

Документ подписан Министерством науки и высшего образования Российской Федерации
Информация о владельце:
ФИО: Макаренко Елена Николаевна
Должность: Ректор
Дата подписания: 18.04.2024 08:53:15
Уникальный программный ключ:
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

УТВЕРЖДАЮ
Директор Института магистратуры
Иванова Е.А.
«01» июня 2023г.

Рабочая программа дисциплины
Реализация систем с параллельной обработкой данных и облачных систем

Направление 09.04.04 Программная инженерия
магистерская программа 09.04.04.01 "Системное и прикладное программное
обеспечение"

Для набора 2023 года

Квалификация
магистр

КАФЕДРА Информационные технологии и защита информации**Распределение часов дисциплины по семестрам**

Семестр (<Курс>.<Семестр на курсе>)	4 (2.2)		Итого	
	4 4/6			
Неделя	4 4/6			
Вид занятий	УП	РП	УП	РП
Лекции	8	8	8	8
Лабораторные	16	16	16	16
Практические	16	16	16	16
Итого ауд.	40	40	40	40
Контактная работа	40	40	40	40
Сам. работа	68	68	68	68
Итого	108	108	108	108

ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 28.03.2023 протокол № 9.

Программу составил(и): к.ф.-м.н., доц., Карнаухов С.Н.

Зав. кафедрой: к.э.н., доц. Ефимова Е.В.

Методическим советом направления: д.э.н., проф., Е.Н. Тищенко

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

- | | |
|-----|---|
| 1.1 | формирование знаний, умений и навыков, необходимых для организации параллельных вычислений; формирование у магистров необходимого объема теоретических и практических знаний о технологии облачных вычислений, умений и навыков практической реализации облачных технологий в современном бизнесе, изучение инструментальных средств данной технологии. |
|-----|---|

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

УК-4:Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия

ПК-7:Способен осуществлять разработку компонентов систем управления базами данных

ПК-8:Способен проводить интеграцию разработанных компонентов системного программного обеспечения

В результате освоения дисциплины обучающийся должен:

Знать:

современные коммуникативные технологии на государственном и иностранном языках; закономерности деловой устной и письменной коммуникации (соотнесено с индикатором УК-4.1).
основные модели данных и их организации, методы обработки данных, основы современных систем управления базами данных(соотнесено с индикатором ПК-7.1).
основы архитектуры, устройства и функционирования вычислительных систем, типичный процесс интеграции, подходы к интеграции компонентов системного программного обеспечения (соотнесено с индикатором ПК-8.1).

Уметь:

применять на практике коммуникативные технологии, методы и способы делового общения (соотнесено с индикатором УК -4.2).
применять языки программирования, определенные в техническом задании на разработку системы управления базами данных, для написания программного кода, обнаруживать ошибки в работе системы управления базами данных, готовить документацию по разработанной системе управления базами данных (соотнесено с индикатором ПК-7.2).
определять порядок сборки разработанных компонентов системного программного обеспечения с учетом зависимостей в компонентах, устанавливать и настраивать серверы интеграции (соотнесено с индикатором ПК-8.2).

Владеть:

методикой межличностного делового общения на государственном и иностранном языках, с применением профессиональных языковых форм и средств (соотнесено с индикатором УК-4.3).
методами анализа ошибок в компонентах системы управления базами данных по данным эксплуатации, методами анализа результатов тестирования разрабатываемых компонентов системы управления базами данных (соотнесено с индикатором ПК-7.3).
методами выбора стратегии интеграции и практикуемых способов сборки разработанного системного программного обеспечения, способами определения порядка управления версиями сборок разработанного системного программного обеспечения (соотнесено с индикатором ПК-8.3).

