

Рабочая программа дисциплины

Практикум по большим данным

<i>Направление подготовки</i>	Информационные системы и технологии
<i>Код</i>	09.03.02
<i>Направленность (профиль)</i>	Информационные технологии и большие данные
<i>Квалификация (степень) выпускника</i>	бакалавр

1. Перечень кодов компетенций, формируемых дисциплиной в процессе освоения образовательной программы

Группа компетенций	Категория компетенций	Код
Профессиональные	-	ПК-1
Профессиональные	-	ПК-4

2. Компетенции и индикаторы их достижения

Код компетенции	Формулировка компетенции	Индикаторы достижения компетенции
ПК-1	ПК-1 Способен использовать математический аппарат, методологию программирования и современные компьютерные технологии для решения практических задач получения, хранения, обработки и передачи информации.	ПК-1.1. Подбирает парадигму программирования под решение конкретной прикладной задачи; ПК-1.2. Модифицирует стандартные алгоритмы обработки информации для оптимизации решения прикладных задач. ПК-1.6. Создает модели основных объектов изучения естественнонаучных дисциплин и реализовывать их в компьютерных моделях.
ПК-4	ПК-4. Способен применять методы и средства проектирования программного обеспечения и баз данных	ПК-4.2. Сбор данных о запросах и потребностях заказчика применительно к типовой ИС. Документирование собранных данных в соответствии с регламентами организации. ПК-4.4. Разработка прототипов ИС на базе типовой ИС в соответствии с требованиями. ПК-4.5. Интеграция ИС с существующими ИС заказчика

3. Описание планируемых результатов обучения по дисциплине

3.1. Описание планируемых результатов обучения по дисциплине

Планируемые результаты обучения по дисциплине представлены дескрипторами (знания, умения, навыки).

Дескрипторы по дисциплине	Знать	Уметь	Владеть
Код компетенции	ПК-1		
	- математический аппарат, методологию программирования и современные	- использовать математический аппарат, методологию программирования и	- навыками использования математического аппарата, методологии

	компьютерные технологии для решения практических задач получения, хранения, обработки и передачи информации; - основные идеи, понятия и методы, определяющие стиль написания, отладки и сопровождения программ; - характеристики основных парадигм программирования;	современные компьютерные технологии для решения практических задач получения, хранения, обработки и передачи информации; - применять современные компьютерные технологии для решения практических задач; - делать обоснованный выбор инструментария для решения прикладных задач;	программирования и современных компьютерных технологий для решения практических задач получения, хранения, обработки и передачи информации; - математическим аппаратом для построения вычислительных моделей практических задач; - навыками использования стандартных алгоритмических моделей для решения задач хранения и обработки информации
Код компетенции	ПК-4		
	- методы оценки объемов и сроков выполнения работ, технологии выполнения работ в организации; - архитектуру, устройство и функционирование вычислительных систем, коммуникационное оборудование, сетевые протоколы; - основы современных операционных систем, основы современных систем управления базами данных, устройство и функционирование современных ИС; теорию баз данных, системы хранения и	- работать с современными системами программирования, конструировать программное обеспечение и базы данных, разрабатывать основные программные документы; - оценивать объемы и сроки выполнения работ; - разрабатывать руководство программиста к модифицированным элементам типовой ИС; - разрабатывать руководство администратора к модифицированным элементам типовой ИС; - разрабатывать руководство	- навыками конструирования программного обеспечения и баз данных; - навыками разработки интерфейсов обмена данными, форматов обмена данными, технологий обмена данными между ИС и существующими системами в соответствии с техническим заданием.

	анализа баз данных; - основы программирования, современные объектно-ориентированные языки программирования, современные структурные языки программирования, языки современных бизнес-приложений; - современные методики тестирования разрабатываемых ИС, современные стандарты информационного взаимодействия систем; - программные средства и платформы инфраструктуры информационных технологий организаций; - современные подходы и стандарты автоматизации организации (например, CRM, MRP, ERP, ITIL, ITSM); - основы теории систем и системного анализа; - методики описания и моделирования бизнес-процессов, средства моделирования бизнес-процессов; - системы классификации и кодирования информации, в том числе присвоение кодов документам и элементам	пользователя к модифицированным элементам типовой ИС; - разрабатывать ТЗ.	
--	--	--	--

	справочников; - отраслевую нормативную техническую документацию.		
--	--	--	--

4. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Практикум по большим данным» относится к части, формируемой участниками образовательных отношений учебного плана ОПОП.

