пакет прикладных офисных программ
3. Лицензионное программное обеспечение Интернет-браузер

## **12 Материально-техническое обеспечение дисциплины**

1. Учебная аудитория для проведения лекционных занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Оснащение: комплект учебной мебели, рабочее место преподавателя, доска. Мультимедийное оборудование (в том числе переносное): мультимедийный проектор, экран, акустическая система, компьютер с выходом в интернет.
2. Учебная аудитория для проведения лабораторных/практических (семинарских) занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Компьютерная аудитория должна быть оснащена свободно

распространяемыми программными продуктами: Microsoft SQL Server, MongoDB, Apache Hadoop.