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература
	Раздел 1. Параллельная обработка больших массивов данных				
1.1	Тема 1.1. «MapReduce: приемы и стратегии реализации» Феномен Big Data. Модель программирования MapReduce. Принципы параллельной реализации вычислений. Область применения и примеры задач. Принципы распределенной реализации MapReduce на кластерных системах. Платформа Apache Hadoop. Интерфейсы прикладного программирования и реализация программ для Hadoop. Локальная отладка и запуск программ на кластере. /Лек/	4	4	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3Л2.1 Л2.3 Л2.4 Л2.5 Л2.6

1.2	Тема 1.1. «MapReduce: приемы и стратегии реализации» Феномен Big Data. Модель программирования MapReduce. Принципы параллельной реализации вычислений. Область применения и примеры задач. Принципы распределенной реализации MapReduce на кластерных системах. /Лаб/	4	6	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
1.3	Тема 1.1. «MapReduce: приемы и стратегии реализации» Феномен Принципы распределенной реализации MapReduce на кластерных системах. Платформа Apache Hadoop. Интерфейсы прикладного программирования и реализация программ для Hadoop. Локальная отладка и запуск программ на кластере. Приемы и стратегии реализации MapReduce- программ. Изучение структуры кодов /Пр/	4	6	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
1.4	Приемы и стратегии реализации MapReduce-программ. /Ср/	4	10	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
1.5	Тема 1.2. «MapReduce: инструменты и практические примеры» Высокоуровневые языки и инструментарии для работы с Hadoop. Практические примеры использования MapReduce. Ограничения модели MapReduce, расширения и альтернативные подходы. Система Apache Spark. Понятие Resilient Distributed Dataset (RDD), операции над RDD. Принципы распределенного выполнения Spark-программы на кластере. Запуск, отладка и оптимизация Spark- программ. /Лек/	4	4	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
1.6	Тема 1.2. «MapReduce: инструменты и практические примеры» Высокоуровневые языки и инструментарии для работы с Hadoop. Практические примеры использования MapReduce. Ограничения модели MapReduce, расширения и альтернативные подходы. Система Apache Spark. /Лаб/	4	4	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
1.7	Тема 1.2 «MapReduce: инструменты и практические примеры». Фильтрация данных (например, «Найти все записи с IP-адреса 123.123.123.123» в логах web-сервера);Преобразование данных («Удалить колонку в csv-логах»); Загрузка и выгрузка данных из внешнего источника («Вставить все записи из лога в базу данных»).Изучение регистра WordCount (Определение класса сопоставления, класс редуктора. Технологии очистки данных /Пр/	4	4	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
1.8	Понятие Resilient Distributed Dataset (RDD), операции над RDD. Принципы распределенного выполнения Spark-программы на кластере. Запуск, отладка и оптимизация Spark- программ. /Ср/	4	10	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
	Раздел 2. Реализация облачных систем				
2.1	Тема 2.1 Azure Службы, доступные в Azure, в том числе службы вычислений, сетевые службы, службы хранения и базы данных. Сетевые ресурсы Azure, такие как виртуальные сети, VPN-шлюзы и Azure ExpressRoute. Сводные сведения о службах хранилища Azure, таких как Хранилище BLOB-объектов Azure, Хранилище дисков Azure и хранилище файлов Azure /Ср/	4	4	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3Л2.1 Л2.3 Л2.4 Л2.5 Л2.6