Данная дисциплина взаимосвязана с другими дисциплинами, такими как: «Математический анализ», «Линейная алгебра», «Алгоритмизация и методы программирования», «Математическая логика и дискретная математика», «Дифференциальные уравнения», «Теория вероятностей и математическая статистика», «Численные методы», «Системы искусственного интеллекта», «Базы данных», «Машинное обучение».

В рамках освоения программы бакалавриата выпускники готовятся к решению задач профессиональной деятельности следующих типов:

Профиль (направленность) программы установлена путем ее ориентации на сферу профессиональной деятельности выпускников: Информационные технологии и большие данные

5. Объем дисциплины

<i>Виды учебной работы</i>	<i>Формы обучения</i>
	<i>Очно-заочное</i>
Общая трудоемкость: зачетные единицы/часы	4/144
Контактная работа:	
Занятия лекционного типа	12
Занятия семинарского типа	16
Промежуточная аттестация: зачет	0,1
Самостоятельная работа (СРС)	115,9

6. Содержание дисциплины (модуля), структурированное по темам / разделам с указанием отведенного на них количества академических часов и видов учебных занятий

6.1. Распределение часов по разделам/темам и видам работы

6.1.1. Очно-заочная форма обучения

№ п/п	Раздел/тема	Виды учебной работы (в часах)					
		Контактная работа				Самосто ятельная работа	
		Занятия лекционного типа		Занятия семинарского типа			
		Лекции	Иные учебные занятия	Практи ческие занятия	Семи нары		Лабора торные работы

1.	Основы Big Data: концепции, архитектура и экосистема	2			3			30
2.	Инфраструктура хранения и распределённые файловые системы	2			3			30
3.	Обработка и анализ больших данных с Apache Spark	2			3			30
4.	Машинное обучение и аналитика на больших данных	3			3			30
5.	Практические кейсы, безопасность и этика работы с Big Data	3			4			30,9
	Промежуточная аттестация	0,1						
	Итого	12			16			115,9

6.2 Программа дисциплины, структурированная по темам / разделам

6.2.1 Содержание лекционного курса

№ п/п	Наименование раздела дисциплины	Содержание раздела дисциплины
1.	Основы Big Data: концепции, архитектура и экосистема	определение больших данных, 3V-концепция (Volume, Velocity, Variety) и дополнительные характеристики (Veracity, Value, Variability); источники больших данных: IoT, социальные сети, логи серверов, финансовые транзакции, телеметрия; архитектура Big Data-систем: слои сбора, хранения, обработки, аналитики и визуализации; экосистема инструментов: Hadoop, Spark, Kafka, NoSQL-базы; облачные платформы для Big Data: AWS, GCP, Azure (сервисы EMR, Dataproc, HDInsight); кейсы применения Big Data в разных отраслях: ритейл, финансы, здравоохранение, промышленность.
2.	Инфраструктура хранения и распределённые файловые системы	распределённые файловые системы: HDFS (архитектура NameNode/DataNode, репликация, блоки данных); NoSQL-базы данных: типы (ключ-значение, документо-ориентированные, колоночные, графовые), примеры (MongoDB, Cassandra, HBase); форматы хранения данных: Parquet, ORC, Avro — преимущества для Big Data;

		Data Lakes vs Data Warehouses: сравнение концепций, сценариев использования; управление метаданными: Apache Hive, Metastore; практические задания: развёртывание HDFS-кластера, загрузка данных, работа с Hive.
3.	Обработка и анализ больших данных с Apache Spark	архитектура Spark: драйвер, исполнители, кластерный менеджер (Standalone, YARN, Mesos); RDD (Resilient Distributed Dataset): создание, трансформации и действия, ленивые вычисления; DataFrames и Spark SQL: работа с табличными данными, выполнение SQL-запросов; Spark Streaming: обработка потоковых данных в режиме near real-time; оптимизация Spark-приложений: партиционирование, кэширование, сериализация (Kryo); интеграция Spark с другими компонентами: Kafka, HDFS, NoSQL; практическое задание: анализ логов веб-сервера с помощью Spark SQL.
4.	Машинное обучение и аналитика на больших данных	особенности применения ML к большим данным: масштабируемость, распределённое обучение; Spark MLlib: алгоритмы классификации, регрессии, кластеризации, отбора признаков; предобработка больших данных для ML: масштабирование, кодирование категориальных признаков, работа с пропусками; оценка моделей на больших выборках: кросс-валидация, метрики качества (accuracy, precision, recall, RMSE); онлайн-обучение и инкрементальное обучение для потоковых данных; библиотеки для распределённого ML: Dask-ML, TensorFlow с распределённым режимом; кейс: прогнозирование спроса на основе больших исторических данных.
5.	Практические кейсы, безопасность и этика работы с Big Data	кейс 1: анализ пользовательского поведения в e-commerce (сегментация, рекомендации); кейс 2: прогнозирование нагрузки на IT-инфраструктуру по логам серверов; кейс 3: обработка и анализ данных IoT-устройств (датчики, телеметрия); кейс 4: мониторинг социальных сетей: sentiment analysis на больших данных; безопасность данных: шифрование (в покое и при передаче), контроль доступа (RBAC), аудит; этические и юридические аспекты: GDPR, ФЗ-152, анонимизация данных (k-anonymity, differential privacy); современные тенденции: федеративное обучение, приватный анализ данных, AutoML для Big Data.

