МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Рязанский государственный радиотехнический университет»

Кафедра «Безопасность жизнедеятельности и экология»

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

по дисциплине

«Продвинутые технологии обработки больших данных»

Направление

09.03.04 Программная инженерия

Профиль

Программная инженерия систем искусственного интеллекта
Программное обеспечение систем искусственного интеллекта

Квалификация выпускника – бакалавр

Форма обучения – очная

Рязань

1. ОБЩИЕ ПОЛОЖЕНИЯ

Оценочные материалы – это совокупность учебно-методических материалов и процедур, предназначенных для оценки качества освоения обучающимися данной дисциплины как части основной образовательной программы.

Цель – оценить соответствие знаний, умений и уровня приобретённых компетенций обучающихся целям и требованиям основной образовательной программы в ходе текущего контроля и промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности компетенций и индикаторов их достижения, приобретаемых обучающимися в соответствии с требованиями ФГОС ВО.

Контроль знаний обучающихся проводится в форме текущего контроля (лабораторные работы) и промежуточной аттестации (зачёт).

2. ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ С УКАЗАНИЕМ ЭТАПОВ ИХ ФОРМИРОВАНИЯ

При освоении дисциплины формируются следующие компетенции:

- **ПК-9** – способен организовывать хранение данных, выбирая адекватные технологические решения (BD-3);
  - **ПК-9.1** – разрабатывает, отлаживает и тестирует прикладные решения с элементами ИИ с применением различных технологий хранения структурированных данных, оценивает качество построенных прикладных решений (BD-3.1);
  - **ПК-9.2** – разрабатывает, отлаживает и тестирует прикладные решения с элементами ИИ с применением различных технологий хранения неструктурированных данных, оценивает качество (BD-3.2).
- **ПК-10** – способен применять различные модели и (или) технологии обработки данных (BD-4);
  - **ПК-10.1** – осуществляет выбор технологий обработки больших данных, приемлемых для создания прикладной системы ИИ с заданными требованиями (BD-4.1);
  - **ПК-10.2** – разрабатывает и отлаживает прикладные решения с элементами ИИ с применением различных технологий обработки данных (BD-4.2);
  - **ПК-10.3** – тестирует, испытывает и оценивает качество решений с элементами ИИ, реализованных с использованием технологий обработки данных (BD-4.3).
- **ПК-16** – способен применять JVM-совместимые языки программирования для решения задач в области ИИ (PL-2);
  - **ПК-16.2** – разрабатывает и поддерживает системы обработки больших данных различной степени сложности (PL-2.2).

- **ПК-25** – способен применять технологии организации инфраструктуры больших данных (BD-5);
  - **ПК-25.1** – осуществляет выбор направления вспомогательных технологических решений для формирования единого стека работы с большими данными для решения поставленной задачи (BD-5.1);
  - **ПК-25.2** – разрабатывает и отлаживает прикладные решения с элементами ИИ с применением различных технологий организации инфраструктуры БД (BD-5.2).

Указанные компетенции формируются в соответствии со следующими этапами:

- формирование и развитие теоретических знаний (лекционные занятия, самостоятельная работа);
- приобретение и развитие практических умений (лабораторные работы, самостоятельная работа);
- закрепление знаний и навыков (выполнение и защита лабораторных работ, подготовка к зачёту).

3. ПОКАЗАТЕЛИ И КРИТЕРИИ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ

Сформированность каждой компетенции оценивается по трёхуровневой шкале:

- **пороговый уровень** – обязательный для всех обучающихся; соответствует освоению более 40% приведённых знаний, умений и навыков;
- **продвинутый уровень** – соответствует освоению более 60% знаний, умений и навыков;
- **эталонный уровень** – соответствует освоению более 80% знаний, умений и навыков.

При освоении менее 40% компетенция считается несформированной.

Таблица 1. Показатели достижения индикаторов компетенций

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
ПК-9.1	Знать основные принципы организации реляционных систем хранения данных, принципы организации распределённых реляционных систем хранения данных. Уметь создавать базы данных в реляционных СУБД, заполнять данными реляционные хранилища, писать запросы на языке SQL. Владеть навыками создания	8	Зачёт