2.2	Тема 2.1. «Azure. Основных решения и средств управления в Azure». Службы искусственного интеллекта Azure для решения различных видов бизнес-задач. Выполните идентификацию служб виртуализации, таких как "Виртуальные машины Azure", "Экземпляры контейнеров Azure", "Служба Azure Kubernetes" и "Виртуальный рабочий стол Azure". Сравнение служб баз данных Azure, таких как Azure Cosmos DB, Azure SQL, база данных Azure для MySQL, база данных Azure для PostgreSQL, а также большие данные Azure и службы аналитики. Средства и службы разработки программного обеспечения для конкретного бизнес-сценария. Выбор подходящей службы мониторинга облака для решения различных видов бизнес-задач. Выбор подходящего средства управления Azure для решения различных технических задач и проблем. /Ср/	4	4	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3 Л2.1 Л2.2 Л2.3 Л2.4 Л2.5 Л2.6
2.3	Тема 2.1. «Azure. Разработка веб приложения» Разработка простого веб - приложения для загрузки изображений в хранилище Windows Azure Blob: подготовка приложения; загрузка и отображение изображений; удаление сущностей; копирование сущностей. Разработка примера, демонстрирующей основы работы с очередями Windows Azure , на примере рабочей и веб - ролей /Ср/	4	4	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3 Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
2.4	РaaS-сервис для облачного анализа больших данных (big data) на базе технологий Hadoop, Spark, HBase, Kafka. /Ср/	4	10	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3 Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
2.5	Тема 2.2. «Работа с компонентами MSC» Архитектура сервиса. Описание и версии компонентов. Шаблоны и компоновка кластеров. Apache Hive — система хранилища данных для Hadoop. Hadoop MapReduce - фреймворк для создания приложений, обрабатывающих большие объемы данных в парадигме MapReduce. Apache Spark - платформа для параллельной обработки данных в памяти. Apache Pig — платформа для обработки больших объемов данных с помощью скриптов. /Лаб/	4	6	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3 Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
2.6	Тема 2.2. «Работа с компонентами MSC» Архитектура сервиса. Описание и версии компонентов. Шаблоны и компоновка кластеров. Apache Hive — система хранилища данных для Hadoop. Hadoop MapReduce - фреймворк для создания приложений, обрабатывающих большие объемы данных в парадигме MapReduce. Apache Spark - платформа для параллельной обработки данных в памяти. Apache Pig — платформа для обработки больших объемов данных с помощью скриптов. /Пр/	4	6	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3 Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
2.7	Вопросы для самостоятельного изучения: 1. Введение в машинное обучение и быстрый старт с Azure ML. 2. Поддержка языка R внутри Azure ML. 3. Ценообразование на средство машинного обучения, позволяющее создавать, тестировать, активировать решения прогнозируемой аналитики в облаке и легко управлять ими. 4. Развертывание веб-сайта в Microsoft /Ср/	4	26	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3 Л2.1 Л2.3 Л2.4 Л2.5 Л2.6
2.8	/Зачёт/	4	0	ПК-8 УК-4 ПК-7	Л1.1 Л1.2 Л1.3 Л2.1 Л2.2 Л2.3 Л2.4 Л2.5 Л2.6

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущей и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**5.1. Основная литература**

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л1.1	Губарев В. В., Савульчик С. А., Чистяков Н. А.	Введение в облачные вычисления и технологии: учебное пособие	Новосибирск: Новосибирский государственный технический университет, 2013	https://biblioclub.ru/index.php?page=book&id=228962 неограниченный доступ для зарегистрированных пользователей
Л1.2	Соснин В. В.	Облачные вычисления в образовании	Москва: Национальный Открытый Университет «ИНТУИТ», 2016	https://biblioclub.ru/index.php?page=book&id=429074 неограниченный доступ для зарегистрированных пользователей
Л1.3	Сухорукова М. В., Тябин И. В.	Предпринимательство в области мобильных приложений и облачных сервисов: курс лекций	Москва: Национальный Открытый Университет «ИНТУИТ», 2016	https://biblioclub.ru/index.php?page=book&id=429874 неограниченный доступ для зарегистрированных пользователей