6.2.2 Содержание практических занятий

№ п/п	Наименование раздела дисциплины	Содержание практических занятий
1.	Основы Big Data: концепции, архитектура и экосистема	определение больших данных, 3V-концепция (Volume, Velocity, Variety) и дополнительные характеристики (Veracity, Value, Variability); источники больших данных: IoT, социальные сети, логи серверов, финансовые транзакции, телеметрия; архитектура Big Data-систем: слои сбора, хранения, обработки, аналитики и визуализации; экосистема инструментов: Hadoop, Spark, Kafka, NoSQL-базы; облачные платформы для Big Data: AWS, GCP, Azure (сервисы EMR, Dataproc, HDInsight); кейсы применения Big Data в разных отраслях: ритейл, финансы, здравоохранение, промышленность.
2.	Инфраструктура хранения и распределённые файловые системы	распределённые файловые системы: HDFS (архитектура NameNode/DataNode, репликация, блоки данных); NoSQL-базы данных: типы (ключ-значение, документо-ориентированные, колоночные, графовые), примеры (MongoDB, Cassandra, HBase); форматы хранения данных: Parquet, ORC, Avro — преимущества для Big Data; Data Lakes vs Data Warehouses: сравнение концепций, сценариев использования; управление метаданными: Apache Hive, Metastore; практические задания: развёртывание HDFS-кластера, загрузка данных, работа с Hive.
3.	Обработка и анализ больших данных с Apache Spark	архитектура Spark: драйвер, исполнители, кластерный менеджер (Standalone, YARN, Mesos); RDD (Resilient Distributed Dataset): создание, трансформации и действия, ленивые вычисления; DataFrames и Spark SQL: работа с табличными данными, выполнение SQL-запросов; Spark Streaming: обработка потоковых данных в режиме near real-time; оптимизация Spark-приложений: партиционирование, кэширование, сериализация (Kryo); интеграция Spark с другими компонентами: Kafka, HDFS, NoSQL; практическое задание: анализ логов веб-сервера с помощью Spark SQL.
4.	Машинное обучение и аналитика на больших данных	особенности применения ML к большим данным: масштабируемость, распределённое обучение; Spark MLlib: алгоритмы классификации, регрессии, кластеризации, отбора признаков;

		предобработка больших данных для ML: масштабирование, кодирование категориальных признаков, работа с пропусками; оценка моделей на больших выборках: кросс-валидация, метрики качества (accuracy, precision, recall, RMSE); онлайн-обучение и инкрементальное обучение для потоковых данных; библиотеки для распределённого ML: Dask-ML, TensorFlow с распределённым режимом; кейс: прогнозирование спроса на основе больших исторических данных.
5.	Практические кейсы, безопасность и этика работы с Big Data	кейс 1: анализ пользовательского поведения в e-commerce (сегментация, рекомендации); кейс 2: прогнозирование нагрузки на IT-инфраструктуру по логам серверов; кейс 3: обработка и анализ данных IoT-устройств (датчики, телеметрия); кейс 4: мониторинг социальных сетей: sentiment analysis на больших данных; безопасность данных: шифрование (в покое и при передаче), контроль доступа (RBAC), аудит; этические и юридические аспекты: GDPR, ФЗ-152, анонимизация данных (k-anonymity, differential privacy); современные тенденции: федеративное обучение, приватный анализ данных, AutoML для Big Data.

