Министерство науки и высшего образования Российской Федерации

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

Кафедра автоматизации и информационных систем

УТВЕРЖДАЮ
Директор института технологий устойчивого развития
_____ А.В. Феоктистов
подпись
«____» _____ 20___ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Технологии сбора и обмена информацией в системах больших данных

09.03.02 «Информационные системы и технологии»
(направленность (профиль): «Экономика и аналитика интеллектуального бизнеса»)

Квалификация выпускника
Бакалавр

Форма обучения
Очная форма

Срок обучения: 4 года

Год начала подготовки 2025

Новокузнецк
2025

СибГИУ
Подлинник электронного документа хранится в системе электронного согласования документов СибГИУ
ДОКУМЕНТ ПОДПИСАН ЭП
Сертификат: 1b8241b8e11f629c673def0838264027810374a8
Владелец: Директор ИЭиМ Феоктистов Андрей Владимирович
Действителен: с 11 сентября 2024 г. по 5 декабря 2025 г.
Дата и время: 05-05-2025 10:58:28 ПРОВЕРИТЬ ПОДПИСЬ

1 Цели и задачи освоения учебной дисциплины

Целями учебной дисциплины являются:
– формирование представлений об основных технологиях и стандартах сбора и обмена информацией в больших данных;
– ознакомление с соответствующим программным обеспечением.

Задачами учебной дисциплины являются:
– формирование систематизированного представления о концепциях, моделях и принципах технологий сбора и обмена информацией в больших данных.

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

Учебная дисциплина относится к учебным дисциплинам части, формируемой участниками образовательных отношений **Блока 1 «Дисциплины (модули)»** ООП по направлению подготовки (специальности) 09.03.02 «Информационные системы и технологии».

Учебная дисциплина базируется на предварительном усвоении обучающимися учебных дисциплин:
– Архитектура вычислительных систем;
– Основы программирования;
– Базы данных.

Учебная дисциплина дополняет знания и умения, получаемые по одновременно изучаемым и последующим учебным дисциплинам:
– Инфокоммуникационные системы и сети;
– Информационная безопасность и защита информации;
– Корпоративные информационные системы.

3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

– Профессиональные компетенции

Наименование категории (группы) ПК	Код и наименование ПК	Код и наименование индикатора достижения ПК	Планируемые результаты обучения
	ПК-4: Способен осуществлять подготовку данных для проведения аналитических работ по исследованию	ПК-4.1 Проводит определение источников больших данных для анализа, идентификации внешних и внутренних источников данных	– знать: виды источников данных: созданные человеком, созданные машинами; источники информации, в том числе информации, необходимой для обеспечения деятельности в

	больших данных		предметной области заказчика исследования; методы извлечения информации и знаний из гетерогенных, мультиструктурированных, неструктурированных источников, в том числе при потоковой обработке. – уметь: определять требования к поставщикам данных из гетерогенных источников; осуществлять взаимодействие с внутренними и внешними поставщиками данных из гетерогенных источников.
		ПК-4.3 Оценивает соответствие набора данных предметной области и задачам аналитических работ	– знать: режимы получения и обработки данных, поддержка режима реального времени. – уметь: оценивать соответствие наборов данных задачам анализа больших данных.

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, лабораторные работы, коллоквиумы), 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***	***6 семестр***
Форма промежуточной аттестации			*экзамен*
Трудоёмкость	*академ. час.*	***144***	*144*
	зачетных единиц	***4***	*4*
Лекции, *академ. час.*		***16***	*16*
в форме практической подготовки		***0***	*0*

Лабораторные работы, *академ. час.*	***16***	*16*
в форме практической подготовки	***0***	*0*
Практические занятия, *академ. час.*	***16***	*16*
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.*	***60***	*60*
в форме практической подготовки	***0***	*0*
Контроль, *академ. час.*	***36***	*36*
в форме практической подготовки	***0***	*0*

Содержание учебной дисциплины

Раздел 1 Технологии сбора и передачи больших данных;

Тема 1.1 Основные понятия больших данных (Терминология, понятия, возможности, виды больших данных);

Тема 1.2 Технология Apache Kafka (Конфигурирование, настройка, работа с очередями Apache Kafka);

Тема 1.3 Технология RabbitMQ (Конфигурирование, настройка, работа с очередями RabbitMQ);

Раздел 2 Технологии хранения и обработки больших данных;

Тема 2.1 Задачи хранения и обработки больших данных. (Способы хранения и обучения больших данных. Фреймворки для хранения больших данных. Файловая система hdfs, rdd);

Тема 2.2 Задача машинного обучения больших данных (Фреймворки хранения и обучения больших данных. Обработка больших данных.);

Тема 2.3 Технология Apache Hadoop (Фреймворк Apache Hadoop. Работа с фреймворком Hadoop);

Тема 2.4 Технология Apache Spark (Фреймворк Apache Spark. Работа с фреймворком Spark.);

Тема 2.5 Обработка данных с использованием Apache Spark (Технологии обработки больших данных).