5.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л2.1		Функциональное программирование. Лекция 30. Асинхронные и параллельные вычисления. Презентация	Москва: Национальный Открытый Университет «ИНТУИТ», 2014	http://biblioclub.ru/index.php?page=book&id=237183 неограниченный доступ для зарегистрированных пользователей
Л2.2	Рытенкова О.	Информационная безопасность: журнал	Москва: ГРОТЕК, 2014	https://biblioclub.ru/index.php?page=book&id=238446 неограниченный доступ для зарегистрированных пользователей
Л2.3	Алексеев А. А.	Основы параллельного программирования с использованием Visual Studio 2010: курс лекций	Москва: Национальный Открытый Университет «ИНТУИТ», 2016	https://biblioclub.ru/index.php?page=book&id=428829 неограниченный доступ для зарегистрированных пользователей
Л2.4	Биллиг В. А.	Параллельные вычисления и многопоточное программирование	Москва: Национальный Открытый Университет «ИНТУИТ», 2016	https://biblioclub.ru/index.php?page=book&id=428948 неограниченный доступ для зарегистрированных пользователей
Л2.5	Туральчук К. А.	Параллельное программирование с помощью языка С	Москва: Национальный Открытый Университет «ИНТУИТ», 2016	https://biblioclub.ru/index.php?page=book&id=429098 неограниченный доступ для зарегистрированных пользователей
Л2.6	Николаев Е. И.	Параллельные вычисления: учебное пособие	Ставрополь: Северо- Кавказский Федеральный университет (СКФУ), 2016	https://biblioclub.ru/index.php?page=book&id=459124 неограниченный доступ для зарегистрированных пользователей

5.3 Профессиональные базы данных и информационные справочные системы

ИСС «КонсультантПлюс»

ИСС «Гарант» <http://www.internet.garant.ru/>

Официальная документация <https://hadoop.apache.org/docs/current/>

Официальная документация <https://azure.microsoft.com/ru-ru/>

5.4. Перечень программного обеспечения

Notepad ++

свободно распространяемый набор утилит, библиотек и фреймворк:

Apache Hadoop

Apache Spark

Apache Hive

Apache Pig

5.5. Учебно-методические материалы для студентов с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:

- столы, стулья;

- персональный компьютер / ноутбук (переносной);

- проектор;

- экран / интерактивная доска

Лабораторные занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными и/или свободно распространяемыми программными средствами и выходом в Интернет.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1 Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

1.1 Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
УК-4: Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия			
З. современные коммуникативные технологии на государственном и иностранном языках; закономерности деловой устной и письменной коммуникации (соотнесено с индикатором УК-4.1).	выбор методов распределенного хранения и обработки данных для решения профессиональных задач	полнота и обоснованность выбора методов распределенного хранения и обработки данных на основе изученной литературы	УО (1-35) З (1-35)
У. применять на практике коммуникативные технологии, методы и способы делового общения. (соотнесено с индикатором УК-4.2).	решение практико-ориентированных и лабораторных заданий на основе официальной документации: Apache Hadoop	правильность применения технологий, описанных в официальной документации Apache Hadoop	ЛЗ (1-4) ПЗ (1-4) ПОЗ (1)
В. методикой межличностного делового общения на государственном и иностранном языках, с применением профессиональных языковых форм и средств. (соотнесено с индикатором УК-4.3).	использование различных подходов к анализу кодов	правильность и обоснованность различных подходов к анализу кодов	ЛЗ (1-4) ПЗ (1-4) ПОЗ (1)
ПК-7: Способен осуществлять разработку компонентов систем управления базами данных.			