6.2.3 Содержание самостоятельной работы

№ п/п	Наименование раздела дисциплины	Содержание самостоятельной работы
1.	Основы Big Data: концепции, архитектура и экосистема	определение больших данных, 3V-концепция (Volume, Velocity, Variety) и дополнительные характеристики (Veracity, Value, Variability); источники больших данных: IoT, социальные сети, логи серверов, финансовые транзакции, телеметрия; архитектура Big Data-систем: слои сбора, хранения, обработки, аналитики и визуализации; экосистема инструментов: Hadoop, Spark, Kafka, NoSQL-базы; облачные платформы для Big Data: AWS, GCP, Azure (сервисы EMR, Dataproc, HDInsight); кейсы применения Big Data в разных отраслях: ритейл, финансы, здравоохранение, промышленность.
2.	Инфраструктура хранения и распределённые файловые системы	распределённые файловые системы: HDFS (архитектура NameNode/DataNode, репликация, блоки данных);

		NoSQL-базы данных: типы (ключ-значение, документо-ориентированные, колоночные, графовые), примеры (MongoDB, Cassandra, HBase); форматы хранения данных: Parquet, ORC, Avro — преимущества для Big Data; Data Lakes vs Data Warehouses: сравнение концепций, сценариев использования; управление метаданными: Apache Hive, Metastore; практические задания: развёртывание HDFS-кластера, загрузка данных, работа с Hive.
3.	Обработка и анализ больших данных с Apache Spark	архитектура Spark: драйвер, исполнители, кластерный менеджер (Standalone, YARN, Mesos); RDD (Resilient Distributed Dataset): создание, трансформации и действия, ленивые вычисления; DataFrames и Spark SQL: работа с табличными данными, выполнение SQL-запросов; Spark Streaming: обработка потоковых данных в режиме near real-time; оптимизация Spark-приложений: партиционирование, кэширование, сериализация (Kryo); интеграция Spark с другими компонентами: Kafka, HDFS, NoSQL; практическое задание: анализ логов веб-сервера с помощью Spark SQL.
4.	Машинное обучение и аналитика на больших данных	особенности применения ML к большим данным: масштабируемость, распределённое обучение; Spark MLlib: алгоритмы классификации, регрессии, кластеризации, отбора признаков; предобработка больших данных для ML: масштабирование, кодирование категориальных признаков, работа с пропусками; оценка моделей на больших выборках: кросс-валидация, метрики качества (accuracy, precision, recall, RMSE); онлайн-обучение и инкрементальное обучение для потоковых данных; библиотеки для распределённого ML: Dask-ML, TensorFlow с распределённым режимом; кейс: прогнозирование спроса на основе больших исторических данных.
5.	Практические кейсы, безопасность и этика работы с Big Data	кейс 1: анализ пользовательского поведения в e-commerce (сегментация, рекомендации); кейс 2: прогнозирование нагрузки на IT-инфраструктуру по логам серверов; кейс 3: обработка и анализ данных IoT-устройств (датчики, телеметрия); кейс 4: мониторинг социальных сетей: sentiment analysis на больших данных; безопасность данных: шифрование (в покое и при передаче), контроль доступа (RBAC), аудит;

		этические и юридические аспекты: GDPR, ФЗ-152, анонимизация данных (k-anonymity, differential privacy); современные тенденции: федеративное обучение, приватный анализ данных, AutoML для Big Data.
--	--	--

7. Текущий контроль по дисциплине (модулю) в рамках учебных занятий

В рамках текущего контроля преподаватель самостоятельно может проводить следующие мероприятия:

№ п/п	Контролируемые разделы (темы)	Наименование оценочного средства
1.	Основы Big Data: концепции, архитектура и экосистема	Опрос, информационный проект.
2.	Инфраструктура хранения и распределённые файловые системы	Опрос, информационный проект, тестирование.
3.	Обработка и анализ больших данных с Apache Spark	Опрос, информационный проект.
4.	Машинное обучение и аналитика на больших данных	Опрос, информационный проект.
5.	Практические кейсы, безопасность и этика работы с Big Data	Опрос, информационный проект, тестирование.

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

8.1. Основная литература:

- Монгуш, Ч. М. Введение в анализ данных : учебное пособие для студентов / Ч. М. Монгуш. — Кызыл : Издательство Тувинского государственного университета, 2022. — 51 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/149460.html>
- Татарникова, Т. М. Интеллектуальный анализ данных : учебное пособие / Т. М. Татарникова. — Москва, Вологда : Инфра-Инженерия, 2024. — 172 с. — ISBN 978-5-9729-1772-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/143351.html>
- Зуев, С. В. Методы анализа данных : учебное пособие / С. В. Зуев. — Москва : Ай Пи Ар Медиа, 2023. — 132 с. — ISBN 978-5-4497-2107-5. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/129059.html>
- Маккинли, У. Python и анализ данных / У. Маккинли ; перевод А. Слинкина. — 3-е изд. — Саратов : Профобразование, 2024. — 482 с. — ISBN 978-5-4488-0046-7. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/145897.html>

8.2. Дополнительная литература:

- Маккини, У. Python и анализ данных / У. Маккини ; перевод А. А. Слинкин. — 2-е изд. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст :

электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/125361.html>

2. Бюиссон, Ф. Анализ поведенческих данных на R и Python / Ф. Бюиссон ; перевод А. В. Логунов. — Москва : ДМК Пресс, 2022. — 368 с. — ISBN 978-5-97060-992-7. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/125286.html>

8.3. Периодические издания

1. Вестник Московского государственного технического университета имени Н.Э. Баумана. Серия Естественные науки. ISSN 1812-3368. <https://www.iprbookshop.ru/23124.html>
2. Открытые Системы. СУБД. ISSN 1028-7493. <https://www.iprbookshop.ru/76383.html>
3. Информационные технологии моделирования и управления. ISSN 1813-9744. <https://www.iprbookshop.ru/43350.html>

9. Перечень ресурсов информационно-телекоммуникационной сети "Интернет" (далее - сеть "Интернет"), необходимых для освоения дисциплины (модуля)

1. Федеральный портал «Российское образование» <http://www.edu.ru>
2. Электронно-библиотечная система «Научная электронная библиотека eLIBRARY.RU» <https://www.elibrary.ru>
3. Электронно-библиотечная система IPR BOOKS <https://www.iprbookshop.ru>
4. Электронная библиотечная система <https://biblioclub.ru>

10. Методические указания для обучающихся по освоению дисциплины (модуля)

Успешное освоение данного курса базируется на рациональном сочетании нескольких видов учебной деятельности – лекций, семинарских занятий, самостоятельной работы. При этом самостоятельную работу следует рассматривать одним из главных звеньев полноценного высшего образования, на которую отводится значительная часть учебного времени.

Самостоятельная работа студентов складывается из следующих составляющих:

1. Работа с основной и дополнительной литературой, с материалами интернета и конспектами лекций;
2. Внеаудиторная подготовка к контрольным работам, выполнение докладов, рефератов и курсовых работ;
3. Выполнение самостоятельных практических работ;
4. Подготовка к экзаменам (зачетам) непосредственно перед ними.

Для правильной организации работы необходимо учитывать порядок изучения разделов курса, находящихся в строгой логической последовательности. Поэтому хорошее усвоение одной части дисциплины является предпосылкой для успешного перехода к следующей. Задания, проблемные вопросы, предложенные для изучения дисциплины, в том числе и для самостоятельного выполнения, носят междисциплинарный характер и базируются, прежде всего, на причинно-следственных связях между компонентами окружающего нас мира. Важным составляющим в изучении данного курса является решение ситуационных задач и работа над проблемно-аналитическими заданиями, что предполагает знание соответствующей научной терминологии и т.д.

Для лучшего запоминания материала целесообразно использовать индивидуальные особенности и разные виды памяти: зрительную, слуховую, ассоциативную. Успешному запоминанию также способствует приведение ярких свидетельств и наглядных примеров. Учебный материал должен постоянно повторяться и закрепляться.

При выполнении докладов, творческих, информационных, исследовательских проектов особое внимание следует обращать на подбор источников информации и методику работы с ними.

Для успешной сдачи экзамена (зачета) рекомендуется соблюдать следующие правила:

1. Подготовка к экзамену (зачету) должна проводиться систематически, в течение всего семестра.
2. Интенсивная подготовка должна начаться не позднее, чем за месяц до экзамена.
3. Время непосредственно перед экзаменом (зачетом) лучше использовать таким образом, чтобы оставить последний день свободным для повторения курса в целом, для систематизации материала и доработки отдельных вопросов.

На экзамене высокую оценку получают студенты, использующие данные, полученные в процессе выполнения самостоятельных работ, а также использующие собственные выводы на основе изученного материала.

Учитывая значительный объем теоретического материала, студентам рекомендуется регулярное посещение и подробное конспектирование лекций.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

1. Microsoft Windows Server;
2. Семейство ОС Microsoft Windows;
3. Libre Office свободно распространяемый офисный пакет с открытым исходным кодом;
4. Информационно-справочная система: Система КонсультантПлюс (КонсультантПлюс);
5. Информационно-правовое обеспечение Гарант: Электронный периодический справочник «Система ГАРАНТ» (Система ГАРАНТ);

Перечень используемого программного обеспечения указан в п.12 данной рабочей программы дисциплины.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

12.1. Учебная аудитория для проведения учебных занятий, предусмотренных образовательной программой, оснащенная оборудованием и техническими средствами обучения.

