

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики

УТВЕРЖДАЮ
декан факультета вычислительной
математики и кибернетики


/И.А. Соколов /
«27» сентября 2022г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине
Основы обработки больших данных

Уровень высшего образования:
бакалавриат

Направление подготовки / специальность:
01.03.02 "Прикладная математика и информатика" (3++)

Направленность (профиль) ОПОП:
Искусственный интеллект и анализ данных

Форма обучения:
очная

Рассмотрен и утвержден
на заседании Ученого совета факультета ВМК
(протокол №7, от 27 сентября 2022 года)

Москва 2022

1. ФОРМЫ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

В процессе и по завершении изучения дисциплины оценивается формирование у студентов следующих компетенций:

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	Индикатор (показатель) достижения компетенции	Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций
ПК-8. Способен разрабатывать системы анализа больших данных	ПК-8.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных	Знать основные понятия и архитектуру системы хранения и обработки больших данных Hadoop; языки высокого уровня анализа больших данных; основные понятия, потенциальные возможности и особенности применения No-SQL баз данных. Уметь разрабатывать программы для решения задач анализа больших данных на основе парадигмы MapReduce; разрабатывать программы для решения задач анализа больших данных с помощью языков высокого уровня на платформе Hadoop. Владеть основными навыками выбора платформы хранения больших данных и выбора средств для решения задач анализа больших данных.

1.1. Текущий контроль успеваемости

Текущий контроль успеваемости осуществляется путем оценки результатов выполнения заданий практических (семинарских) занятий, самостоятельной работы, предусмотренных учебным планом и посещения занятий/активность на занятиях.

В качестве оценочных средств текущего контроля успеваемости предусмотрены:

выполнение заданий на практических (семинарски) занятиях

Примерные практические контрольные задания для текущего контроля успеваемости.

1. Реализовать операции Select, Project и Join в виде MapReduce- приложения.
2. Реализовать алгоритм TF*IDF в виде MapReduce- приложения.
3. Реализовать алгоритм поиска кратчайшего пути в графе в виде MapReduce- приложения.

4. Реализовать метод кластеризации K-means в виде MapReduce- приложения.
5. Написать представление на языке AQL для извлечения информации (список стран и протяженность границы) о граничащих странах
6. Написать представление на языке AQL для извлечения информации о продолжительности железных дорог по типу колеи
7. Написать представление на языке AQL для извлечения списка международных организаций, упорядоченных по числу стран, принимающих участие в этой организации
8. Написать представление на языке AQL для извлечения количества вхождений используемых природных ресурсов
9. Написать представление на языке AQL для извлечения количества импортируемых товаров для каждой страны

1.2. Промежуточная аттестация

Промежуточная аттестация осуществляется в форме экзамена

В качестве средств, используемых на промежуточной аттестации предусматривается:

Билеты

1.3. Типовые задания для проведения промежуточной аттестации

Вопросы к экзамену

1. Что такое большие данные (3v, 4v)
2. Чем знаменит IBM Watson
3. Что называют традиционными архитектурами хранения данных
4. В чем отличие систем хранения больших данных (масштабируемость, отказоустойчивость)
5. Что такое Hadoop (HDFS+MapReduce)
6. На каком языке написан Hadoop
7. Какие технологии, связанные с Hadoop, Вы знаете
8. Хорошо ли использовать Hadoop для транзакций
9. Что такое блок в hadoop
10. Что такое репликация
11. Назовите любую команду для работы с Hadoop
12. Какие узлы есть в Hadoop кластере
13. Что на вход принимает map
14. Что на вход принимает reduce
15. Что происходит между Map и reduce
16. Отказоустойчивость, что если упадет DataNode, NameNode, TaskTracker, JobTracker
17. Зачем нужен Combiner
18. Что такое Hadoop 2
19. Основные отличия Hadoop 2 от Hadoop 1
20. Что такое Yarn, основная идея
21. Какие языки высокоуровневого программирования Вы знаете
22. Что такое Pig
23. Что такое Hive, пример команды на Hive
24. Что такое Jaql
25. Что такое Spark, основная идея
26. Зачем нужны NoSQL базы данных
27. Модель данных HBase
28. Основные этапы извлечения информации из текстов
29. Подходы к распознаванию сущностей из текстов
30. Язык AQL. Компоненты языка AQL

Пример экзаменационного билета

1. Какие узлы есть в Hadoop кластере
2. Зачем нужны NoSQL базы данных
3. Задача

Примерное практическое контрольное задание для промежуточной аттестации.

Тема – Разработка приложений MapReduce на языке Java.

Реализация алгоритма TF*IDF в виде MapReduce-приложения на языке Java (допускается использование и других языков программирования). Входными данными являются: директория, содержащая несколько текстовых файлов, и поисковая строка, состоящая из нескольких слов. Формат представления входных и выходных данных выбирается магистрами.

Тема – Разработка и выполнение программ на языке AQL

Написать представление для извлечения информации (список стран и протяженность границы) о граничащих странах из заданного файла. Представление должно содержать 3 поля: country, bordered_country, distance. Пример фрагмента текста с извлекаемой информацией:

Geography Afghanistan

...

Land boundaries: total: 5,529 km border countries: China 76 km, Iran 936 km, Pakistan 2,430 km, Tajikistan 1,206 km, Turkmenistan 744 km, Uzbeki-stan 137 km

2. КРИТЕРИИ ОЦЕНКИ ПО ДИСЦИПЛИНЕ

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Оценка	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)
виды оценочных средств				
Знания (виды оценочных средств: приведены в п. 1.2.)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения (виды оценочных средств: приведены в п. 1.2.)	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности не принципиального характера)	Успешное и систематическое умение
Навыки (владения, опыт деятельности) (виды оценочных средств: приведены в п. 1.2..)	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач