МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ
В.Ф. УТКИНА»

Кафедра «Вычислительная и прикладная математика»

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ

«Продвинутые технологии обработки больших данных»

Направление подготовки – 09.03.04 Программная инженерия

ОПОП «Программная инженерия систем искусственного интеллекта»

Квалификация (степень) выпускника – бакалавр

Форма обучения – очная (4 года)

Рязань

1. СПИСОК ТЕОРЕТИЧЕСКИХ ВОПРОСОВ К ЗАЧЁТУ

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.

2. ПЕРЕЧЕНЬ ПРАКТИЧЕСКИХ И ЛАБОРАТОРНЫХ РАБОТ ПО ДИСЦИПЛИНЕ

№	Тема лабораторной работы	Кол-во часов
1	Оптимизация ETL-задач в Spark	4
2	Разработка приложения для потоковой агрегации	4
3	Работа с ACID-транзакциями в хранилищах данных (Delta Lake, Data Lakehouse)	4
4	Построение DAG для ML-конвейера в Apache Airflow	4
5	Интеграция Feature Store в конвейер данных	4
6	Итоговый проект по выбору студента (выбирается на основе кейсов индустриальных партнёров)	4
Итого		**24**

3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Самостоятельная работа как вид учебной работы может использоваться на лабораторных занятиях, а также иметь самостоятельное значение – внеаудиторная самостоятельная работа обучающихся – при подготовке к лабораторным работам и при подготовке к зачёту.

Самостоятельная работа включает в себя следующие этапы:

- изучение теоретического материала (работа над конспектом лекции);
- самостоятельное изучение дополнительных информационных ресурсов (доработка конспекта лекции);
- выполнение заданий текущего контроля успеваемости (подготовка к лабораторным работам);
- итоговая аттестация по дисциплине (подготовка к зачёту).

При подготовке к лабораторным работам необходимо:

1. Изучить конспект лекций и методические указания к данной работе.
2. Ознакомиться с рекомендуемой основной и дополнительной литературой.
3. Выполнить индивидуальное задание.
4. Оформить отчёт в соответствии с требованиями кафедры.
5. Защитить работу, ответив на вопросы преподавателя.

Важным этапом является защита лабораторной работы. В процессе защиты студент отвечает на вопросы преподавателя, касающиеся теоретического материала, относящегося к данной работе, и проекта, реализующего её задание, комментирует полученные результаты. При подготовке к защите рекомендуется ознакомиться со списком вопросов по изучаемой теме и попытаться самостоятельно на них ответить, используя конспект лекций и рекомендуемую литературу.

4. РЕКОМЕНДАЦИИ ПО ПЛАНИРОВАНИЮ И ОРГАНИЗАЦИИ ВРЕМЕНИ, НЕОБХОДИМОГО ДЛЯ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ. ОПИСАНИЕ ПОСЛЕДОВАТЕЛЬНОСТИ ДЕЙСТВИЙ СТУДЕНТА («СЦЕНАРИЙ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ»)

Рекомендуется следующим образом организовать время, необходимое для изучения дисциплины.

Для освоения лекционного материала следует: изучить конспект лекции в тот же день, после лекции: 10–15 минут, повторно прочитать конспект лекции за день перед следующей лекцией: 10–15 минут. Также следует изучить теоретический лекционный материал по рекомендуемому учебнику/учебному пособию: 1 час в неделю.

Следует максимально использовать лекционное время для изучения дисциплины, понимания лекционного материала и написания конспекта лекций. В процессе лекционного занятия студент должен уметь выделять важные моменты и основные положения. При написании конспекта лекций следует придерживаться следующих правил и рекомендаций:

1. При ведении конспекта рекомендуется структурировать материал по разделам, главам, темам. Вести нумерацию формул. Выделять по каждой теме постановку задачи, основные положения, выводы. Кратко записывать те пояснения лектора, которые показались особенно важными. Это позволит при подготовке к сдаче зачёта не запутаться в структуре лекционного материала.
2. Лекционный материал следует записывать в конспект лишь после того, как излагаемый лектором тезис будет вами дослушан до конца и понят.
3. При конспектировании следует отмечать непонятные на данном этапе положения, доказательства и пр.
4. Рекомендуется по каждой теме выразить своё мнение, комментарий, вывод.