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
	решений распределённых хранилищ под конкретную прикладную задачу.		
ПК-9.2	Знать основные технологии NoSQL, популярные хранилища класса ключ-значение, документные, колоночные и графовые базы данных. Уметь выбирать технологии и механизмы хранения и доступа к большим данным. Владеть основными командами работы с хранилищами класса ключ-значение, документными, колоночными и графовыми базами данных.	8	Зачёт
ПК-10.1	Знать принципы организации распределённых хранилищ данных для различных задач. Уметь выбирать технологии обработки больших данных для решения поставленной задачи. Владеть пониманием выбора технологий Grid, MapReduce, Hadoop, Spark.	8	Зачёт
ПК-10.2	Знать принципы организации параллельной обработки хранилищ на базе современных технологий больших данных. Уметь разрабатывать решения обработки больших данных, приемлемых для создания системы по заданным требованиям. Владеть навыками создания решений на базе параллельной, потоковой и пакетной обработки данных.	8	Зачёт
ПК-10.3	Знать принципы тестирования и оценивания качества решений параллельной, потоковой и пакетной	8	Зачёт

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
	обработки больших данных. Уметь тестировать и оценивать качество решений обработки больших данных. Владеть навыками тестирования и оценивания качества решений обработки больших данных.		
ПК-16.2	Знать фреймворки для пакетной обработки данных (Spark, MapReduce), инструменты организации потоковых вычислений (Kafka, Flink), аналитические БД (ClickHouse, Greenplum). Уметь адаптировать алгоритмы для вычислений на малых объёмах данных под большие данные, разрабатывать и поддерживать ETL-алгоритмы в пайплайнах обработки данных. Владеть принципами использования фреймворков для пакетной обработки данных, инструментами организации потоковых вычислений, навыками использования аналитических баз данных.	8	Зачёт
ПК-25.1	Знать методологию создания инфраструктуры больших данных. Уметь выполнять отдельные функции в проектах по созданию инфраструктуры БД. Владеть навыками выбора технологических решений для формирования единого стека работы с большими данными.	8	Зачёт
ПК-25.2	Знать компоненты инфраструктуры больших данных. Уметь применять прикладные решения с элементами ИИ и различными технологиями организации инфраструктуры больших данных. Владеть	8	Зачёт

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
	навыками участия в разработке решений с элементами ИИ с применением различных технологий организации инфраструктуры больших данных.		

4. ПАСПОРТ ОЦЕНОЧНЫХ МАТЕРИАЛОВ ПО ДИСЦИПЛИНЕ

Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или её части)	Наименование оценочного средства
Тема 1. Современный стек Big Data для AI	ПК-9.1, ПК-9.2, ПК-10.3, ПК-16.2, ПК-25.1, ПК-25.2	Зачёт
Тема 2. Apache Spark: углублённое изучение	ПК-9.1, ПК-9.2, ПК-10.3, ПК-16.2, ПК-25.1, ПК-25.2	Зачёт
Тема 3. Обработка потоковых данных	ПК-9.1, ПК-9.2, ПК-10.3, ПК-16.2, ПК-25.1, ПК-25.2	Зачёт
Тема 4. Хранение данных: Data Warehouse (DWH), Data Lake и Data Lakehouse	ПК-9.1, ПК-9.2, ПК-10.3, ПК-16.2, ПК-25.1, ПК-25.2	Зачёт
Тема 5. Оркестрация конвейеров данных	ПК-9.1, ПК-9.2, ПК-10.3, ПК-16.2, ПК-25.1, ПК-25.2	Зачёт
Тема 6. Big Data и MLOps	ПК-9.1, ПК-9.2, ПК-10.3, ПК-16.2, ПК-25.1, ПК-25.2	Зачёт
Лабораторный практикум (6 лабораторных работ)	ПК-9.1, ПК-9.2, ПК-10.1, ПК-10.2, ПК-10.3, ПК-16.2, ПК-25.1, ПК-25.2	Защита лабораторных работ

5. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ

5.1. Промежуточная аттестация (зачёт)

Зачёт проводится в форме устного ответа по билетам, включающим теоретический вопрос и практическое задание (задачу). Также может использоваться тестирование.

5.1.1. Тестовые вопросы закрытого типа (20 вопросов)

1. Что такое Apache Spark?
а) Фреймворк для распределённой обработки больших данных;
б) Реляционная СУБД;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

2. Какие компоненты входят в архитектуру Apache Spark?
а) Spark Core, Spark SQL, Structured Streaming, MLlib;
б) Только Spark Core;
в) Только Spark SQL;
г) Только MLlib.
Правильный ответ: а)

3. Что такое Data Lake?
а) Хранилище для сырых, неструктурированных данных;
б) Реляционная база данных;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

4. Что такое Data Warehouse?
а) Централизованное хранилище структурированных, очищенных данных для аналитики;
б) Хранилище сырых данных;
в) Операционная система;
г) Язык программирования.
Правильный ответ: а)