<p>3. основные модели данных и их организации, методы обработки данных, основы современных систем управления базами данных (соотнесено с индикатором ПК-7.1).</p>	<p>выбор методов распределенного хранения и обработки данных для решения профессиональных задач</p>	<p>полнота и обоснованность выбора методов распределенного хранения и обработки данных на основе изученной литературы</p>	<p>УО (Раздел 1 вопросы 1-15, Раздел 2 вопросы 1-4) 3 (1-23, 33)</p>
<p>У. - применять языки программирования, определенные в техническом задании на разработку системы управления базами данных, для написания программного кода, обнаруживать ошибки в работе системы управления базами данных, готовить документацию по разработанной системе управления базами данных (соотнесено с индикатором ПК-7.2).</p>	<p>решение практико-ориентированных и лабораторных заданий: применяет высокоуровневые языки и инструментарии Выбор подходящей службы мониторинга облака для решения различных видов бизнес-задач.</p>	<p>правильность применения высокоуровневые языки и инструментарии, средств управления для решения различных технических задач и проблем.</p>	<p>ЛЗ (1,2,4) ПЗ (1,2,4) ПОЗ (2,3)</p>
<p>В. методами анализа ошибок в компонентах системы управления базами данных по данным эксплуатации, методами анализа результатов тестирования разрабатываемых компонентов системы управления базами данных (соотнесено с индикатором ПК-7.3).</p>	<p>решение практико-ориентированных и лабораторных заданий различными способами применяет высокоуровневые языки и инструментарии для работы с Hadoop. Ограничения модели, расширения и альтернативные подходы</p>	<p>правильность и обоснованность выбора высокоуровневые языки и инструментарии для работы с Hadoop средств управления для решения различных технических задач и проблем.</p>	<p>ЛЗ (1,2,4) ПЗ (1,2,4) ПОЗ (2,3)</p>
<p>ПК -8 Способен проводить интеграцию разработанных компонентов системного программного обеспечения.</p>			
<p>3. основы архитектуры, устройства и функционирования вычислительных систем, типичный процесс интеграции, подходы к интеграции компонентов системного программного обеспечения (соотнесено с индикатором ПК-8.1).</p>	<p>выбор методов распределенного хранения и обработки данных для решения профессиональных задач</p>	<p>полнота и обоснованность выбора методов интеграции компонентов системного программного обеспечения на основе изученной литературы</p>	<p>УО (Раздел 1 вопросы 1-15, Раздел 2 вопросы 5-12) 3 (13-29)</p>

У. определять порядок сборки разработанных компонентов системного программного обеспечения с учетом зависимостей в компонентах, устанавливать и настраивать серверы интеграции (соотнесено с индикатором ПК-8.2).	решение практико-ориентированных и лабораторных заданий: применяет высокоуровневые языки и инструментарии для работы с Hadoop. Применяет базовые технологии разработки простого веб - приложения для загрузки изображений в хранилище	правильность применения: высокоуровневых языков и инструментария для работы с Hadoop	ЛЗ (3,4) ПЗ (1,3,4) ПОЗ (2,3)
В. методами выбора стратегии интеграции и практикуемых способов сборки разработанного системного программного обеспечения, способами определения порядка управления версиями сборок разработанного системного программного обеспечения (соотнесено с индикатором ПК-8.3).	решение практико-ориентированных и лабораторных заданий различными способами применяет высокоуровневые языки и инструментарии для работы с Hadoop. Ограничения модели, расширения и альтернативные подходы Применяет различные технологии разработки веб - приложения хранилище	правильность и обоснованность выбора высокоуровневых языков и инструментарии для работы с Hadoop	ЛЗ (3,4) ПЗ (1,3,4) ПОЗ (2,3)

З – вопросы к зачету, ЛЗ – лабораторное задание, ПОЗ-практико-ориентированное задание к зачету, УО- устный опрос

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале:

50-100 баллов (зачет)

0-49 баллов (незачет)

2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Вопросы к зачету

1. Модели программирования.
2. Принципы параллельной реализации вычислений. Область применения и примеры задач.
3. Принципы распределенной реализации MapReduce на кластерных системах.
4. Платформа Apache Hadoop.
5. Интерфейсы прикладного программирования и реализация программ для Hadoop.
6. Локальная отладка и запуск программ на кластере.
7. Приемы и стратегии реализации MapReduce-программ.
8. Высокоуровневые языки и инструментарии для работы с Hadoop.
9. Практические примеры использования MapReduce.
10. Ограничения модели MapReduce, расширения и альтернативные подходы.
11. Система Apache Spark.
12. Понятие Resilient Distributed Dataset (RDD), операции над RDD.