5 Перечень тем лекций

№ раздела / темы дисциплины	**Темы лекций**	**Трудоемкость, *академ.час***	
		всего	**в форме практической подготовки**
Раздел 1.	Технологии сбора и передачи больших данных		
Тема 1.1.	Основные понятия больших данных	2	
Тема 1.2.	Технология Apache Kafka	2	
Тема 1.3.	Технология RabbitMQ	2	

Раздел 2.	Технологии хранения и обработки больших данных		
Тема 2.1.	Задачи хранения и обработки больших данных.	2	
Тема 2.2.	Задача машинного обучения больших данных	2	
Тема 2.3.	Технология Apache Hadoop	2	
Тема 2.4.	Технология Apache Spark	2	
Тема 2.5.	Обработка данных с использованием Apache Spark	2	
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.	Технологии сбора и передачи больших данных		
Тема 1.1.	Основные понятия больших данных	2	
Тема 1.2.	Технология Apache Kafka	2	
Тема 1.3.	Технология RabbitMQ	2	
Раздел 2.	Технологии хранения и обработки больших данных		
Тема 2.1.	Задачи хранения и обработки больших данных.	2	
Тема 2.2.	Задача машинного обучения больших данных	2	
Тема 2.3.	Технология Apache Hadoop	2	
Тема 2.4.	Технология Apache Spark	2	
Тема 2.5.	Обработка данных с использованием Apache Spark	2	
Итого:		**16**	**0**

7 Перечень тем лабораторных работ

№ раздела / темы дисциплины	**Темы лабораторных работ**	**Трудоемкость, *академ.час***	
		всего	**в форме практической**

			подготовки
Раздел 1.	Технологии сбора и передачи больших данных		
Тема 1.1.	Основные понятия больших данных	2	
Тема 1.2.	Технология Apache Kafka	2	
Тема 1.3.	Технология RabbitMQ	2	
Раздел 2.	Технологии хранения и обработки больших данных		
Тема 2.1.	Задачи хранения и обработки больших данных.	2	
Тема 2.2.	Задача машинного обучения больших данных	2	
Тема 2.3.	Технология Apache Hadoop	2	
Тема 2.4.	Технология Apache Spark	2	
Тема 2.5.	Обработка данных с использованием Apache Spark	2	
Итого:		**16**	**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. Оформление отчета о практической работе; 3. Оформление отчета по лабораторной работе; 4. Подготовка к лабораторной работе; 5. Подготовка к практическому занятию; 6. Прохождение тестирования.	30	

Раздел 2.	1. Изучение лекционного материала; 2. Оформление отчета о практической работе; 3. Оформление отчета по лабораторной работе; 4. Подготовка к лабораторной работе; 5. Подготовка к практическому занятию; 6. Прохождение тестирования.	30	
Контроль	*Подготовка к экзамену*	36	
Итого:		**96**	**0**

10 Матрица компетентностных задач по учебной дисциплине

Контролируемые разделы (темы) дисциплины	Контролируемые компетенции	Оценочные средства
Раздел 1.	ПК-4	1. Вопросы для устного обсуждения (опроса); 2. Наблюдение за работой на лабораторной работе; 3. Наблюдение за работой на практическом занятии; 4. Отчет о лабораторной работе; 5. Отчет о практической работе; 6. Тестовые задания
Раздел 2.	ПК-4	1. Вопросы для устного обсуждения (опроса); 2. Наблюдение за работой на лабораторной работе; 3. Наблюдение за работой на практическом занятии; 4. Отчет о лабораторной работе; 5. Отчет о практической работе; 6. Тестовые задания
Экзамен.	ПК-4	1. Вопросы к экзамену

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 – индивидуальный или фронтальный устный опрос. В ходе проведения входного контроля обучающимся предлагается ответить на вопросы по материалу, изученному в процессе освоения ООП ВО.

Шкала оценивания устных ответов обучающихся

Критерии оценки:

– правильность ответа по содержанию задания (учитывается количество и характер ошибок при ответе);

– полнота и глубина ответа (учитывается количество усвоенных фактов, понятий и т.п.);

– сознательность ответа (учитывается понимание излагаемого материала);

– логика изложения материала (учитывается умение строить целостный, последовательный рассказ, грамотно пользоваться специальной терминологией);

– сопровождение ответа примерами (учитывается умение грамотно и с пользой применять фактический материал для подкрепления теоретического при устном ответе);

– рациональность использования времени, отведенного на ответ (не одобряется затянутость выполнения задания, устного ответа во времени).

Оценку «зачтено» получает обучающийся, если он при собеседовании обнаруживает знание и понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры, в том числе и самостоятельно составленные; при ответе на вопросы излагает материал последовательно и правильно, полно и аргументировано отвечает по содержанию задания; рационально использует время, отведенное на ответ.

Оценку «не зачтено» получает обучающийся, который не рационально использует время, отведенное на ответ; обнаруживает незнание ответа на соответствующий вопрос; допускает ошибки в формулировке определений и правил, искажающие их смысл; беспорядочно и неуверенно излагает материал. Оценка «не зачтено» отмечает такие недостатки в подготовке обучающегося, которые требуют проведения дополнительных индивидуальных консультаций.