5. Что такое Data Lakehouse?
а) Гибридная архитектура, сочетающая преимущества Data Lake и Data Warehouse;
б) Только Data Lake;
в) Только Data Warehouse;
г) Реляционная база данных.
Правильный ответ: а)

6. Что такое Apache Hudi?
а) Технология для управления данными в Data Lake;
б) Реляционная СУБД;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

7. Что такое Delta Lake?
а) Технология хранения данных с ACID-транзакциями для Data Lake;
б) Реляционная СУБД;
в) Операционная система;
г) Язык программирования.
Правильный ответ: а)

8. Что такое Apache Airflow?
а) Инструмент для оркестрации конвейеров данных;
б) Реляционная СУБД;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

9. Что такое DAG в Apache Airflow?
а) Направленный ациклический граф задач;
б) Тип базы данных;
в) Метод оптимизации;
г) Алгоритм шифрования.
Правильный ответ: а)

10. Что такое потоковая обработка данных (stream processing)?
а) Обработка данных в реальном времени по мере их поступления;
б) Пакетная обработка данных;
в) Хранение данных;
г) Шифрование данных.
Правильный ответ: а)

11. Что такое оконные агрегации (window aggregations)?
а) Агрегация данных в пределах определённого временного окна;
б) Тип базы данных;
в) Метод оптимизации;
г) Алгоритм шифрования.
Правильный ответ: а)

12. Что такое ETL?
а) Extract, Transform, Load — процесс извлечения, преобразования и загрузки данных;
б) Язык программирования;
в) Сетевой протокол;

г) Тип базы данных.
Правильный ответ: а)

13. Что такое MLOps?
а) Практика управления жизненным циклом ML-моделей;
б) Тип базы данных;
в) Язык программирования;
г) Сетевой протокол.
Правильный ответ: а)

14. Что такое Feature Store?
а) Хранилище признаков для ML-моделей;
б) Реляционная СУБД;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

15. Что такое ClickHouse?
а) Аналитическая колоночная СУБД;
б) Реляционная СУБД;
в) Операционная система;
г) Язык программирования.
Правильный ответ: а)

16. Что такое Greenplum?
а) Аналитическая база данных на основе PostgreSQL;
б) Реляционная СУБД;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

17. Что такое Apache Kafka?
а) Распределённый брокер сообщений для потоковой обработки данных;
б) Реляционная СУБД;
в) Операционная система;
г) Язык программирования.
Правильный ответ: а)

18. Что такое Apache Flink?
а) Фреймворк для потоковой обработки данных;
б) Реляционная СУБД;
в) Операционная система;
г) Сетевой протокол.
Правильный ответ: а)

19. Что такое Grid-вычисления?
а) Распределённые вычисления с использованием множества компьютеров;

б) Тип базы данных;
в) Метод оптимизации;
г) Алгоритм шифрования.
Правильный ответ: а)

20. Что такое MapReduce?
а) Модель параллельных вычислений для обработки больших данных;
б) Тип базы данных;
в) Сетевой протокол;
г) Алгоритм шифрования.
Правильный ответ: а)

5.1.2. Тестовые вопросы открытого типа (10 вопросов)

1. Опишите современный стек технологий Big Data для задач искусственного интеллекта.

2. Объясните архитектуру Apache Spark и его основные компоненты (Spark Core, Spark SQL, Structured Streaming, MLlib).

3. Сравните пакетную и потоковую обработку данных. Приведите примеры использования.

4. Опишите различия между Data Warehouse, Data Lake и Data Lakehouse.

5. Что такое ACID-транзакции в контексте хранилищ данных? Приведите примеры (Delta Lake, Apache Hudi).

6. Опишите процесс оркестрации конвейеров данных с использованием Apache Airflow. Что такое DAG?

7. Что такое Feature Store? Опишите его роль в MLOps-подходе.

8. Опишите методы оптимизации ETL-задач в Apache Spark.

9. Сравните Apache Kafka и Apache Flink по назначению и областям применения.

10. Какие практики мониторинга и управления Big Data-процессами вы знаете? Приведите примеры инструментов.

5.1.3. Типовые практические задания (задачи) к зачёту

1. Разработать ETL-конвейер с использованием Apache Spark для заданного набора данных.
2. Оптимизировать ETL-задачу в Spark (настроить параметры, выбрать оптимальные операторы).
3. Разработать приложение для потоковой агрегации данных с использованием Apache Spark Structured Streaming.