13. Принципы распределенного выполнения Spark-программы на кластере.
14. Запуск, отладка и оптимизация Spark-программ.
15. Infra: виртуальные машины и дополнительные инструменты их настройки.
16. Cloud storage: объектное хранилище с поддержкой S3 API.
17. Графические адаптеры: облачные вычисления на базе GPU.
18. Containers: контейнеры Kubernetes в облаке.
19. Big data: облачная обработка больших данных
20. Архитектура сервиса.
21. Описание и версии компонентов.
22. Шаблоны и компоновка кластеров.
23. Apache Hive — система хранилища данных для Hadoop.
24. Hadoop MapReduce - фреймворк для создания приложений, обрабатывающих большие объемы данных в парадигме MapReduce.
25. Apache Spark - платформа для параллельной обработки данных в памяти.
26. Apache Pig — платформа для обработки больших объемов данных с помощью скриптов.
27. Состав и возможности программного комплекса Apache Hadoop.
28. Языки поисковых запросов для Hadoop.
29. Принципы работы Apache Spark.
30. Компоненты экосистемы Hadoop.
31. MapReduce — модель распределённых вычислений
32. PaaS-сервис для анализа больших данных на базе Apache Hadoop
33. Azure . Службы вычислений, сетевые службы, службы хранения и базы данных.
34. Сетевые ресурсы Azure
35. Службы хранилища Azure

Практико-ориентированные задания к зачету

Задание 1

Предположим, у нас есть данные о сотрудниках в четырех разных файлах — А, В, С и D. Предположим также, что во всех четырех файлах есть дубликаты записей о сотрудниках из-за повторного импорта данных о сотрудниках из всех таблиц базы данных. (См рисунок).

name, salary	name, salary	name, salary	name, salary
satish, 26000	gopal, 50000	satish, 26000	satish, 26000
Krishna, 25000	Krishna, 25000	kiran, 45000	Krishna, 25000
Satishk, 15000	Satishk, 15000	Satishk, 15000	manisha, 45000
Raju, 10000	Raju, 10000	Raju, 10000	Raju, 10000

Задание 2

Создание аккаунта хранилищ и сервисного компонента

Задание 3

Выполните идентификацию служб виртуализации, таких как "Виртуальные машины", "Экземпляры контейнеров", "Служба Kubernetes" и "Виртуальный рабочий стол".

Ключ для контроля правильности выполнения практико-ориентированного задания к зачету

Задание 1.

Фаза Мар обрабатывает каждый входной файл и предоставляет данные о сотруднике в парах ключ-значение (<k, v>: <emp name, salary>).

```
<k: employee name, v: salary>
Max= the salary of an first employee. Treated as max salary

if(v(second employee).salary > Max){
    Max = v(salary);
}

else{
    Continue checking;
}
```

Задание 2.

```
# put resource group in a variable so you can use the same group name going forward,
# without hard-coding it repeatedly
$resourceGroup = "storage-resource-group"
$location = "westus"
New-AzResourceGroup -Name $resourceGroup -Location $location

New-AzStorageAccount -ResourceGroupName $resourceGroup `
    -Name <account-name> `
    -Location $location `
    -SkuName Standard_RAGRS `
    -Kind StorageV2
$resourceGroupName = Read-Host -Prompt "Enter the Resource Group name"
$location = Read-Host -Prompt "Enter the location (i.e. centralus)"

New-AzResourceGroup -Name $resourceGroupName -Location "$location"
New-AzResourceGroupDeployment -ResourceGroupName $resourceGroupName -TemplateUri
"https://raw.githubusercontent.com/Azure/azure-quickstart-templates/master/101-storage-account-cr
eate/azuredeploy.json"
```

Задание 3

```
"$schema": "https://schema.management.azure.com/schemas/2015-01-01/deploymentTemplate.json#",
"contentVersion": "1.0.0.0",
"parameters": {
    "storageAccounts_mysourceaccount_name": {
        "defaultValue": "mytargetaccount",
        "type": "String"
    }
},
"$schema": "https://schema.management.azure.com/schemas/2015-01-01/deploymentTemplate.json#",
"contentVersion": "1.0.0.0",
"parameters": {
    "storageAccounts_mysourceaccount_name": {
        "defaultValue": "mytargetaccount",
        "type": "String"
    }
},
},
```