Вопросы 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 – вопросы для устного обсуждения (опроса); наблюдение за работой на лабораторной работе; наблюдение за работой на практическом занятии; отчет о лабораторной работе; отчет о практической работе; тестовые задания.

12.2 Описание оценочных средств

Шкалы оценивания (методики оценки)

12.2.1 Рекомендации по оцениванию устных ответов обучающихся

С целью контроля и подготовки обучающихся к изучению новой темы в начале занятия лекционного типа (лекции) педагогическим работником проводится индивидуальный или фронтальный устный опрос по выполненным заданиям предыдущего раздела (темы).

Критерии оценки:

– правильность ответа по содержанию задания (учитывается количество и характер ошибок при ответе);

– полнота и глубина ответа (учитывается количество усвоенных фактов, понятий и т.п.);

– сознательность ответа (учитывается понимание излагаемого материала);

– логика изложения материала (учитывается умение строить целостный, последовательный рассказ, грамотно пользоваться специальной терминологией);

– рациональность использованных приемов и способов решения поставленной учебной задачи (учитывается умение использовать наиболее прогрессивные и эффективные способы достижения цели);

– своевременность и эффективность использования наглядных пособий и технических средств при ответе (учитывается умение грамотно и с пользой применять наглядность и демонстрационный опыт при устном ответе);

– использование дополнительного материала (обязательное условие);

– рациональность использования времени, отведенного на задание (не одобряется затянутость выполнения задания, устного ответа во времени).

Оценка «отлично» ставится, если обучающийся:

– полно и аргументировано отвечает по содержанию задания;

– обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры, в том числе самостоятельно составленные;

– излагает материал последовательно и правильно.

Оценка «хорошо» ставится, если обучающийся дает ответ, удовлетворяющий тем же требованиям, что и для оценки «отлично», но допускает 1 – 2 ошибки, которые сам же исправляет.

Оценка «удовлетворительно» ставится, если обучающийся обнаруживает знание и понимание основных положений данного задания, но:

– излагает материал неполно и допускает неточности в определении понятий или формулировке правил;

– не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры;

– излагает материал непоследовательно и допускает ошибки.

Оценка «неудовлетворительно» ставится, если обучающийся обнаруживает незнание ответа на соответствующее задание, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал. Оценка «неудовлетворительно» отмечает такие недостатки в подготовке обучающегося, которые являются серьезным препятствием к успешному освоению последующего материала.

12.2.2 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

12.2.3 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».

12.2.4 Рекомендации по оцениванию отчета о практической работе

Оформление отчета о проделанной практической работе обучающиеся осуществляют в рамках самостоятельной работы.

Критерии оценки отчета о практической работе:

– правильность выполнения работы (учитывается полнота отчета о практической работе, наличие всех необходимых данных, наличие выводов, наличие ошибок и неточностей);

– аккуратность оформления отчета о практической работе.

Критерии оценки отчета о практической работе

Наименование параметра оценки	Максимальное количество баллов
Отчет о практической работе содержит все необходимые данные и выводы.	2,5
Отчет о практической работе не содержит ошибок и неточностей.	1,25
Отчет о практической работе оформлен аккуратно.	1,25
Итого	**5**

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

12.2.5 Рекомендации по оцениванию в процессе наблюдения за работой на лабораторном занятии

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

Критерии оценки в процессе наблюдения за работой на лабораторном 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 лабораторной работы (учитывается индивидуальная работа и работа в группе).

Критерии оценки по результатам наблюдения за работой на лабораторном 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	1,25

Наименование параметра оценки	Максимальное количество баллов
инструментами и/или средствами вычислительной техники.	
Рационально использует время, отведенное на задание; проявляет активность при выполнении лабораторной работы.	1,25
Итого	**5**

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

12.2.6 Рекомендации по оцениванию отчета о лабораторной работе

Оформление отчета о проделанной лабораторной работе обучающиеся осуществляют в рамках самостоятельной работы.

Критерии оценки отчета о лабораторной работе:

– правильность выполнения работы (учитывается полнота отчета о лабораторной работе, наличие всех необходимых данных, наличие выводов, наличие ошибок и неточностей);

– аккуратность оформления отчета о лабораторной работе.

Критерии оценки отчета о лабораторной работе

Наименование параметра оценки	Максимальное количество баллов
Отчет о лабораторной работе содержит все необходимые данные и выводы.	2,5
Отчет о лабораторной работе не содержит ошибок и неточностей.	1,25
Отчет о лабораторной работе оформлен аккуратно.	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: вопросы к экзамену.

