

Паспорт

оценочных материалов для проведения текущего контроля и
промежуточной аттестации обучающихся по дисциплине (модулю)

Б1.В.ДВ.01.01 Технологии обработки и анализа больших массивов данных

Перечень оценочных материалов и индикаторов достижения компетенций,
сформированность которых они контролируют

Наименование оценочного средства	Коды индикаторов достижения формируемых компетенции	Номер приложения
Выполнение практических заданий	ИД-1 ПК-1; ИД-2 ПК-1	1
Курсовой проект	ИД-1 ПК-1; ИД-2 ПК-1	2
Экзамен	ИД-1 ПК-1; ИД-2 ПК-1	3

Утверждено на заседании кафедры «Вычислительная техника»
протокол №3 от «11» октября 2021 года

Заведующий кафедрой _____ К.В.Святов

Выполнение практических заданий

1. Процедура выполнения практических работ

Количество проводимых практических работ в течение всего периода освоения дисциплины	8 работ
Формат проведения результатов	Электронный
Методические рекомендации (при необходимости)	Рекомендуется представлять выполненную практическую работу в виде ссылки на репозиторий github (для практических работ по созданию приложения) и PDF-файла отчета, оформленного в системе Latex. Приведенный код должен быть аккуратно оформлен и прокомментирован для удобства его проверки. В случае включения графического материала в файлы, их содержание должно быть читаемо и понятно.

2. Шкала оценивания с учетом срока сдачи¹

Показатель качества исполнения работы	Балл (по 100-балльной шкале)
Практическая работа выполнена в полном объеме и в срок; результаты работы скрипта/приложения корректны на тестовых примерах или проведен требуемый вычислительный эксперимент; результаты работы представлены преподавателю; исполнитель может объяснить действия команд и внести простые изменения в алгоритм по требованию преподавателя.	90-100
Практическая работа выполнена практически в полном объеме и в срок; результаты работы скрипта/приложения корректны на тестовых примерах или проведен требуемый вычислительный эксперимент; результаты работы	60-89

¹ За несвоевременную сдачу обучающемуся могут быть начислены штрафные баллы.

представлены преподавателю; исполнитель может объяснить действия команд программы, возникают проблемы с быстрой модификацией алгоритмов по требованию преподавателя.	
Работа выполнена не в полном объеме (скрипт/приложение работает некорректно); исполнитель может объяснить действия команд программы, однако не может модифицировать код по требованию преподавателя.	40-59
Работа не выполнена или выполнена не в полном объеме (не проведены заданные вычислительные эксперименты); результаты работы не представлены преподавателю или представлены с существенным нарушением срока; исполнитель не может объяснить действия команд программы и не может внести простые изменения в алгоритм по требованию преподавателя.	< 40

3. Перечень практических работ

Результатом работы должен быть обзор в виде отчета систем хранения и обработки больших данных. Конкретные технологии выбираются исходя из целесообразности применения для решения предметной задачи, которая определяется тематикой магистерской работы студента.

1. Знакомство с понятием Data Mining, OLTP, OLAP, технологиями Apache Hadoop MapReduce
2. Использование технологий непрерывного развертывания и интеграции (github/ gitlab, jenkins)
3. Применение микросервисной архитектуры в потоковой обработке Big Data
4. Использование реляционных хранилищ данных для big data (PostgreSQL) и NoSql хранилищ данных для big data (Neo4J, CouchDB, Redis, MongoDB)
5. Аналитика и визуализация больших данных с помощью инструмента Grafana.
6. Использование облачных вычислений при помощи Apache Spark, и Apache Ignite.
7. Знакомство с контейнерами Docker, LXC и Kubernetes.
8. Системы виртуализации Hyper-V, Proxmox, Amazon.

