Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Сибирский государственный индустриальный университет»

Кафедра прикладных информационных технологий и программирования

УТВЕРЖДАЮ
Директор института
информационных технологий и
автоматизированных систем
_____ А.В. Зимин
подпись
«____» _____ 20___ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Введение в инженерию больших данных

09.04.01 «Информатика и вычислительная техника»
(направленность (профиль): «Программная инженерия»)

Квалификация выпускника
Магистр

Форма обучения
Очная форма

Срок обучения: 2 года

Год начала подготовки 2025

Новокузнецк
2025

СибГИУ
Подлинник электронного документа хранится в системе электронного согласования документов СибГИУ
ДОКУМЕНТ ПОДПИСАН ЭП
Сертификат: daba56d7b892efca3bf5c04ed0251c98b19b099c
Владелец: Директор ИИТиАС Зимин Алексей Валерьевич
Действителен: с 15 января 2025 г. по 10 апреля 2026 г.
Дата и время: 03-06-2025 10:06:23 ПРОВЕРИТЬ ПОДПИСЬ

1 Цели и задачи освоения учебной дисциплины

Целями учебной дисциплины являются:

– формирование базовых компетенций студентов в области анализа, обработки и хранения больших объемов разнородных данных (Big Data), включая изучение современных технологий и инструментов, необходимых для эффективного управления такими объемами данных.

Задачами учебной дисциплины являются:

– освоение основных понятий и принципов работы с большими данными, понимание особенностей и возможностей данной технологии;

– изучение ключевых подходов и методов обработки и анализа больших данных, освоение наиболее популярных фреймворков и платформ для распределённой обработки данных (Apache Spark, Hadoop и др.).;

– формирование практических навыков работы с инструментами сбора, очистки, хранения и визуализации больших объёмов данных, знакомство с практическими кейсами использования технологий Big Data в реальных проектах.

2 Место учебной дисциплины в структуре ООП по направлению подготовки (специальности)

Учебная дисциплина относится к учебным дисциплинам обязательной части **Блока 1 «Дисциплины (модули)»** ООП по направлению подготовки (специальности) 09.04.01 «Информатика и вычислительная техника».

Учебная дисциплина базируется на предварительном усвоении обучающимися учебных дисциплин:

– Математические и инструментальные методы анализа данных;
– Визуализация данных для принятия управленческих решений;
– Анализ и оптимизация бизнес-процессов;
– Методы и инструменты цифровой трансформации;
– Облачные технологии.

Учебная дисциплина дополняет знания и умения, получаемые по одновременно изучаемым и последующим учебным дисциплинам:

– Анализ и оптимизация бизнес-процессов;
– Методы и инструменты цифровой трансформации;
– Теория оптимизации;
– Гибкие методологии управления проектами.

3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

– Общепрофессиональные компетенции

Наименование категории (группы) ОПК	Код и наименование ОПК	Код и наименование индикатора достижения ОПК	Планируемые результаты обучения
	ОПК-2: Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ОПК-2.1 Осуществляет выбор современных сред, средств и технологий разработки программного обеспечения, в том числе интеллектуальных	– знать: методологии проектирования и разработки ПО, включая Agile, Scrum, Kanban. – уметь: анализировать поставленную задачу и выбирать подходящую среду разработки, исходя из требований проекта.
		ОПК-2.2 Разрабатывает математическое и алгоритмическое обеспечение решения прикладных задач	– знать: современные алгоритмы оптимизации, поиска, сортировки и обработки данных. – уметь: разрабатывать оригинальные алгоритмы для решения конкретных прикладных задач, основываясь на фундаментальных знаниях.
		ОПК-2.3 Применяет типовые решения и разрабатывает оригинальные программы для решения профессиональных задач	– знать: наиболее распространённые библиотеки и платформы для разработки ПО и способы их адаптации под собственные нужды. – уметь: оценивать целесообразность использования готового решения или создание оригинального продукта.
	ОПК-4: Способен применять на практике новые	ОПК-4.2 Осуществляет сбор и обработку данных,	– знать: современные методы сбора