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 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 Тузовский, А. Ф. Проектирование и разработка web-приложений : учебное пособие для вузов / А. Ф. Тузовский. – Москва : Юрайт, 2023. – 218 с. – ISBN 978-5-534-00515-8. – URL: https://urait.ru/bcode/530767 (дата обращения: 26.04.2025);

2 Сысолетин, Е. Г. Разработка интернет-приложений : учебное пособие для вузов / Е. Г. Сысолетин, С. Д. Ростунцев, Л. Г. Доросинский ; под научной редакцией Л. Г. Доросинского. – Москва : Юрайт, 2023. – 90 с. – ISBN 978-5-9916-9975-4. – URL: https://urait.ru/bcode/514303 (дата обращения: 26.04.2025);

3 Кузнецова, Л. В. Лекции по современным веб-технологиям : учебное пособие / Л. В. Кузнецова. – Москва : ИНТУИТ, 2016. – URL: https://www.studentlibrary.ru/book/intuit_131.html (дата обращения: 26.04.2025);

4 Кудряшев, А. В. Введение в современные веб-технологии : учебное пособие / А. В. Кудряшев, П. А. Светашков. – Москва : ИНТУИТ, 2016. – URL: https://www.studentlibrary.ru/book/intuit_085.html (дата обращения: 26.04.2025).

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 Электронная библиотека УМЦ ЖДТ : [коллекция «Эксплуатация железных дорог»] / ФГБУ ДПО «Учебно-методический центр по образованию на железнодорожном транспорте». – Москва, [2013 –]. – URL: https://umczdt.ru/books/. – Режим доступа: для авторизир. пользователей;

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

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

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), оснащенную компьютерной и мультимедийной техникой;
- 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) 09.03.02 «Информационные системы и технологии».

Составитель(и):

доцент Добрынин Алексей Сергеевич (кафедра автоматизации и информационных систем).

Рабочая программа дисциплины рассмотрена и утверждена на заседании кафедры.

Согласована:
Начальник РЦС С/
начальник Регионального центра
сопровождения Сибирь ООО Евразтехника _____ Кириенко Е.В.
должность, степень, звание представителя работодателя — подпись — инициалы, фамилия

Приложение

Тестовые задания для входного контроля знаний обучающихся

Ответьте на один из вопросов:

1 Что такое IP-адрес?

2 Что такое объектно-ориентированное программирование?

3 Что такое HTTP?

4 Как происходит обмен данными между веб-сервером и веб-браузром?

5 Что такое DNS (в контексте сетевых технологий)?

6 Что такое веб-сервер, какие основные технологии (программное обеспечение) для создания веб-серверов вы знаете?

7 Какие языки и технологии разработки веб-приложений вы знаете, в чем их отличие?

9 Какие форматы обмена данными вам известны?

10 Что такое Linux, в чем его отличия от ОС Windows?

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

Раздел 1. Общие понятия сбора и обмена информацией в больших данных.

1 Устный опрос на лекции

Перечень вопросов для устного опроса:

1 Что такое большие данные «big data»?

2 Каковы особенности CAP-теоремы для «big data»?

3 Какие эффекты CAP-теоремы распространены в «big data»?

4 Приведите пример технологий, применяемых в больших данных.

5 Каким образом хранить большие данные?

6 Каким образом обрабатывать большие данные?

7 Преимущества и недостатки больших данных?

8 Системы и технологии сбора больших данных?

9 Системы обмена информацией в больших данных?

10 Методы обработки больших данных.

2 Выполнение практических заданий

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

Раздел 1. Технологии сбора и передачи больших данных.

Тема 1.1 Основные понятия больших данных.

1. Анализ типовых форматов хранения данных. Достоинства и недостатки типовых форматов хранения данных.

2. Текстовый формат хранения данных с разделителем. Формирование и обработка текстовых данных с разделителем.

3. Формирование и обработка данных формата *csv*.
4. Формирование и обработка данных формата *ini*.
5. Формирование и обработка данных формата *XML*.
6. Формирование и обработки данных формата *json*.
7. Бинарные форматы хранения данных.
8. Формирование и обработка конфигурационных файлов.
9. Формирование и обработка данных формата *avro*.
10. Формирование и обработка данных формата *parquet*.

Тема 1.2 Технология Apache Kafka.
1. Настройка Apache Kafka в заданной конфигурации.
2. Конфигурирование Apache Kafka.
3. Работа с очередью сообщений Apache Kafka.
4. Настройка Producers и Consumers в Apache Kafka.
5. Настройка механизмов репликации и контроля целостности Apache Kafka.
6. Использование облачных решений Apache Kafka.
7. Механизмы контейнеризации Apache Kafka.
8. Использования HIVE механизмов работы с данными Apache Kafka.
9. Развертывание Kafka в связке docker+kubernetes.
10. Механизмы горизонального масштабирования и отказоустойчивости Kafka.

Тема 1.3 Технология RabbitMQ.
1. Настройка RabbitMQ в заданной конфигурации.
2. Конфигурирование RabbitMQ.
3. Работа с очередью сообщений RabbitMQ.
4. Настройка Producers и Consumers в RabbitMQ.
5. Настройка механизмов репликации RabbitMQ.
6. Использование облачных решений RabbitMQ.
7. Конвейерная обработка данных RabbitMQ.
8. Использования HIVE механизмов работы с данными RabbitMQ.
9. Развертывание RabbitMQ в связке docker+kubernetes.
10. Механизмы горизонтального масштабирования и отказоустойчивости RabbitMQ.

