

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
(121)»

УТВЕРЖДЕНА:

на заседании совета института ИТиАД им. Е.И.Попова
Протокол №8 от 16 февраля 2026 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: 11.06.2026

Документ подписан простой
электронной подписью
Утвердил: Говорков Алексей
Сергеевич
Дата подписания: 14.06.2026

Документ подписан простой
электронной подписью
Согласовал: Аршинский
Вадим Леонидович
Дата подписания: 11.06.2026

Год набора – 2026

Иркутск, 2026 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	Знать особенности классификации систем баз данных и назначение каждой из них. Уметь проектировать модели данных для различных типов баз данных; разрабатывать и оптимизировать спроектированные базы данных. Владеть 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	Генерация данных и моделирование действий пользователей в базе данных							1, 3, 4	7	Отчет по лаборатор ной работе
2	Оптимизация запросов. Низкоуровневая архитектура баз данных. Мониторинг производительнос ти сервера	2	6	1, 2	9			1, 3, 4	15	Отчет по лаборатор ной работе
3	Хранилища данных и их проектирование	1	2	3	4			1, 3, 4	3	Отчет по лаборатор ной работе
4	NoSQL базы данных	3	4	4	8			1, 3, 4	13	Отчет по лаборатор ной работе
5	Распределенная обработка данных на базе Apache Hadoop	4	4	5	11			1, 2, 3, 4	22	Отчет по лаборатор ной работе
	Промежуточная аттестация									Зачет
	Всего		16		32				60	

4.2 Краткое содержание разделов и тем занятий

Семестр № 6

№	Тема	Краткое содержание
1	Генерация данных и моделирование действий пользователей в базе данных	Проектирование базы данных для выбранной предметной области и реализация для нее алгоритма генерации данных. Реализация многопоточного приложения для проведения моделирования действий пользователей с данной

		базой данных.
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	Хранилища данных и их проектирование	Рассматриваются отличия в подходах к проектированию хранилищ данных и транзакционных баз
4	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
5	Распределенная обработка данных на базе Apache Hadoop	Реализация распределенной файловой системы и приложения, включающее алгоритм распределенного интеллектуального анализа данных

4.3 Перечень лабораторных работ

Семестр № 6

№	Наименование лабораторной работы	Кол-во академических часов
1	Генератор данных и моделирование действий пользователей в базе данных	1
2	Оптимизация запросов. Оценка производительности посредством мониторинга	8
3	Проектирование реляционного хранилища данных и документо-ориентированной базы	4
4	Реализация распределенной NoSQL базы данных на примере MongoDB	8
5	Распределенная обработка данных на базе Apache Hadoop	11

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	5
2	Подготовка к зачёту	8
3	Подготовка к практическим занятиям (лабораторным работам)	32

4	Подготовка к сдаче и защите отчетов	15
---	-------------------------------------	----

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 три практических вопроса 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. Приведите 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. Каково место искусственного интеллекта в генерации данных?
6. Какие готовые решения для генерации данных есть на рынке? В чем их недостатки?
7. Расскажите про библиотеку Faker (или ее аналоги). Приведите основные классы и методы для генерации данных.
8. Какова роль транзакций в процессе генерации данных?

К лабораторной работе №2:

1. Опишите структуру B-tree. Каково ее преимущество при хранении большого объема

данных?

2. Опишите процесс добавления нового уровня в B-tree. Что происходит с уже имеющимися данными?
3. Что такое кластеризованный индекс? Каково его назначение? Какие есть ограничения на использование?
4. Что такое некластеризованный индекс? Каково его назначение? Какие есть ограничения на использование? В чем минусы построения таких индексов?
5. Приведите основные способы проведения мониторинга сервера баз данных.
6. Опишите назначение статистики для сервера. Для каких целей оптимизатор использует ее?
7. Как оценить производительность выполнения запроса?
8. Почему важно самому корректно описывать запрос? Как это влияет на процесс его выполнения?

К лабораторной работе №3:

1. Опишите особенности проектирования Data Warehouse. Какие основные подходы используются для этого?
2. Опишите особенности проектирования OLTP.
3. Что такое OLAP? Каково его назначение и место в промышленных системах обработки и хранения данных?
4. Каково влияние некорректно выбранного типа данных? Как это влияет на процесс хранения?
5. Приведите преимущества и недостатки Data Warehouse.
6. Приведите 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 №4:

1. Приведите архитектуру хранения данных в MongoDB.
2. Какие существуют подходы к оптимизации запросов в документо-ориентированных базах данных?
3. Каково назначение колоночных баз данных? Приведите их основные реализации.
4. В чем суть горизонтального масштабирования баз данных? Какие существуют подходы к данному процессу?
5. Опишите основы синтаксиса языка запросов в MongoDB.
6. Приведите основные виды нереляционных баз данных.
7. Опишите основные процессы, которые возникают при поступлении запроса в распределенную базу данных на примере MongoDB.
8. В чем отличие шарда от чанка?
9. По какому принципу производится выбор ключа шардирования?
10. Для каких целей используется ReplicaSet?

К лабораторной работе №5:

1. Приведите отличительные черты BigData.
2. Опишите концепцию парадигмы MapReduce. Приведите ее основные этапы.
3. Что такое Datanode и Namenode? Чем они отличаются?
4. Опишите архитектуру HDFS.
5. В чем цель репликации данных? Какой коэффициент репликации является наиболее широкоприменяемым на практике?

6. Какие преимущества дает распределенная модель хранения данных?
7. Какое место занимает база HBase в системе Apache Hadoop?
8. Какие задачи могут быть решены с использованием парадигмы MapReduce?
9. Как используется Apache Spark для распределенной интеллектуальной обработки данных?
10. В чем состоят особенности реализации распределенного алгоритма интеллектуальной обработки данных?

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

Индикатор достижения компетенции	Критерии оценивания	Средства (методы) оценивания промежуточной аттестации
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

Зачтено	Не зачтено
демонстрирует понимание основ проектирования, разработки и администрирования хранилищ данных. Владеет инструментальными средствами для достижения этих целей.	наличие грубых ошибок в ответах на вопросы, непонимание сущности излагаемого вопроса.

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.

[Сайт] – URL: <https://e.lanbook.com/book/322664>

2. Hari Seetha. Modern Technologies for Big Data Classification and Clustering / Hari Seetha, M. Narasimha Murty, B. K. Tripathy, 2018. - 382.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-35088.pdf>

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.

4. Изучаем Spark. Молниеносный анализ данных / Х. Карау [и др.], 2015. - 303.

5. Marin I. Big Data Analysis with Python. Combine Spark and Python to Unlock the Powers of Parallel Computing and Machine Learning / I. Marin, A. Shukla, S. Vvk, 2019. - 276.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-37559.pdf>

6. Mengle S. S. R. Mastering Machine Learning on AWS. Advanced Machine Learning in Python Using SageMaker, Apache Spark, and TensorFlow / S. S. R. Mengle, M. Gurmendez, 2019. - 293.

[Сайт] – URL: <http://elib.istu.edu/viewer/view.php?file=/files3/er-37675.pdf>

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
4. Свободно распространяемое программное обеспечение Docker

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.