4. Реализовать работу с ACID-транзакциями в Delta Lake.
5. Построить DAG для ML-конвейера в Apache Airflow.
6. Интегрировать Feature Store в конвейер данных.
7. Разработать итоговый проект по выбору студента (на основе кейсов индустриальных партнёров).
8. Настроить и развернуть распределённый кластер для обработки данных.
9. Выполнить мониторинг и профилирование распределённого приложения.
10. Спроектировать архитектуру хранения данных для аналитических и ML-задач.

5.1.4. Вопросы к зачёту (30 вопросов)

1. Современный стек Big Data для AI. Основные компоненты и технологии.
2. Архитектура Apache Spark: Spark Core, Spark SQL, Structured Streaming, MLlib.
3. Углублённое изучение Apache Spark: RDD, DataFrame, Dataset.
4. Оптимизация ETL-задач в Spark. Основные методы и приёмы.
5. Потоковая обработка данных: принципы и инструменты (Kafka, Flink).
6. Оконные агрегации в потоковой обработке.
7. Data Warehouse (DWH): назначение, архитектура, особенности.
8. Data Lake: назначение, архитектура, особенности.
9. Data Lakehouse: гибридная архитектура, преимущества.
10. Технологии хранения данных с ACID-транзакциями (Delta Lake, Apache Hudi).
11. Оркестрация конвейеров данных: Apache Airflow, Prefect.
12. DAG в Apache Airflow: построение, выполнение, мониторинг.
13. Big Data и MLOps: интеграция, практики, инструменты.
14. Feature Store: назначение, архитектура, применение.
15. Аналитические базы данных: ClickHouse, Greenplum.
16. Пакетная обработка данных: MapReduce, Hadoop, Spark.
17. Потоковая обработка данных: Kafka, Flink.
18. Grid-вычисления: принципы и применение.
19. Оптимизация производительности Big Data-приложений.
20. Развёртывание и настройка распределённых кластеров.
21. Облачные сервисы для Big Data (Google BigQuery, Amazon Redshift).
22. Интеграция Big Data-технологий с системами машинного обучения.
23. Мониторинг и управление Big Data-процессами.
24. Методы отладки и профилирования распределённых приложений.
25. Инструменты для оркестрации сложных, зависимых задач обработки данных.
26. Выбор технологий обработки больших данных под конкретную задачу.
27. Принципы тестирования и оценивания качества решений обработки больших данных.
28. Инфраструктура больших данных: компоненты и методология создания.
29. Выбор технологических решений для формирования единого стека работы с большими данными.
30. Перспективные направления развития технологий обработки больших данных.

6. КРИТЕРИИ ОЦЕНИВАНИЯ

6.1. Шкала оценивания тестирования

Шкала	Критерий
3 балла (эталонный уровень)	Процент верных ответов от 85 до 100%
2 балла (продвинутый уровень)	Процент верных ответов от 75 до 84%
1 балл (пороговый уровень)	Процент верных ответов от 60 до 74%
0 баллов	Процент верных ответов от 0 до 59%

6.2. Шкала оценивания теоретического вопроса

Шкала	Критерий
3 балла (эталонный уровень)	Дан полный ответ, показаны глубокие систематизированные знания, приведены примеры, даны ответы на дополнительные вопросы.
2 балла (продвинутый уровень)	Дан полный ответ, но на некоторые дополнительные вопросы ответы даны с помощью наводящих вопросов.
1 балл (пороговый уровень)	Дан неполный ответ, на дополнительные вопросы ответы даны только с помощью преподавателя.
0 баллов	Студент не смог ответить на вопрос.

6.3. Шкала оценивания практического задания

Шкала	Критерий
3 балла (эталонный уровень)	Задание решено верно, получены корректные результаты, сделаны выводы.
2 балла (продвинутый уровень)	Задание решено верно, но имеются незначительные технические неточности.

Шкала	Критерий
1 балл (пороговый уровень)	Задание решено верно с дополнительными наводящими вопросами преподавателя.
0 баллов	Задание не решено.

6.4. Итоговая оценка

На зачёт выносится тестовое задание, практическое задание и теоретический вопрос.
Максимальный балл – 9.
Оценка «зачтено» выставляется при сумме баллов не менее 5 и выполнении всех предусмотренных лабораторных работ.
Оценка «не зачтено» – при сумме менее 5 баллов или наличии несданных лабораторных работ.

Оператор ЭДО ООО "Компания "Тензор"
ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
СОГЛАСОВАНО **ФГБОУ ВО "РГРТУ", РГРТУ,** Овечкин Геннадий Владимирович, Заведующий кафедрой ВПМ **05.11.26** 14:46 (MSK) Простая подпись