	научные принципы и методы исследований	формулирует по результатам выводы	первичных данных и способы формирования репрезентативных выборок. – уметь: организовывать процесс сбора данных, соблюдать требования к качеству и полноте собранной информации.
	ОПК-5: Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем	ОПК-5.2 Участвует в разработке алгоритмического и программного обеспечения для информационных и автоматизированных систем	– знать: современные подходы и стандарты разработки программного обеспечения, языки программирования и технологические стеки, применяемые для создания информационных и автоматизированных систем. – уметь: анализировать техническое задание и предлагать оптимальные пути реализации функциональных требований.
		ОПК-5.3 Осуществляет интеграцию и/или модернизацию программного обеспечения с существующим программным обеспечением	– знать: принципы интеграции программного обеспечения и современные подходы к обеспечению совместимости между системами. – уметь: анализировать существующие системы и определять стратегию интеграции или модернизации программного

			обеспечения.

4 Объем и содержание учебной дисциплины

Учебные занятия по учебной дисциплине проводятся в форме контактной работы и в форме самостоятельной работы обучающихся.

Контактная работа включает в себя занятия лекционного типа (лекции), занятия семинарского типа (семинары, практические занятия, практикумы), промежуточную аттестацию обучающихся и иные формы взаимодействия обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации ООП на иных условиях, в том числе при проведении промежуточной аттестации обучающихся. Контактная работа может проводиться с применением электронного обучения, дистанционных образовательных технологий.

Объем учебной дисциплины

Семестр / курс		***ИТОГО***	***5 семестр***
Форма промежуточной аттестации			*экзамен*
Трудоёмкость	*академ. час.*	***180***	*180*
	зачетных единиц	***5***	*5*
Лекции, *академ. час.*		***16***	*16*
в форме практической подготовки		***0***	*0*
Лабораторные работы, *академ. час.*		***0***	*0*
в форме практической подготовки		***0***	*0*
Практические занятия, *академ. час.*		***32***	*32*
в форме практической подготовки		***0***	*0*
Курсовая работа / проект, *академ. час.*		***0***	*0*
в форме практической подготовки		***0***	*0*
Консультации, *академ. час.*		***0***	*0*
в форме практической подготовки		***0***	*0*
Самостоятельная работа, *академ. час.*		***114***	*114*
в форме практической подготовки		***0***	*0*
Контроль, *академ. час.*		***18***	*18*
в форме практической подготовки		***0***	*0*

Содержание учебной дисциплины

Раздел 1 Основы больших данных;

Тема 1.1 Что такое Big Data (Знакомство с технологиями Big Data. Что такое Big Data + возможности. Что такое Big Data - Сложности.);

Тема 1.2 Ключевые технологии и подходы (Ключевые технологии и подходы);

Тема 1.3 Инструменты получения данных (Инструменты получения данных. А что если данные лежат в базе данных? Apache Flume и Apache Kafka);

Раздел 2 Обработка и хранение данных;

Тема 2.1 Обработка данных (Обработка данных в Apache Spark. Обработка данных. Apache Hive. Cloudera Impala);

Тема 2.2 Хранение данных (Хранение данных. Apache Solr).

5 Перечень тем лекций

№ раздела / темы дисциплины	Темы лекций	Трудоемкость, *академ.час*	
		всего	в форме практической подготовки
Раздел 1.	Основы больших данных	8	
Раздел 2.	Обработка и хранение данных	8	
Итого:		**16**	**0**

6 Перечень тем практических занятий (семинаров)

№ раздела / темы дисциплины	Темы практических занятий (семинаров)	Трудоемкость, *академ.час*	
		всего	в форме практической подготовки
Раздел 1.	Современные технологии работы с Big Data	16	
Раздел 2.	Обработка и хранение данных	16	
Итого:		**32**	**0**

7 Перечень тем лабораторных работ

№ раздела / темы дисциплины	Темы лабораторных работ	Трудоемкость, *академ.час*	
		всего	в форме практической подготовки
	Отсутствуют		
Итого:		**0**	**0**

8 Перечень тем курсовых работ (проектов)

№ раздела / темы дисциплины	Темы курсовых работ (проектов)	Трудоемкость, *академ.час*	
		всего	в форме практической подготовки
	Отсутствуют		
Итого:		**0**	**0**

9 Виды самостоятельной работы

№ раздела / темы дисциплины	Виды самостоятельной работы	Трудоемкость, *академ.час*	
		всего	в форме практической

			подготовки
Раздел 1.	1. Изучение лекционного материала; 2. Подготовка к практическому занятию; 3. Прохождение тестирования.	57	
Раздел 2.	1. Изучение лекционного материала; 2. Подготовка к практическому занятию; 3. Прохождение тестирования.	57	
Контроль	*Подготовка к экзамену*	18	
Итого:		**132**	**0**

10 Матрица компетентностных задач по учебной дисциплине

Контролируемые разделы (темы) дисциплины	Контролируемые компетенции	Оценочные средства
Раздел 1.	ОПК-2; ОПК-4; ОПК-5	1. Наблюдение за работой на практическом занятии; 2. Тестовые задания
Раздел 2.	ОПК-2; ОПК-4; ОПК-5	1. Наблюдение за работой на практическом занятии; 2. Тестовые задания
Экзамен.	ОПК-2; ОПК-4; ОПК-5	1. Экзамен на основе результатов текущего контроля

11 Входной контроль

11.1 Цель входного контроля − определить начальный уровень подготовленности обучающихся и выстроить индивидуальную траекторию обучения. В условиях личностно-ориентированной образовательной среды результаты входного оценивания обучающегося используются как начальные значения в индивидуальном профиле академической успешности обучающегося.

11.2 Описание оценочных средств

Форма проведения входного контроля – тестирование.

Длительность тестирования – 20 минут.

Тестирование проводится в течение первых двух недель изучения учебной дисциплины в СУО «Moodle».

Для проведения тестирования используются тестовые задания для входного контроля (для ответа на 1 вопрос, как правило, выделяется 1 минута).

Шкала оценивания результатов тестирования

За каждый правильный ответ выставляется 1 балл. Оценка выставляется в зависимости от процента правильных ответов:

Оценка формируется в соответствии с критериями таблицы:

Процент результативности (правильных ответов)	Оценка
86 до 100	«отлично»
от 67до 85	«хорошо»
от 50 до 66	«удовлетворительно»
менее 50	«неудовлетворительно»

Основываясь на результатах тестирования, формируется индивидуальная траектория обучающегося (группы обучающихся), исходя из следующих соображений:

– 86 – 100 % – обучающийся готов к изучению учебной дисциплины без дополнительного повторения изученного ранее материала;

– 50 – 67 % – обучающийся в целом готов к изучению учебной дисциплины, требуется повторение отдельных вопросов(тем) изученного ранее материала, необходимого для успешного освоения учебной дисциплины;

– 49% и менее – обучающийся имеет слабую подготовку к изучению учебной дисциплины, необходимы корректирующие занятия и консультации.

Тестовые задания для входного контроля знаний обучающихся приведены в приложении.

12 Текущий контроль

12.1 Текущий контроль знаний используется для оперативного и регулярного управления учебной деятельностью (в том числе самостоятельной) обучающихся. Данный вид контроля стимулирует у обучающихся стремление к систематической самостоятельной работе по изучению учебной дисциплины и позволяет оценить степень освоения изученного материала.

Текущий контроль успеваемости осуществляется в течение семестра, в ходе повседневной учебной работы по индивидуальной инициативе педагогического работника, в том числе посредством использования СУО «Moodle».

Формы проведения текущего контроля – наблюдение за работой на практическом занятии; тестовые задания.

12.2 Описание оценочных средств

Шкалы оценивания (методики оценки)

12.2.1 Рекомендации по оцениванию результатов тестирования обучающихся

Тестирование проводится по изученным разделам (темам) учебной дисциплины для выявления уровня знаний и умений обучающегося, мотивирования обучающегося к активизации самостоятельной работы по усвоению учебного материала, выявления и устранения пробелов в знаниях, повышения дисциплины и организации учебной деятельности обучающегося.