Примерные вопросы при собеседовании при сдаче отчета:

1. Как устроен Apache Hadoop MapReduce: принцип работы?
2. В какой последовательности технология Apache Hadoop MapReduce использует в рабочем процессе задачи-распределители и задачи-редукторы?
3. Какие функции в Apache Hadoop MapReduce запускает главный контроллер-Мастер?
4. В чем состоит рекурсивное обобщение Apache Hadoop MapReduce?
5. Подходит ли для реализации микросервисной архитектуры и интеграции разрозненных систем Apache Spark?
6. Можно ли анализировать данные, хранящиеся в Apache Hadoop, с помощью стандартного инструментария SQL-запросов?

7. Какая технология больше всего подойдет автоматизации запуска пакетных задач в рамках конвейера обработки больших данных по расписанию?
8. Для полнотекстового интеллектуального поиска и аналитики полуструктурированным данным в формате JSON отлично подходит СУБД?
9. Что такое дедупликация данных?
10. Какой способ хранения данных используется в MongoDB?
11. Apache Hadoop - это..?
12. Достоинства Amazon?
13. Какая реализация MapReduce является закрытой?
14. Что является процессом создания и выбора модели для предсказания вероятности наступления некоторого события?
15. Дайте определение термину "Предиктивное моделирование"?
16. Основная идея NoSQL БД?
17. Принцип 3Vs расшифровывается как?
18. Какой из принципов работы не применяется к Big Data?
19. Что относится к средствам интеграции в «Business Intelligence»?
20. Что такое жизненный цикл аналитики данных?
21. Что является плюсом репликации?
22. Какая операция в NoSQL использует в аргументах пару key,value?

Курсовой проект

1. Процедура проведения

Этапы проведения КП (КР) с указанием сроков выполнения:

1. Выдача задания (3 неделя семестра).
2. Изучение алгоритмов ИИ, которые могут быть использованы для решения поставленной задачи. (4 - 5 недели семестра)
3. Составление реферата по изученным алгоритмам. (6 неделя семестра)
4. Проектирование и разработка программного комплекса с использованием технологий, изученных в ходе выполнения практических работ. (7 - 9 недели семестра)
5. Написание пояснительной записки к курсовому проекту. (14 неделя семестра)

2. Шкала оценивания с учетом срока сдачи

Критерии оценки уровня сформированности компетенций	Балл
<ul style="list-style-type: none"> – курсовой проект выполнен в полном объеме и соответствует заданию; – пояснительная записка составлена аккуратно, последовательно с учетом требований стандартов по составлению текстовых документов; – практическая часть курсового проекта выполнена в полном объеме, приложение – надёжно, эффективно, имеет удобный пользовательский интерфейс; – выполнение курсового проекта проходило в полном соответствии с графиком курсового проектирования; – защита курсового проекта проведена грамотно с демонстрацией всех возможностей разработанного приложения. 	Отлично
<p>По отношению к оценке «отлично»:</p> <ul style="list-style-type: none"> – возможны некоторые отступления от графика выполнения курсового проектирования; – существуют незначительных погрешностей в оформлении пояснительной записки и приложения (практической части курсового проекта). – преподавателем отмечены небольшие замечания к интерфейсу и работоспособности приложения (практической части курсового проекта). 	Хорошо
<p>По отношению к оценке «хорошо»:</p> <ul style="list-style-type: none"> – существование ошибок, неточностей и непоследовательности при составлении пояснительной 	Удовлетворительно

записки; – отсутствие самостоятельности и творческого подхода при разработке приложения; – существование незначительных погрешностей в работе приложения; – значительное отступление от сроков выполнения курсового проекта; – недостаточно грамотная защита и неполная демонстрация возможностей разработанного приложения.	
– несоответствие курсового проекта заданию; – отсутствие учета требований стандартов по оформлению текстовых документов при составлении пояснительной записки; – существование ошибок и непоследовательности в работе приложения; – значительное отступление от сроков выполнения курсового проекта; – неспособность грамотно защитить курсовой проект.	Неудовлетворительно

3. Задание по курсовому проекту

Наполнение курсового проекта составляют выполненные студентами 8 практических заданий, объединенных одной целью разработать систему хранения и обработки больших данных с помощью изученных технологий. Вариативность заданий обеспечивается спецификой темы магистерской диссертации.