Критерии оценивания:

- 50-100 баллов (зачет) – изложенный материал фактически верен, наличие глубоких исчерпывающих знаний; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе; практико-ориентированное задание выполнено правильно и прокомментировано; наличие

- твердых и достаточно полных знаний, правильные действия по применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности, неуверенность и неточность ответов на дополнительные и наводящие вопросы; практико-ориентированное задание выполнено правильно, но не прокомментировано; при неполном ответе на вопросы; затрудняется ответить на дополнительные вопросы; практико-ориентированное задание выполнено с ошибками и отсутствуют комментарии;
- 0-49 баллов (незачет) – ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы; практико-ориентированное задание не выполнено.

Лабораторные задания

Раздел 1. Параллельная обработка больших массивов данных

Лабораторное задание 1 (8 баллов).

«MapReduce: приемы и стратегии реализации»

Феномен Big Data. Модель программирования MapReduce. Принципы параллельной реализации вычислений. Область применения и примеры задач. Принципы распределенной реализации MapReduce на кластерных системах.

Лабораторное задание 2 (12 баллов).

«MapReduce: инструменты и практические примеры»

Высокоуровневые языки и инструментарии для работы с Hadoop. Практические примеры использования MapReduce. Ограничения модели MapReduce, расширения и альтернативные подходы. Система Apache Spark.

Максимальное количество баллов по лабораторным заданиям раздела 1 – 20

Раздел 2. Реализация облачных систем

Лабораторное задание 3 (8 баллов).

«Основных решения и средств управления ". Службы искусственного интеллекта для решения различных видов бизнес-задач.

Выполните идентификацию служб виртуализации, таких как "Виртуальные машины ", "Экземпляры контейнеров", "Служба Kubernetes" и "Виртуальный рабочий стол ".

Сравнение служб баз данных, а также большие данные и службы аналитики. Средства и службы разработки программного обеспечения для конкретного бизнес-сценария.

Выбор подходящей службы мониторинга облака для решения различных видов бизнес-задач.

Выбор подходящего средства управления для решения различных технических задач и проблем.

Лабораторное задание 4 (12 баллов).

«Работа с компонентами MSC»

Архитектура сервиса. Описание и версии компонентов. Шаблоны и компоновка кластеров. Apache Hive — система хранилища данных для Hadoop. Hadoop MapReduce - фреймворк для создания приложений, обрабатывающих большие объемы данных в парадигме MapReduce. Apache Spark - платформа для параллельной обработки данных в памяти. Apache Pig — платформа для обработки больших объемов данных с помощью скриптов.

Максимальное количество баллов по лабораторным заданиям раздела 2 – 20

Практические задания

Раздел 1. Параллельная обработка больших массивов данных

Практическое задание 1 (5 баллов).

«MapReduce: приемы и стратегии реализации»

Феномен Принципы распределенной реализации MapReduce на кластерных системах.
Платформа Apache Hadoop.
Интерфейсы прикладного программирования и реализация программ для Hadoop.
Локальная отладка и запуск программ на кластере. Приемы и стратегии реализации MapReduce-программ.
Изучение структуры кодов

Практическое задание 2 (15 баллов).

«MapReduce: инструменты и практические примеры». Фильтрация данных (например, «Найти все записи с IP-адреса 123.123.123.123» в логах web-сервера); Преобразование данных («Удалить колонку в csv-логах»);

Загрузка и выгрузка данных из внешнего источника («Вставить все записи из лога в базу данных»). Изучение регистра WordCount (Определение класса сопоставления, класс редуктора. Технологии очистки данных

Максимальное количество баллов по практическим заданиям – 20

Раздел 2. Реализация облачных систем

Практическое задание 3 (10 баллов).