Тестирование проводится в СУО «Moodle» в рамках часов, отведенных на самостоятельную работу обучающихся. Для тестирования используется 20 вопросов из базы типовых тестовых заданий по изученным разделам (темам) учебной дисциплины. Обучающийся выбирает один ответ из предложенных вариантов ответа. В случае выбора неправильного ответа обучающийся получает за данный вопрос 0 баллов; в случае правильного ответа – 1 балл.

Оценка формируется в соответствии с критериями таблицы:

Максимальный балл	Проходной балл	Оценка
20	не менее 18	отлично
17	не менее 14	хорошо
13	не менее 10	удовлетворительно
9	–	неудовлетворительно

12.2.2 Рекомендации по оцениванию в процессе наблюдения за работой на практическом занятии

Наблюдение за работой на практическом занятии – важнейший компонент контроля деятельности обучающихся и качества их обучения.

Критерии оценки в процессе наблюдения за работой на практическом занятии:

– самостоятельность выполнения задания и работы с методическими указаниями (учитывается работа в течение занятия, быстрота и способность нахождения в методических указаниях нужной информации);

– правильность выполнения работы (учитывается последовательность выполняемых действий, умение работать с инструментами и/или средствами вычислительной техники);

– рациональность использования времени, отведенного на задание (не одобряется затянутость выполнения работы с учетом индивидуальных особенностей обучающихся);

– активность обучающегося при выполнении практической работы (учитывается индивидуальная работа и работа в группе).

Критерии оценки по результатам наблюдения за работой на практическом занятии

Наименование параметра оценки	Максимальное количество баллов
Самостоятельно выполняет задания; умеет работать с методическими указаниями.	2,5
Указанные в методических указаниях действия выполняет в правильной последовательности; умеет работать с инструментами и/или средствами вычислительной техники.	1,25
Рационально использует время, отведенное на задание; проявляет активность при выполнении практической работы.	1,25
Итого	**5**

Максимальное количество – 5 баллов соответствует оценке «отлично», 4 балла – «хорошо», 3 балла – «удовлетворительно», 2 балла – «неудовлетворительно».

Задания и вопросы для текущего контроля знаний обучающихся по разделам (темам) учебной дисциплины приведены в приложении.

13 Промежуточная аттестация

13.1 Оценочные материалы для промежуточной аттестации обучающихся по учебной дисциплине предназначены для оценки степени достижения запланированных результатов обучения по завершению изучения учебной дисциплины в установленной учебным планом форме и позволяют определить качество усвоения изученного материала.

Итоговой формой контроля сформированности компетенций у обучающихся по учебной дисциплине является: экзамен.

Оценочные материалы для промежуточной аттестации обучающихся включают: экзамен на основе результатов текущего контроля.

13.2 Оценивание обучающегося на экзамене

Оценка	Требования к знаниям
отлично	Оценка «отлично» выставляется обучающемуся, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, использует в ответе материал различной литературы, правильно обосновывает принятое нестандартное решение, обладает приемами выполнения практических задач по формированию компетенций.
хорошо	Оценка «хорошо» выставляется обучающемуся, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении

Оценка	Требования к знаниям
	практических вопросов и задач, обладает необходимыми умениями и приемами их выполнения.
удовлетворительно	Оценка «удовлетворительно» выставляется обучающемуся, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, испытывает сложности при выполнении практических работ и затрудняется связать теорию вопроса с практикой.
неудовлетворительно	Оценка «неудовлетворительно» выставляется обучающемуся, который не знает значительной части программного материала, неуверенно отвечает, допускает серьезные ошибки, не имеет представлений по методике выполнения практической работы. Как правило, оценка «неудовлетворительно» ставится обучающимся, которые не могут продолжить обучение без дополнительных занятий по данной учебной дисциплине.

Задания и вопросы для промежуточной аттестации обучающихся приведены в приложении.