3 Выполнение лабораторных работ

Задания для лабораторных работ, теоретический и методический материал для самостоятельного изучения и вопросы для собеседования представлены в методических указаниях к лабораторным работам.

Раздел 1. Технологии сбора и передачи больших данных.

Тема 1.1 Основные понятия больших данных.
1. Основы статистического анализа данных.
2. Механизмы подготовки данных к анализу и обработке.

3. Статистический анализ данных формата *csv*.
4. Статистический анализ данных формата *ini*.
5. Статистический анализ данных формата *XML*.
6. Статистический анализ данных формата *json*.
7. Статистический анализ двоичных данных.
8. Статистический анализ конфигурационных файлов.
9. Статистический анализ данных формата *avro*.
10. Статистический анализ данных формата *parquet*.

Тема 1.2 Технология Apache Kafka.
1. Обработка сообщений Kafka.
2. Работа с очередями Kafka с использованием GraphQL.
4. Работа с очередями Kafka с использованием Hive.
5. Использование Kafka в стеке технологий MEAN.
6. Использование Kafka в стеке технологий Core.
7. Разработка Producers & Consumers для Apache Kafka на языке программирования Python.
8. Разработка Producers & Consumers для Apache Kafka на языке программирования C#.
9. Разработка Producers & Consumers для Apache Kafka на языке программирования JavaScript.
10. Работа с несколькими источниками данных Apache Kafka.

Тема 1.3 Технология RabbitMQ.
1. Обработка сообщений RabbitMQ.
2. Работа с очередями RabbitMQ с использованием GraphQL.
4. Работа с очередями RabbitMQ с использованием Hive.
5. Использование RabbitMQ в стеке технологий MEAN.
6. Использование RabbitMQ в стеке технологий Core.
7. Разработка Producers & Consumers для RabbitMQ на языке программирования Python.
8. Разработка Producers & Consumers для RabbitMQ на языке программирования C#.
9. Разработка Producers & Consumers для RabbitMQ на языке программирования JavaScript.
10. Работа с несколькими источниками данных RabbitMQ.

4 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». Примерный перечень вопросов показан ниже. Тест закрытого типа.

1. Отметьте все способы интеграции Apache Kafka и Spark Streaming использование внешней системы, например, Hadoop или Hive, в качестве временного хранилища данных для обмена

а) непосредственная передача данных путем отслеживания смещений в разделах (partition) топика (topic);

б) Кафка через контрольные точки (checkpoints) передача данных через механизм SQL-запросов;

в) получение сообщений через службу синхронизации Zookeeper с помощью высокоуровневого API Kafka;

2.Можно ли на базе Apache Kafka создавать stateful-приложения в микросервисной архитектуре?

а) нет, Kafka предназначена для потоковой обработки событий в рамках одной монолитной системы и не подходит для микросервисной архитектуры;

б) да, эта платформа отлично подходит для хранения состояний микросервисов;

в) нет, возможна реализация только stateless-приложений

г) да, но с подключением внешнего хранилища состояний, например, Hadoop HDFS.

3. Выберите все механизмы обеспечения отказоустойчивости брокеров передачи сообщений (Kafka, RabbitMQ)

а) распределенный характер работы с данными;

б) репликация данных по узлам кластера с синхронизацией реплик

в) разбиение топиков на разделы;

г) возможность восстановить необработанное сообщение по смещению (offset) в разделе топика.

4. Выберите способы повышения производительности Apache Kafka.

а) увеличение коэффициента репликации;

б) ограничение размера сообщений с помощью параметра message.max.bytes;

в) замена HDD-дисков на SSD;

г) увеличение плотности разделов на каждом брокере

5. Отметьте альтернативы Apache Kafka, которые могут заменить эту платформу потоковой обработки событий в высоконагруженных Big Data проектах

а) Apache Storm;

б) Apache Pulsar;

в) Apache Spark;

г) RabbitMQ, почему нет?

6. Может ли Apache Kafka работать без службы Zookeeper?

а) да, Apache Kafka является самодостаточной отказоустойчивой распределенной системой и не нуждается в дополнительных сервисах.

б) нет, ZooKeeper обязателен для любых распределенных систем, т.к. обеспечивает синхронизацию кластерных узлов.

в) да, если используется другой механизм обеспечения надежной согласованности состояния кластера и его конфигурации.

г) нет, ZooKeeper является встроенным механизмом платформы Apache Kafka.

7. С какими форматами данных может работать Apache Kafka?

а) бинарный формат (binary), raw;

б) только JSON и AVRO;

в) только parquet;

г) с любыми в принципе, благодаря механизму работы с разными схемами данных, реализованному в компоненте Schema Registry.

8. где Apache Kafka хранит данные?

а) в распределенной файловой системе Hadoop HDFS;

б) в распределенно-разъединенной системе RDD;

в) нигде, это платформа потоковой обработки событий, а не хранения данных;

г) в разделах своих топиков;

д) во внешней базе данных.

9. Как связать Apache Kafka с внешними источниками данных?

а) через механизм SQL-запросов, реализованный в KSQL;

б) через механизмы HIVE или GraphQL;

в) с помощью команды потокового чтения данных;

г) с помощью коннекторов на базе компонента Connect.