Специализированная мебель:

Комплект учебной мебели (стол, стул) по количеству обучающихся; комплект мебели для преподавателя; доска (маркерная).

Технические средства обучения:

Компьютер в сборе для преподавателя, проектор, экран, колонки

Перечень лицензионного программного обеспечения, в том числе отечественного производства:

Windows 10, КонсультантПлюс, Kaspersky Endpoint Security.

Перечень свободно распространяемого программного обеспечения:

Яндекс Браузер, LibreOffice, МТС Линк.

Подключение к сети «Интернет» и обеспечение доступа в электронную информационно-образовательную среду ММУ.

12.2. Помещение для самостоятельной работы обучающихся.

Специализированная мебель:

Комплект учебной мебели (стол, стул) по количеству обучающихся; комплект мебели для преподавателя; доска (маркерная).

Технические средства обучения:

Компьютер в сборе для преподавателя; компьютеры в сборе для обучающихся; колонки; проектор, экран.

Перечень лицензионного программного обеспечения, в том числе отечественного производства:

Windows Server 2016, Windows 10, Microsoft Office, КонсультантПлюс, Kaspersky Endpoint Security.

Перечень свободно распространяемого программного обеспечения:

Яндекс Браузер, LibreOffice, МТС Линк, Gimp, Paint.net, AnyLogic, Inkscape.

Помещение для самостоятельной работы обучающихся оснащено компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду ММУ.

13. Образовательные технологии, используемые при освоении дисциплины

Для освоения дисциплины используются как традиционные формы занятий – лекции (типы лекций – установочная, вводная, текущая, заключительная, обзорная; виды лекций – проблемная, визуальная, лекция конференция, лекция консультация); и семинарские (практические) занятия, так и активные и интерактивные формы занятий – деловые и ролевые игры, решение ситуационных задач и разбор конкретных ситуаций.

На учебных занятиях используются технические средства обучения мультимедийной аудитории: компьютер, монитор, колонки, настенный экран, проектор, микрофон, пакет программ Microsoft Office для демонстрации презентаций и медиафайлов, видеопроектор для демонстрации слайдов, видеосюжетов и др. Тестирование обучаемых может осуществляться с использованием компьютерного оборудования университета.

13.1. В освоении учебной дисциплины используются следующие традиционные образовательные технологии:

- чтение проблемно-информационных лекций с использованием доски и видеоматериалов;
- семинарские занятия для обсуждения, дискуссий и обмена мнениями;
- контрольные опросы;
- консультации;
- самостоятельная работа студентов с учебной литературой и первоисточниками;
- подготовка и обсуждение рефератов (проектов), презентаций (научно-исследовательская работа);
- тестирование по основным темам дисциплины.

13.2. Активные и интерактивные методы и формы обучения

Из перечня видов: («мозговой штурм», анализ НПА, анализ проблемных ситуаций, анализ конкретных ситуаций, инциденты, имитация коллективной профессиональной деятельности, разыгрывание ролей, творческая работа, связанная с освоением дисциплины, ролевая игра, круглый стол, диспут, беседа, дискуссия, мини-конференция и др.) используются следующие:

- диспут;
- анализ проблемных, творческих заданий, ситуационных задач;
- ролевая игра;
- круглый стол;
- мини-конференция;

- дискуссия;
- беседа.

13.3. Особенности обучения инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ)

При организации обучения по дисциплине учитываются особенности организации взаимодействия с инвалидами и лицами с ограниченными возможностями здоровья (далее – инвалиды и лица с ОВЗ) с целью обеспечения их прав. При обучении учитываются особенности их психофизического развития, индивидуальные возможности и при необходимости обеспечивается коррекция нарушений развития и социальная адаптация указанных лиц.

Выбор методов обучения определяется содержанием обучения, уровнем методического и материально-технического обеспечения, особенностями восприятия учебной информации студентами с инвалидностью и студентами с ограниченными возможностями здоровья и т.д. В образовательном процессе используются социально-активные и рефлексивные методы обучения, технологии социокультурной реабилитации с целью оказания помощи в установлении полноценных межличностных отношений с другими студентами, создании комфортного психологического климата в студенческой группе.