Доработка конспекта лекции с применением учебника, методической литературы, дополнительной литературы, интернет-ресурсов: этот вид самостоятельной работы студентов особенно важен в том случае, когда одну и ту же задачу можно решать различными способами, а на лекции изложен только один из них. Кроме того, рабочая программа предполагает рассмотрение некоторых относительно несложных тем только во время самостоятельных занятий, без чтения лектором.

Подготовка к лабораторным работам состоит в теоретической подготовке (изучение конспекта лекций, методических указаний к данной работе и дополнительной литературы) и выполнении индивидуального задания. Выполнение каждой из запланированных работ заканчивается предоставлением отчёта. Требования к форме и содержанию отчёта приведены в методических указаниях к лабораторным работам или определяются преподавателем на первом занятии. Допускаясь к лабораторной работе, каждый студент должен представить преподавателю «заготовку» отчёта, содержащую: оформленный титульный лист, цель работы, задание, проект решения, полученные результаты, выводы.

Подготовка к сдаче зачёта.

Зачёт – форма промежуточной проверки знаний, умений, навыков, степени освоения дисциплины. Главная задача зачёта состоит в том, чтобы у студента по окончании изучения данной дисциплины сформировались определённое представление об общем содержании дисциплины, определённые теоретические знания и практические навыки, определённый кругозор. Готовясь к зачёту, студент приводит в систему знания, полученные на лекциях, на лабораторных занятиях, разбирается в том, что осталось непонятным, и тогда изучаемая им дисциплина может быть воспринята в полном объёме с присущей ей строгостью и логичностью, её практической направленностью.

Зачёт даёт возможность преподавателю определить теоретические знания студента и его практические навыки при решении определённых прикладных задач. Оцениваются: понимание и степень усвоения теоретического материала; степень знакомства с основной и дополнительной литературой, а также с современными публикациями; умение применить теорию к практике, решать определённые практические задачи данной предметной области, правильно проводить расчёты и т. д.; знакомство с историей данной науки; логика, структура и стиль ответа, умение защищать выдвигаемые положения.

Планируйте подготовку к зачёту, учитывая сразу несколько факторов: неоднородность в сложности учебного материала и степени его проработки в ходе обучения, свои индивидуальные способности. Рекомендуется делать перерывы в занятиях через каждые 50–60 минут на 10 минут. После 3–4 часов занятий следует сделать часовой перерыв. Чрезмерное утомление приведёт к снижению тонуса интеллектуальной деятельности. Целесообразно разделять весь рабочий день на три рабочих периода – с утра до обеда, с обеда до ужина и с ужина до сна. Каждый рабочий период дня должен заканчиваться отдыхом не менее 1 часа. Работая в

сессионном режиме, студент имеет возможность увеличить время занятий с 10 (как требовалось в семестре) до 12 часов в сутки.

Подготовку к зачёту следует начинать с общего планирования своей деятельности. С определения объёма материала, подлежащего проработке, необходимо внимательно сверить конспекты с программой дисциплины, чтобы убедиться, все ли разделы отражены в лекциях, отсутствующие темы изучить по учебнику. Второй этап предусматривает системное изучение материала по данному предмету с обязательной записью всех выкладок, выводов, формул. На третьем этапе – этапе закрепления – полезно чередовать углублённое повторение особенно сложных вопросов с беглым повторением всего материала.

5. РЕКОМЕНДАЦИИ ПО РАБОТЕ С ЛИТЕРАТУРОЙ

Основная литература, литература для дополнительного чтения и методические указания к лабораторным работам указаны в рабочей программе данной дисциплины и доступны посредством ЭБС.

Теоретический материал курса становится более понятным, когда дополнительно к прослушиванию лекции и изучению конспекта изучаются и книги по данному предмету. Литературу по дисциплине рекомендуется читать как в бумажном, так и в электронном виде (если отсутствует бумажный аналог). Полезно использовать несколько учебников и пособий по дисциплине. Рекомендуется после изучения очередного параграфа ответить на несколько вопросов по данной теме. Кроме того, полезно мысленно задать себе следующие вопросы (и попробовать ответить на них): «о чём этот параграф?», «какие новые понятия введены, каков их смысл?», «зачем мне это нужно по специальности?».

Рекомендуется самостоятельно изучать материал, который ещё не прочитан на лекции и не применялся на практическом занятии, тогда занятия будут гораздо понятнее. В течение недели рекомендуется выбрать время (1 час) для работы с литературой.