д) Через непосредственную запись сообщений в топики Kafka

10. Можно создавать Kafka-приложения на Python?

а) да, универсальность Kafka позволяет писать приложения на любом языке программирования;

б) нет, Python не подходит для потоковой обработки событий;

в) да, с использованием специальных библиотек, предоставляющих API Python для Kafka;

г) нет, Kafka предоставляет только Java и Scala API.

Раздел 2. Технологии хранения и обработки больших данных.

1 Устный опрос на лекции

Перечень вопросов для устного опроса:

1 Что такое веб сервисы?

2 В чём разница между SOAP и веб-сервисом?

3 Что такое SOAP, SOAP-сервис?

4 Что такое RESTfull-сервис?

5 Что такое XML, XML-RPC?

6 Различия и сферы применения стандартов обмена данными: JSON и XML?

7 Области применимости различных видов веб-сервисов?

8 Что такое брокер сообщений (для сбора и обмена данными)?

9 Что такое Apache Spark?

10 Что такое Apache Kafka?

11 Опишите (в целом) организацию загрузки/выгрузки информации в системах Apache Spark и Apache Kafka?

12 Опишите способы организации асинхронного взаимодействия нескольких вычислительных задач в системах Apache Spark и Apache Kafka?

2 Выполнение практических заданий

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

Тема 2.1. Задачи хранения и обработки больших данных.

а) Установка Python – инфраструктуры на целевой машине (Anaconda, SpiderX, Jupiter Notebook, VS Code).

б) Настройка Python – инфраструктуры на целевой машине (Anaconda, SpiderX, Jupiter Notebook, VS Code).

в) Установка фреймворка Apache Hadoop в связке (внутри) со средой Jupiter Notebook;

г) Установка фреймворка Apache Spark в связке (внутри) со средой Jupiter Notebook.

д) Тестирование фреймворков работы с большими данными (big data).

Тема 2.2. Задача машинного обучения больших данных

а) Решение задачи классификации цветов Ириса с использованием библиотеки Pandas.

б) Решение задачи классификации цветов Ириса с использованием mpp – фреймворка Apache Hadoop.

в) Решение задачи классификации цветов Ириса с использованием mpp – фреймворка Apache Spark.

г) Решение задачи классификации пассажиров Титаника с использованием mpp – фреймворка Apache Hadoop.

д) Решение задачи классификации пассажиров Титаника с использованием mpp – фреймворка Apache Spark.

Тема 2.3. Технология Apache Hadoop.

а) Изучение основных коллекций Apache Hadoop;

б) Технологии работы с коллекциями Apache Hadoop;

в) Использование механизма MapReduce в Apache Hadoop.

г) Использование механизма parallelize в Apache Hadoop.

д) Изучение типовых задач обработки больших данных.

Тема 2.4. Технология Apache Spark.

а) Изучение основных коллекций Apache Spark;

б) Технологии работы с коллекциями Apache Spark;

в) Использование механизма MapReduce в Apache Spark.

г) Использование механизма parallelize в Apache Spark.

д) Изучение типовых задач обработки больших данных.

Тема 2.5. Обработка данных с использованием Apache Spark.

а) Управление задачами, распределение данных, планирование в Apache Spark.

3 Выполнение лабораторных работ

Задания для лабораторных работ, теоретический и методический материал для самостоятельного изучения и вопросы для собеседования представлены в методических указаниях к лабораторным работам.

Тема 2.1. Задачи хранения и обработки больших данных.

а) Выполните потоковую mpp – обработку больших данных (big data) с использованием библиотеки pandas;

б) Выполните потоковую mpp – обработку больших данных (big data) с использованием фреймворка hadoop ;

в) Выполните потоковую mpp – обработку больших данных (big data) с использованием фреймворка spark;

г) Выполните потоковую mpp – обработку коллекции данных HDFS;

д) Выполните потоковую mpp – обработку коллекции данных RDD.

Тема 2.2. Задача машинного обучения больших данных

а) Решение задачи машинного обучения с помощью фреймворка обработки больших данных.

б) Задача предварительной загрузки и обработки датасета kaggle в систему Apache Hadoop.

в) Задача предварительной загрузки и обработки датасета kaggle в систему Apache Spark

г) Решение произвольной задачи машинного обучения на основе датасета kaggle в среде обработки больших данных Apache Hadoop;

г) Решение произвольной задачи машинного обучения на основе датасета kaggle в среде обработки больших данных Apache Spark;

Тема 2.3. Технология Apache Hadoop.

а) Решение задачи статистической обработки данных с использованием Apache Hadoop.

б) Решение задачи глубокого машинного обучение (deep learning) с использованием Apache Hadoop на примере задачи многомерной регрессии.

Тема 2.4. Технология Apache Spark.

а) Решение задачи статистической обработки данных с использованием Apache Spark.

б) Решение задачи глубокого машинного обучение (deep learning) с использованием Apache Spark на примере задачи многомерной регрессии.

Тема 2.5. Обработка данных с использованием Apache Spark.