14 Учебно-методическое и информационное обеспечение учебной дисциплины

а) литература:

1 Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 4-е изд., стер. — Санкт-Петербург : Лань, 2024. — 188 с. — ISBN 978-5-507-47346-5. – URL: https://e.lanbook.com/book/362318 (дата обращения: 01.06.2025);

2 Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 89 с. — (Высшее образование). — ISBN 978-5-534-20732-3. – URL: https://urait.ru/bcode/558662 (дата обращения: 02.06.2025);

3 Миркин, Б. Г. Базовые методы анализа данных : учебник и практикум для вузов / Б. Г. Миркин. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 297 с. — (Высшее образование). — ISBN 978-5-534-19709-9. – URL: https://urait.ru/bcode/560414 (дата обращения: 02.06.2025).

б) ресурсы информационно-телекоммуникационной сети «Интернет»:

1 Электронный каталог : сайт / Научно-техническая библиотека СибГИУ. – Новокузнецк, [199 –]. – URL: http://libr.sibsiu.ru. – URL: https://libr.sibsiu.ru;

2 Электронная библиотека // Научно-техническая библиотека СибГИУ : сайт. – Новокузнецк, [200 –]. – URL: http://library.sibsiu.ru/LibrELibraryFullText.asp. – Режим доступа: для

авторизир. пользователей. – URL: https://library.sibsiu.ru/LibrELibraryFullText.asp;

3 Консультант студента : электронно-библиотечная система / ООО «КОНСУЛЬТАНТ СТУДЕНТА». – Москва, [200 –]. – URL: http://www.studentlibrary.ru. – Режим доступа: для авторизир. пользователей;

4 ЛАНЬ : электронно-библиотечная система : [коллекция «Единая профессиональная база знаний для технических вузов - Издательство Лань" ЭБС»] / ООО «Издательство ЛАНЬ». – Санкт-Петербург, [200 –]. – URL: http://e.lanbook.com. – Режим доступа: для авторизир. пользователей;

5 Образовательная платформа ЮРАЙТ / ООО «Электронное издательство ЮРАЙТ». – Москва, [200 –]. – URL: https://urait.ru. – Режим доступа: для авторизир. пользователей;

6 Университетская библиотека онлайн : электронно-библиотечная система / ООО «Директ-Медиа». – Москва, [200 –]. – URL: https://biblioclub.ru. – Режим доступа: для авторизир. пользователей. – URL: http://www.biblioclub.ru;

7 НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА eLIBRARY.RU : база данных / ООО «НЭБ». – Москва, [200 –]. – URL: http://elibrary.ru. – Режим доступа: по подписке;

8 Электронные периодические издания ИВИС : универсальная база данных / ООО «ИВИС». – Москва, [200 –]. – URL: http://eivis.ru. – Режим доступа: по подписке;

9 Онлайн-курс "Введение в инженерию больших данных". – URL: https://openedu.ru/course/misis/ABD/?session=fall_2024.

в) лицензионное и свободно распространяемое программное обеспечение:

– 7-Zip;
– Adobe Acrobat Reader;
– Astra Linux Special Edition;
– Kaspersky Endpoint Security;
– Microsoft Office;
– Microsoft Windows;
– OnlyOffice;
– Р7-Офис.

г) базы данных и информационно-справочные системы:

1 ГАРАНТ : справочно-правовая система / ООО «Правовой центр «Гарант». – Кемерово, [200 –]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.;

2 КонсультантПлюс : справочно-правовая система / ООО «Информационный центр АНВИК». – Новокузнецк, [199 –]. – Режим доступа: компьютерная сеть библиотеки Сиб. гос. индустр. ун-та.;

3 Техэксперт : информационно-справочная система / ООО «Группа компаний «Кодекс». – Кемерово, [200 –]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.

15 Материально-техническое обеспечение учебной дисциплины

Материально-техническое обеспечение учебной дисциплины включает учебные аудитории, оснащенные оборудованием, компьютерной техникой, и техническими средствами обучения, в том числе:

- учебную аудиторию для проведения занятий лекционного типа, оборудованную учебной доской, экраном и мультимедийным проектором;
- учебную аудиторию для проведения занятий семинарского типа (практических занятий);
- учебную аудиторию (помещения) для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду, научно-техническую библиотеку СибГИУ.