При обучении лиц с ограниченными возможностями здоровья электронное обучение и дистанционные образовательные технологии предусматривают возможность приема-передачи информации в доступных для них формах.

Обучающиеся из числа лиц с ограниченными возможностями здоровья обеспечены печатными и электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.

Автономная некоммерческая организация высшего образования
«МОСКОВСКИЙ МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ»

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ
ПО ДИСЦИПЛИНЕ**

Практикум по большим данным

<i>Направление подготовки</i>	Информационные системы и технологии
<i>Код</i>	09.03.02
<i>Направленность (профиль)</i>	Информационные технологии и большие данные
<i>Квалификация (степень) выпускника</i>	бакалавр

1. Перечень кодов компетенций, формируемых дисциплиной в процессе освоения образовательной программы

Группа компетенций	Категория компетенций	Код
Профессиональные	-	ПК-1
Профессиональные	-	ПК-4

2. Компетенции и индикаторы их достижения

Код компетенции	Формулировка компетенции	Индикаторы достижения компетенции
ПК-1	ПК-1 Способен использовать математический аппарат, методологию программирования и современные компьютерные технологии для решения практических задач получения, хранения, обработки и передачи информации.	ПК-1.1. Подбирает парадигму программирования под решение конкретной прикладной задачи; ПК-1.2. Модифицирует стандартные алгоритмы обработки информации для оптимизации решения прикладных задач. ПК-1.6. Создает модели основных объектов изучения естественнонаучных дисциплин и реализовывать их в компьютерных моделях.
ПК-4	ПК-4. Способен применять методы и средства проектирования программного обеспечения и баз данных	ПК-4.2. Сбор данных о запросах и потребностях заказчика применительно к типовой ИС. Документирование собранных данных в соответствии с регламентами организации. ПК-4.4. Разработка прототипов ИС на базе типовой ИС в соответствии с требованиями. ПК-4.5. Интеграция ИС с существующими ИС заказчика

3. Описание планируемых результатов обучения по дисциплине

3.1. Описание планируемых результатов обучения по дисциплине

Планируемые результаты обучения по дисциплине представлены дескрипторами (знания, умения, навыки).

Дескрипторы по дисциплине	Знать	Уметь	Владеть
Код компетенции	ПК-1		
	- математический аппарат, методологию программирования и современные	- использовать математический аппарат, методологию программирования и	- навыками использования математического аппарата, методологии

	компьютерные технологии для решения практических задач получения, хранения, обработки и передачи информации; - основные идеи, понятия и методы, определяющие стиль написания, отладки и сопровождения программ; - характеристики основных парадигм программирования;	современные компьютерные технологии для решения практических задач получения, хранения, обработки и передачи информации; - применять современные компьютерные технологии для решения практических задач; - делать обоснованный выбор инструментария для решения прикладных задач;	программирования и современных компьютерных технологий для решения практических задач получения, хранения, обработки и передачи информации; - математическим аппаратом для построения вычислительных моделей практических задач; - навыками использования стандартных алгоритмических моделей для решения задач хранения и обработки информации
Код компетенции	ПК-4		
	- методы оценки объемов и сроков выполнения работ, технологии выполнения работ в организации; - архитектуру, устройство и функционирование вычислительных систем, коммуникационное оборудование, сетевые протоколы; - основы современных операционных систем, основы современных систем управления базами данных, устройство и функционирование современных ИС; теорию баз данных, системы хранения и	- работать с современными системами программирования, конструировать программное обеспечение и базы данных, разрабатывать основные программные документы; - оценивать объемы и сроки выполнения работ; - разрабатывать руководство программиста к модифицированным элементам типовой ИС; - разрабатывать руководство администратора к модифицированным элементам типовой ИС; - разрабатывать руководство	- навыками конструирования программного обеспечения и баз данных; - навыками разработки интерфейсов обмена данными, форматов обмена данными, технологий обмена данными между ИС и существующими системами в соответствии с техническим заданием.