Номер этапа	Задание
1	Знакомство с технологией Apache Hadoop MapReduce
2	Использование реляционных хранилищ данных для big data (PostgreSQL)
3,4	Применение NoSql хранилищ данных для big data (Neo4J, CouchDB, Redis, Apache Cassandra)
5,6	Использование NoSql хранилищ для big data (MongoDB, InfluxDB, Elasticsearch, RabbitMQ)
7	Аналитика и визуализация больших данных с помощью инструмента Grafana.
8	Использование облачных вычислений при помощи Apache Spark
9.	Система хранения и обработки больших данных, реализованная студентом с помощью выбранных инструментов и технологий

Конкретные технологии выбираются исходя из целесообразности применения для решения предметной задачи, которая определяется тематикой магистерской работы студента.

Экзамен

1. Процедура проведения

Общее количество вопросов к экзамену	14 вопросов
Количество вопросов в билете	2 вопросов
Наличие задач в билете	нет
Формат проведения	Устно и письменно
Методические рекомендации (при необходимости)	

2. Шкала оценивания с учетом текущего контроля работы обучающегося в семестре

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
выставляется обучающемуся, если он показал глубокие знания материала по поставленному вопросу, грамотно, логично и стройно его излагает	Отлично
Выставляется обучающемуся, если он твердо знает материал, грамотно его излагает, но допускает несущественные неточности в ответе на вопрос	Хорошо
выставляется обучающемуся, если он показывает знания только основных положений по поставленному вопросу, требует в отдельных случаях наводящих вопросов для принятия правильного решения, допускает отдельные неточности	Удовлетворительно
выставляется обучающемуся, если он допускает грубые ошибки в ответе на поставленный вопрос	Неудовлетворительно

Вопросы:

1. Определение больших данных, ключевые характеристики. Примеры задач больших данных. Основные виды данных. Дать краткую сравнительную характеристику инструментария ПО для анализа данных.
2. «Жизненный цикл» проекта по аналитике больших данных. Типовая архитектура проекта в области больших данных. Перечислить используемые технологии, указать степень вовлеченности каждой из технологий на каждом этапе работы над проектом. Перечислить основные роли исполнителей проекта.
3. Хранилища данных. Аналитическая обработка данных (OLAP). Принцип организации многомерного куба. Различия между OLTP и OLAP.

4. Основные задачи и методы Data Mining. Этапы интеллектуального анализа данных. Методы интеллектуального анализа данных.
5. Технология технологиями Apache Hadoop MapReduce. Ключевые понятия. . Описать принцип работы. Нарисовать схему. Перечислить слабые и сильные стороны. Обозначить области применимости. Привести примеры использования.
6. Роль микросервисной архитектуры в потоковой обработке big data.
7. Использование реляционных хранилищ данных для big data. (PostgreSQL)
8. Альтернативные архитектуры баз данных – NoSQL (Not only SQL). Характерные особенности. Типы No-Sql баз данных.
9. Применение NoSql хранилищ данных для big data (Ne4J, CouchDB, Redis, Apache Cassandra)
10. Использование технологий непрерывного развертывания и интеграции больших данных
11. Когнитивные методы анализа больших данных
10. Аналитика и визуализация больших данных. Определение визуализации. Показать важность визуализации в аналитике больших данных. Привести примеры и инструменты для визуализации.
11. Облачные вычисления. Ключевые технологии и понятия.
12. Знакомство и применение контейнеров при работе с большими данными Docker, LXC и Kubernetes.
13. Системы виртуализации больших данных Hyper-V, Proxmox, Amazon.
14. Научные проблемы больших данных. Показать значимость проблем, актуальность, связь с областями математики и инженерии.