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС ВО по направлению подготовки (специальности) 09.04.01 «Информатика и вычислительная техника».

Составитель(и):

доцент Маслова Елена Владимировна (кафедра прикладных информационных технологий и программирования).

Рабочая программа дисциплины рассмотрена и утверждена на заседании кафедры.

Согласован:

Системный аналитик группы
системной архитектуры департамента
информационных технологий
ООО "СИ СОФТ"  Е.А. Шапошникова

Приложение

Тестовые задания для входного контроля знаний обучающихся

Выберите один правильный ответ.

1 Алфавит языка программирования – это…
a. набор символов русского языка, используемых при составлении любых текстов, называемых алфавитом языка
b. набор цифр и знаков препинания, используемых при составлении текстов, называемых алфавитом языка
c. разрешенный к использованию набор символов, с помощью которого могут быть образованы слова и величины данного языка
2 Переменная характеризуется
a. именем, типом, значением
b. именем, значением
c. значением, типом
3 Величина, значение которой меняется в процессе выполнения программы, называется
a. константой
b. переменной
c. меткой
4 Операции DIV и MOD допустимы над величинами
a. целочисленного типа
b. вещественного типа
c. символьного типа
5 Какое значение примет переменная X после выполнения оператора X=17 mod 4?
a. 1
b. 4
c. 4.25
6 Дано трехзначное число x=abc. Операция x div 10 mod 10 определяет
a. единицы
b. десятки
c. сотни
7 Цикл FOR используется в том случае, если
a. необходимо, чтобы тело цикла выполнялось хотя бы один раз
b. точно знаете, сколько раз должно быть выполнено тело цикла
c. хотите, чтобы проверка была произведена прежде, чем выполнено тело цикла
8 Массив - это
a. совокупность конечного числа данных различных типов
b. совокупность конечного числа данных одного типа
c. скалярный тип данных
9 Индекс массива определяет

a. положение элементов массива относительно друг друга
b. номер элемента в массиве
c. количество элементов
10 Разветвляющейся называется алгоритмическая конструкция,…
a. в которой выбирается один из нескольких возможных путей (вариантов) вычислительного процесса;
b. в которой все этапы решения задачи выполняются строго последовательно
c. организующая многократное выполнение указанного действия

Задания и вопросы для текущего контроля знаний обучающихся по разделам (темам) учебной дисциплины

1 Тестирование по результатам освоения лекционного материала и самостоятельной работы

Тестирование проводится для определения степени усвоения материалов лекций, а также выполнения самостоятельной работы в рамках часов, отведенных на самостоятельную работу обучающихся, в СУО «Moodle».

Перечень вопросов для тестирования:

Раздел 1 Основы больших данных

1. Что характеризуют три V больших данных (Volume, Velocity, Variety)?
   a) Скорость, разнообразие, качество
   b) Количество, качество, ценность
   c) Объём, скорость поступления, разнообразие
   d) Объём, разнообразие, доступность
2. Что такое «неструктурированные данные»?
   a) Данные, хранящиеся в таблицах БД
   b) Данные, представленные в виде графиков и диаграмм
   c) Данные, имеющие строгую схему и организацию
   d) Данные, не имеющие заранее определённой структуры
3. Основным источником больших данных являются:
   a) Электронные письма
   b) Документы Word
   c) Социальные сети, сенсоры IoT, транзакции
   d) Файлы PDF
4. Какая проблема характерна для больших данных?
   a) Недостаточная скорость интернета
   b) Проблема избыточности и шумов
   c) Маленький объём данных
   d) Неточность измерительных приборов
5. Какая парадигма описывает свойства больших данных (V^3)?
   a) Качество, консистентность, вариабельность
   b) Volume, Velocity, Variety
   c) Цена, время, место

d) Параметры качества данных

6. Основная характеристика больших данных — это их:
   a) Маленькие размеры
   b) Большие объёмы
   c) Очень быстрая передача
   d) Идеально чистая структура