а) Механизмы автоматизации обработки больших данных (big data) в Apache Spark.

4 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». Примерный перечень вопросов показан ниже. Тест закрытого типа.

1. R не является…

а) языком программирования для статистической обработки данных и работы с графикой, созданный на основе языка S;

б) свободной программной средой вычислений с открытым исходным кодом;

в) высокоуровневым языком программирования общего назначения, ориентированным на повышение производительности разработчика и читаемости кода;

г) языком программирования с динамической типизацией данных.

2. У машинного обучения есть ряд задач. Как называется та, что направлена на предсказание значения той или иной непрерывной числовой величины для входных данных?

а) Кластеризация;

б) Классификация;

в) Переобучение;

г) Регрессия.

3. Какое API было добавлено в Hadoop v2.0?

а) YAWN;

б) YARN;

в) SARN;

г) DARN.

4. Отказоустойчивость Big Data – это, когда…

а) активируются до 1000 компьютеров;

б) сбой в одном звене системы не ведет к сбоям в других звеньях;

в) недостоверные данные удаляются из системы;

г) данные обрабатываются на других серверах.

5. Итак, вы решили работать с большими данными. Какой из этих инструментов вам вряд ли пригодится?

а) SQL;

б) Texmaker;
в) Python;
г) R;
6. Сколько Петабайт в Зеттабайте?
а) 1024;
б) 128;
в) 32;
г) 4;
7. В краудсорсинге вклад в проект каждого участника осуществляется…
а) лишь финансами;
б) лишь затратами времени;
в) ресурсами – компьютер, время и др.;
г) привлечением аутсорсинга.
8. Какая компания создала технологию MapReduce?
а) Google
б) Yahoo
в) EMC
г) Oracle
9. Одна из главных целей Data Mining – это получение…
а) связей малых выборок данных, распространение их на большие массивы;
б) скриншотов всех обработок данных;
в) аудита сайтов, веб-ресурсов;
г) гарантированной сетевой безопасности;
10. Отметьте верное понимание Variety в контексте характеристик Big Data…
а) высокая скорость генерирования данных;
б) разные типы данных в колонках таблиц реляционных СУБД;
в) разнообразие отраслей, являющихся источниками данных;
г) разнообразие типов данных, включающих в себя структурированные, полуструктурированные и неструктурированные;
11. Дайте определение Big Data
а) Комплексный набор инструментов обработки структурированных данных колоссальных объемов;
б) Комплексный набор подходов, инструментов и методов обработки структурированных и неструктурированных данных колоссальных объемов;
в) Комплексный набор методов обработки структурированных данных колоссальных объемов;
12. Эффективно применять Data Mining с целью защиты от…
а) мошенников;
б) инфицирующих ботов;
в) инфицированных клиентов;
г) инсайдеров;

13. Как происходит обучение нейронной сети?

а) эксперты настраивают нейронную сеть;

б) сеть запускается на обучающем множестве, и незадействованные нейроны выкидываются;

в) сеть запускается на обучающем множестве, и подстраиваются весовые значения;

г) сеть запускается на обучающем множестве, и добавляются или убираются соединения между нейронами.

14. До появления Data Mining невозможно было найти связи…

а) студентов вуза и их успеваемости по предмету;

б) течения COVID-19 и осложнений после заболевания по всей РФ;

в) рекламирования и покупательской активности на сайте веб-магазина;

г) сроков доставки заказа через службу доставки и числом курьеров.

15. Примером применения Big Data не может быть…

а) зондирование Земли из космоса;

б) родительский контроль;

в) хранение данных клиентов в соцсетях;

г) доступ к крупнейшим библиотекам.

16. Что такое SQL?

а) Реляционная база данных;

б) Язык неструктурированных запросов;

в) Язык структурированных запросов;

г) Средство для создания пайплайнов;

17. Одна из главных целей Big Data – это…

а) привлечение все больше пользователей;

б) рост числа обработок данных;

в) снижение издержек операций;

г) таргетирование пользователей;

18. Закончите следующее предложение: С точки зрения машины, информация становится структурированной, если...

а) Машина проинструктирована, каким образом её обрабатывать;

б) Информация разделена на части и озаглавлена;

в) Информация имеет логическую взаимосвязь внутри себя;

г) Машина знает из каких частей состоит информация;

19. Отметьте причину создания NoSQL баз данных…

а) высокая стоимость горизонтальной масштабируемости RDBMS при сохранении требования высокой доступности;

б) недостаточная гибкость языка запросов SQL;

в) невозможность хранить большие объёмы;

г) данных дороговизна лицензий RDBMS;

20. Искусственные нейронные сети (ИНС) — модели машинного обучения, использующие комбинации распределенных простых

операций, зависящих от обучаемых параметров, для обработки входных данных. Какого вида ИНС не существует?

а) Наивные;

б) Рекуррентные;

в) Импульсные;

г) Противоборствующие;

6. Что не является целью процесса Business Intelligence?