6. БИБЛИОГРАФИЧЕСКИЙ СПИСОК

6.1. Основная литература

1. Гергель В. П. Теория и практика параллельных вычислений: учебное пособие. – Москва, Саратов: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. – 500 с. – ISBN 978-5-4497-0389-7. – URL: http://www.iprbookshop.ru/89478.html
2. Демидова Л. А. Интеллектуальный анализ данных на языке Python: учебно-методическое пособие. – Москва: РТУ МИРЭА, 2021. – 92 с. – URL: https://e.lanbook.com/book/218693

3. Соснин В. В. Облачные вычисления в образовании: учебное пособие. – Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2024. – 109 с. – ISBN 978-5-4497-2452-6. – URL: https://www.iprbookshop.ru/133953.html
4. Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н. Большие данные. Big Data: учебник для вузов. – Санкт-Петербург: Лань, 2024. – 188 с. – ISBN 978-5-507-47346-5. – URL: https://e.lanbook.com/book/362318
5. Антипов О. В., Дмитриева Т. А., Москвитина О. А., Парфилова Н. И. Алгоритмы и структуры данных: учеб. – Москва: КУРС, 2022. – 216 с. – ISBN 978-5-907535-02-2, 978-5-907535-08-4.
6. Митяков Е. С., Шмелева А. Г., Ладынин А. И. Искусственный интеллект и машинное обучение: учебное пособие для вузов. – Санкт-Петербург: Лань, 2025. – 252 с. – ISBN 978-5-507-51465-6. – URL: https://e.lanbook.com/book/450827
7. Орешков В. И. Хранилища данных и OLAP-технологии: Учебное пособие. – Рязань: РИЦ РГРТУ, 2017. – URL: https://elib.rsreu.ru/ebs/download/595
8. Васильев Е. П., Орешков В. И. Интеллектуальный анализ данных в технологиях принятия решений: учеб. пособие. – Рязань: РИЦ РГРТУ, 2023. – URL: https://elib.rsreu.ru/ebs/download/3845
9. Гринченко Н. Н., Хазриева Н. И., Баранова С. Н. Базы данных. Проектирование моделей данных: учебник. – Москва: КУРС, 2024. – 260 с.; прил. – ISBN 978-5-907064-20-1.
10. Громов А. Ю., Гринченко Н. Н. Проектирование информационных систем: Методические указания. – Рязань: РИЦ РГРТУ, 2019. – URL: https://elib.rsreu.ru/ebs/download/1967
11. Бубнов А. А., Бубнов С. А., Майков К. А. Разработка и анализ требований к программному обеспечению: учебник. – Рязань: КУРС, 2020. – URL: https://elib.rsreu.ru/ebs/download/2678

6.2. Дополнительная литература

1. Зиангирова Л. Ф. Облачные вычисления: учебное пособие. – Саратов: Ай Пи Ар Медиа, 2019. – 168 с. – ISBN 978-5-4497-0175-6. – URL: http://www.iprbookshop.ru/85804.html
2. Баранчиков П. А. Параллельное программирование: метод. указ. к лаб. работам: Методические указания. – Рязань, 2023. – URL: https://elib.rsreu.ru/ebs/download/3561
3. Юрина Т. А. Параллельное программирование. – Омск: СибАДИ, 2023. – 106 с. – URL: https://e.lanbook.com/book/407159

6.3. Методические разработки

1. Ракитский А. А., Дементьева К. И. Методы машинного обучения: учебно-методическое пособие. – Новосибирск: Сибирский государственный университет телекоммуникаций и информатики, 2023. – 35 с. – ISSN 2227-8397. – URL: https://www.iprbookshop.ru/138857.html

6.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Электронная библиотека РГРТУ: http://elib.rsreu.ru/
2. Электронная библиотека IPRBooks: http://iprbookshop.ru/
3. Электронно-библиотечная система издательства «Лань»: http://e.lanbook.com

6.5. Перечень программного обеспечения и информационных справочных систем

Лицензионное и свободно распространяемое программное обеспечение:

Наименование	Описание
Операционная система Windows	Коммерческая лицензия
Adobe Acrobat Reader	Свободное ПО
VirtualBox	Свободное ПО
Anaconda3	Свободно распространяемое программное обеспечение под лицензиями
MongoDB — система управления базами данных NoSQL типа	Свободное ПО

Информационные справочные системы:

1. Справочная правовая система «КонсультантПлюс» (договор об информационной поддержке №1342/455-100 от 28.10.2011 г.)
2. Система КонсультантПлюс: http://www.consultant.ru
3. Информационно-правовой портал ГАРАНТ.РУ: http://www.garant.ru