7. Для какого этапа жизненного цикла больших данных характерны такие задачи, как очистка и обогащение данных?
   a) Сбор данных
   b) Подготовка данных
   c) Анализ данных
   d) Презентация результатов

8. В чём особенность больших данных по сравнению с классическими реляционными БД?
   a) Большие данные всегда имеют чёткую структуру
   b) Большие данные часто бывают неоднородными и плохо структурированными
   c) Большие данные занимают меньше места
   d) Большие данные хранятся локально

9. Одной из главных задач при работе с большими данными является:
   a) Максимальное уменьшение объёма данных
   b) Поиск полезных сигналов среди шума и бесполезных данных
   c) Сохранение всех данных без исключения
   d) Исключение любых анонимизированных данных

10. Технология, которая предназначена для параллельной обработки огромных объёмов данных, называется:
   a) MySQL
   b) PostgreSQL
   c) Hadoop
   d) MongoDB

11. Большой объём данных затрудняет традиционный подход к их обработке потому, что:
   a) Классические БД быстро справляются с большими объёмами
   b) Традиционные системы не предназначены для обработки терабайт и петабайт данных
   c) Большие данные всегда идеально чисты и готовы к анализу
   d) Нет никаких сложностей в хранении больших данных

12. Какая технология используется для горизонтального масштабирования и обработки больших объёмов данных путём разделения на узлы?
   a) RDBMS (Реляционная база данных)
   b) Hadoop
   c) Excel
   d) SQLite

13. Одна из важнейших задач при анализе больших данных — это:
   a) Генерация случайных чисел

b) Найти скрытые корреляции и закономерности
c) Удалить всю ненужную информацию сразу
d) Сделать данные абсолютно прозрачными

14. Основными источниками больших данных являются:
a) Телефоны стационарные
b) Сенсорные устройства, социальные медиа, мобильные устройства
c) Только корпоративные сервера
d) Исключительно бумажные архивы

15. Для решения проблем скорости потока данных используется подход:
a) Обычный последовательный анализ
b) Потоковая обработка (stream processing)
c) Медленная загрузка данных
d) Игнорирование данных низкой важности

16. Понятие «Velocity» в концепции больших данных отражает:
a) Объём данных
b) Скорость поступления данных
c) Однородность данных
d) Размер таблицы

17. Важнейшая характеристика больших данных, влияющая на сложность их обработки, — это:
a) Постоянная однородность
b) Многообразие форматов и типов данных
c) Незаменимость
d) Наличие идеальных метаданных

18. Основоположником идеи обработки больших данных принято считать компанию:
a) Google
b) Yahoo!
c) Facebook
d) Twitter

19. Первичная подготовка данных перед их анализом включает в себя:
a) Игнорирование редких значений
b) Очистку, фильтрацию и нормализацию данных
c) Произвольное удаление записей
d) Изменение масштаба вручную

20. Что входит в концепцию Variety (разнообразие) больших данных?
a) Данные только в одном формате
b) Данные в различных форматах и источниках
c) Единая схема данных
d) Лёгкий доступ к любым данным

21. Третье V в концепции больших данных — это разнообразие данных. Называется оно _____.*Ответ:* Variety

22. Одним из первых шагов в жизненном цикле больших данных является сбор данных. Следующий шаг после сбора — это _____.*Ответ:* подготовка данных

23. Самая известная платформа для распределённой обработки больших данных называется _____.*Ответ:* Hadoop
24. Второе V в концепции больших данных связано со скоростью поступления данных и называется _____.*Ответ:* Velocity
25. В характеристиках больших данных третье V (Variety) подчёркивает наличие огромного разнообразия форматов данных. Всего в классических моделях выделяется ____ V.*Ответ:* три

Раздел 2 Обработка и хранение данных

1. Главная цель предварительной обработки данных в большом масштабе:

a) Улучшить читаемость данных

b) Подготовить данные для дальнейшего анализа

c) Увеличить размер набора данных

d) Добавить больше категорий данных

2. Какой инструмент используется для массовой параллельной обработки больших данных в распределённом режиме?