«Разработка веб приложения" Разработка простого веб - приложения для загрузки изображений в хранилище: подготовка приложения; загрузка и отображение изображений; удаление сущностей; копирование сущностей. Разработка примера, демонстрирующего основы работы с очередями, на примере рабочей и веб – ролей

Практическое задание 4. (10 баллов).

«Работа с компонентами MSC»

Архитектура сервиса. Описание и версии компонентов. Шаблоны и компоновка кластеров. Apache Hive — система хранилища данных для Hadoop. Hadoop MapReduce - фреймворк для создания приложений, обрабатывающих большие объемы данных в парадигме MapReduce. Apache Spark - платформа для параллельной обработки данных в памяти. Apache Pig — платформа для обработки больших объемов данных с помощью скриптов.

Максимальное количество баллов по практическим заданиям – 20

Перечень вопросов для устного опроса

Раздел 1. Параллельная обработка больших массивов данных

1. Модель программирования MapReduce.
2. Принципы параллельной реализации вычислений. Область применения и примеры задач.
3. Принципы распределенной реализации MapReduce на кластерных системах.
4. Платформа Apache Hadoop.
5. Интерфейсы прикладного программирования и реализация программ для Hadoop.
6. Локальная отладка и запуск программ на кластере.
7. Приемы и стратегии реализации MapReduce-программ.
8. Высокоуровневые языки и инструментарии для работы с Hadoop.
9. Практические примеры использования MapReduce.
10. Ограничения модели MapReduce, расширения и альтернативные подходы.
11. Система Apache Spark.
12. Понятие Resilient Distributed Dataset (RDD), операции над RDD.
13. Принципы распределенного выполнения Spark-программы на кластере.
14. Запуск, отладка и оптимизация Spark-программ.
15. Infra: виртуальные машины и дополнительные инструменты их настройки.

Критерии оценивания:

Для каждого вопроса:

- 1 балл дан полный, развернутый ответ на поставленный вопрос, изложение материала при ответе – грамотное и логически стройное;
- 0 баллов – обучающийся не владеет материалом по заданному вопросу.

Максимальное количество баллов – 10

Раздел 2. Реализация облачных систем

1. Azure Службы, доступные в Azure, в том числе службы вычислений, сетевые службы, службы хранения и базы данных.
2. Сетевые ресурсы Azure,
3. VPN-шлюзы
4. Azure ExpressRoute.
5. Общие сведения о службах хранилища Azure
6. Хранилище BLOB-объектов Azure
7. Хранилище дисков Azure
8. Хранилище файлов Azure
9. Cloud storage: объектное хранилище с поддержкой S3 API.
10. Графические адаптеры: облачные вычисления на базе GPU.
11. Containers: контейнеры Kubernetes в облаке.
12. Big data: облачная обработка больших данных

Критерии оценивания:

Для каждого вопроса:

- 1 балл дан полный, развернутый ответ на поставленный вопрос, изложение материала при ответе – грамотное и логически стройное;
- 0 баллов – обучающийся не владеет материалом по заданному вопросу.

Максимальное количество баллов – 10

3 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

Промежуточная аттестация проводится в форме зачета

Зачет проводится по окончании обучения до начала экзаменационной сессии. Количество вопросов в задании – 3 (2 теоретических вопроса и 1 практико-ориентированное задание). Проверка ответов и объявление результатов производится в день зачета. Результаты аттестации заносятся в зачетную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лекции;
- практические работы;
- лабораторные работы.

В ходе лабораторных и практических работ развиваются навыки практического применения технологий параллельной обработки данных для решения задач профессиональной деятельности.

При подготовке к лабораторным занятиям каждый студент должен:

- изучить рекомендованную учебную литературу;

В процессе подготовки к лабораторным и практическим занятиям студенты могут воспользоваться консультациями преподавателя.

Теоретические вопросы должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется методом устного опроса или выполнения практико-ориентированных заданий. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу по изучаемой теме. Выделить непонятные термины, найти их значение в литературе.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.