а) Интерпретация большого количества данных;

б) Моделирование исходов различных вариантов действий;

в) Модификация существующего программного обеспечения;

г) Отслеживание результатов решений;

21. Что из перечисленного помогает следить за эволюцией документа, над созданием которого работает одновременно большое количество авторов?

а) Пространственный поток;

б) Исторический поток;

в) Визуальный поток;

г) Интерактивный поток;

22. Выберите неверное высказывание…

а) большие данные – это данные объёма свыше 1 Тб;

б) проблема больших данных – это проблема, когда при существующих технологиях хранения и обработки сущностная обработка данных затруднена или невозможна;

в) большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров;

г) большие данные как правило не структурированы;

23. Какие модули по умолчанию yt входят в состав проекта Apache Hadoop?

а) HDFS;

б) Spark;

в) MapReduce;

г) YARN;

24. Hadoop – это…

а) набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах;

б) распределённая СУБД, позволяющая обрабатывать большие данные;

в) язык выполнения заданий в парадигме MapReduce;

г) распределённая файловая система, предназначенная для хранения файлов большого объёма;

25. В каком из приведённых примеров наиболее эффективны NoSQL решения типа ключ-значение?

а) потоковая обработка логов кластера серверов и быстрого сохранения без требования оперативной аналитики;

б) оперативная аналитика сохранённых логов кластера серверов;

в) кластеризация логов кластера серверов на основе заранее известных признаков лог-файла хранение данных о клиентах международной;

26. Отметьте те из вариантов, в которых данные структурированы…

а) данные о продажах компании, представленные в виде помесячных отчётов в формате MS Word;

б) таблица с ежедневными показаниями температуры помещения за год в файле формата csv;

в) текст педагогической поэмы А.С. Макаренко, представленный в формате PDF;

г) библиотека фильмов, представленных в формате mpeg4 на одном жестком диске;

27. Какие из задач решаются Big Data?

а) Мониторинг оборудования;

б) Анализ социальных сетей;

в) Оптимизация автомобильного движения;

г) Все вышеперечисленное;

28. Расписание движения поездов может рассматриваться как пример…

а) Табличной модели;

б) Натурной модели;

в) Математической модели;

г) Графической модели;

29. Краудсорсинг большого проекта невозможен без…

а) соединения его с другими проектами, композиции;

б) выделения его частей, декомпозиции;

в) экспертной оценки;

г) финансирования крупным банком;

Тест открытого типа.

1 Характеристика, которая отвечает за размер данных, называется _____

2 Характеристика, которая отвечает за достоверность данных, называется _____

3 Данные и информация это одно и то же (да или нет)? _____.

4 Лидером в области big data является компания _____.

5 Как можно коротко описать термин «Data Mining» _____.

6 Библиотека Python, для работы с большими данными, называется _____

7 Библиотека визуализации данных в языке Python, называется _____

8 Библиотека работы с большими числами в языке Python, называется _____.

9 Как называется алгоритм, который используется для решения задач прогнозирования? _____

10 Как называется алгоритм, который используется для решения задач классификации? _____

5 Вопросы к экзамену для промежуточной аттестации

1 Что такое веб сервисы?
2 В чём разница между SOAP и веб-сервисом?
3 Что такое SOAP, SOAP-сервис?
4 Что такое RESTfull-сервис?
5 Что такое XML, XML-RPC?
6 Различия и сферы применения стандартов обмена данными: JSON и XML?
7 Области применимости различных видов веб-сервисов?
8 Что такое брокер сообщений (для сбора и обмена данными)?
9 Что такое Apache Spark?
10 Что такое Apache Kafka?
11 Опишите (в целом) организацию загрузки/выгрузки информации в системах Apache Spark и Apache Kafka?
12 Опишите способы организации асинхронного взаимодействия нескольких вычислительных задач в системах Apache Spark и Apache Kafka?
13 Дайте определение термину «Big Data».
14 Опишите инструменты, которые используются для работы с «Big Data»
15 Системы хранения и обработки больших данных.
16 Задача классификации. Постановка задачи. Приведите примеры.
17 Задача регрессии. Постановка задачи. Приведите примеры.
18 Задача кластеризации. Постановка задачи. Приведите примеры.
19 Поясните суть технологии Map Reduce.
20 Поясните суть технологии Hadoop.
21 Поясните суть технологии Apache Sparc.
22 Поясните значение термина «машинное обучение».
23 Какие алгоритмы машинного обучения используются.
24 В чем суть задачи машинного обучения.
25 Понятие искусственной нейронной сети.
26 Типовые структуры искусственной нейронной сети.
27 Выборка данных. Понятия выборки.
28 Требования к выборке данных
29 Назначение выборки данных.
30 Обучение с учителем и без учителя.
31 Алгоритмы обучения искусственных нейронных сетей.
32 Формирование больших данных.
33 Требования к большим данным.
34 Технологии обработки больших данных.

35 Технологии хранения больших данных.
36 Технологии распределенных вычислений.
37 Технологии облачных вычислений.
38 Поясните смысл и назначение коллекций в Apache Hadoop.
39 Поясните смысл и назначение коллекций в Apache Spark.
40 поясните смысл технологии MapReduce.