a) MySQL

b) Apache Hadoop

c) PHP

d) Microsoft Office

3. В контексте обработки больших данных важнейшим преимуществом Hadoop является:

a) Быстрая разработка простых скриптов

b) Горизонтальное масштабирование и отказоустойчивость

c) Высокий уровень шифрования данных

d) Централизованное управление всеми процессами

4. Какая система служит для обработки потоков данных в реальном времени?

a) Relational Database Management System (RDBMS)

b) Apache Storm или Apache Flink

c) Local File Storage

d) Email Server

5. Что понимается под концепцией NoSQL?

a) Система управления базами данных, основанная на SQL

b) Альтернативные подходы к хранению данных вне традиционных реляционных схем

c) Запрещённый стандарт баз данных

d) Новый диалект SQL

6. Какая из перечисленных технологий специально разработана для хранения и обработки неструктурированных данных больших размеров?

a) Apache HBase

b) MongoDB

c) PostgreSQL

d) Excel

7. Основной задачей хранилища данных (Data Warehouse) является:

a) Прямая поддержка оперативных транзакций

b) Агрегирование и хранение исторических данных для анализа

c) Непосредственная связь с пользователями

d) Отправка электронной почты пользователям

8. Что означает понятие "масштабируемость" в контексте больших данных?

a) Уменьшение размера данных

b) Возможность увеличить мощность системы с ростом объёма данных

c) Необходимость постоянного мониторинга

d) Сокращение времени обработки данных

9. Какой тип хранилища лучше подходит для хранения документов с иерархическим вложенным содержимым?

a) Табличная база данных

b) Document-oriented database (например, MongoDB)

c) Графовая база данных

d) Key-value storage

10. Какая стратегия используется для увеличения производительности и надежности при обработке больших данных?

a) Монолитная архитектура

b) Распределённая обработка данных

c) Центральный процессор высокого класса

d) Упрощённое представление данных

11. Первый этап в обработке больших данных, связанный с подготовкой данных для анализа, называется _____.

12. Наибольшее распространение среди технологий распределённой обработки больших данных получил фреймворк _____.

13. Среди технологий обработки больших данных наибольшую популярность получила система массового параллельного распределения данных, называемая _____.

14. Технология, предназначенная для обработки потоков данных в реальном времени, называется _____.

15. В контексте больших данных главный признак отличия между реляционными и NoSQL базами данных заключается в отсутствии жесткой предопределённой _____ в последних.

16. Основной инструмент для осуществления операций импорта-экспорт данных и преобразования данных (ETL-процесс) называется _____.

17. Для обработки больших объёмов данных на основе техники map-reduce широко используется инструмент _____.

18. Как называется модель хранения данных, при которой данные размещаются не в строках, а в колонках, повышая эффективность анализа больших данных?

19. Наиболее популярная технология, обеспечивающая массовую параллельную обработку больших данных в распределённой среде, носит название _____.

20. Задача предварительной обработки данных направлена на улучшение их качества посредством очистки, нормализации и _____.

21. Процесс объединения и трансформации данных из разных источников в единый унифицированный формат называется _____.

22. Распространённым инструментом для управления большими данными и анализа в реальном времени является система _____, ориентированная на потоковую обработку.

23. Одно из преимуществ NoSQL баз данных — это их высокая _____ при росте объёма данных.

24. Главный показатель производительности в распределённых системах обработки больших данных — это способность выдерживать увеличение нагрузки благодаря увеличению количества узлов, называемый _____.

25. Чаще всего встречающийся механизм организации данных в big data — это разделение информации на отдельные фрагменты и распределение их по множеству устройств, называемое _____.

2 Выполнение практических работ

Задания для практических работ, теоретический и методический материал для самостоятельного изучения и вопросы для собеседования представлены в методических указаниях к практическим занятиям.

3 Оценивание обучающегося на экзамене

Промежуточная аттестация производится на основе текущего контроля успеваемости при условии выполнения всех заданий в семестре, предусмотренных фондом оценочных средств по дисциплине.