	анализа баз данных; - основы программирования, современные объектно-ориентированные языки программирования, современные структурные языки программирования, языки современных бизнес-приложений; - современные методики тестирования разрабатываемых ИС, современные стандарты информационного взаимодействия систем; - программные средства и платформы инфраструктуры информационных технологий организаций; - современные подходы и стандарты автоматизации организации (например, CRM, MRP, ERP, ITIL, ITSM); - основы теории систем и системного анализа; - методики описания и моделирования бизнес-процессов, средства моделирования бизнес-процессов; - системы классификации и кодирования информации, в том числе присвоение кодов документам и элементам	пользователя к модифицированным элементам типовой ИС; - разрабатывать ТЗ.	
--	--	--	--

	справочников; - отраслевую нормативную техническую документацию.		
--	--	--	--

3.2. Критерии оценки знаний студентов

Шкала оценив ания	Индикаторы достижения	Показатели оценивания результатов обучения
ОТЛИЧНО/Зачтено	Знает:	- студент глубоко и всесторонне усвоил материал, уверенно, логично, последовательно и грамотно его излагает, опираясь на знания основной и дополнительной литературы, - на основе системных научных знаний делает квалифицированные выводы и обобщения, свободно оперирует категориями и понятиями.
	Умеет:	- студент умеет самостоятельно и правильно решать учебно-профессиональные задачи или задания, уверенно, логично, последовательно и аргументировано излагать свое решение, используя научные понятия, ссылаясь на нормативную базу.
	Владеет:	- студент владеет рациональными методами (с использованием рациональных методик) решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении продемонстрировал навыки - выделения главного, - связкой теоретических положений с требованиями руководящих документов, - изложения мыслей в логической последовательности, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
ХОРОШО/Зачтено	Знает:	- студент твердо усвоил материал, достаточно грамотно его излагает, опираясь на знания основной и дополнительной литературы, - затрудняется в формулировании квалифицированных выводов и обобщений, оперирует категориями и понятиями, но не всегда правильно их верифицирует.
	Умеет:	- студент умеет самостоятельно и в основном правильно решать учебно-профессиональные задачи или задания, уверенно, логично, последовательно и аргументировано излагать свое решение, не в полной мере используя научные понятия и ссылки на нормативную базу.
	Владеет:	- студент в целом владеет рациональными методами решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении смог продемонстрировать достаточность, но не глубинность навыков - выделения главного,

		- изложения мыслей в логической последовательности. - связки теоретических положений с требованиями руководящих документов, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
УДОВЛЕТВОРИТЕЛЬНО/Зачтено	Знает:	- студент ориентируется в материале, однако затрудняется в его изложении; - показывает недостаточность знаний основной и дополнительной литературы; - слабо аргументирует научные положения; - практически не способен сформулировать выводы и обобщения; - частично владеет системой понятий.
	Умеет:	- студент в основном умеет решить учебно-профессиональную задачу или задание, но допускает ошибки, слабо аргументирует свое решение, недостаточно использует научные понятия и руководящие документы.
	Владеет:	- студент владеет некоторыми рациональными методами решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении продемонстрировал недостаточность навыков - выделения главного, - изложения мыслей в логической последовательности. - связки теоретических положений с требованиями руководящих документов, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
НЕУДОВЛЕТВОРИТЕЛЬНО/Не зачтено	Знает:	- студент не усвоил значительной части материала; - не может аргументировать научные положения; - не формулирует квалифицированных выводов и обобщений; - не владеет системой понятий.
	Умеет:	студент не показал умение решать учебно-профессиональную задачу или задание.
	Владеет:	не выполнены требования, предъявляемые к навыкам, оцениваемым “удовлетворительно”.

При ответе на вопросы в рамках прохождения промежуточной аттестации (зачет/ зачет с оценкой/ экзамен) допускается вольная формулировка ответа, по смыслу раскрывающая содержание ответа, указанного в фонде оценочных средств, в качестве верного ответа.

4. Типовые контрольные задания (закрытого, открытого и иного типа) для проведения промежуточной аттестации, необходимые для оценки достижения компетенции, соотнесенной с результатами обучения по дисциплине

ТЕСТИРОВАНИЕ

8 СЕМЕСТР

ПК-1

1. Что означает характеристика «Velocity» в контексте Big Data?
Ответ: высокая скорость поступления и обработки данных.
2. Какой компонент архитектуры Big Data отвечает за сбор данных из различных источников?
Ответ: слой сбора данных (Ingestion Layer).
3. Какой сервис AWS используется для запуска Hadoop и Spark-кластеров?
Ответ: Amazon EMR.
4. Что такое Data Lake?
Ответ: хранилище необработанных данных в исходном формате без предварительной структуризации.
5. Какой пример иллюстрирует характеристику «Variety» в Big Data?
Ответ: сочетание логов серверов, текстовых сообщений из соцсетей и изображений с камер наблюдения.
6. Какую роль выполняет NameNode в HDFS?
Ответ: управляет файловой системой, хранит метаданные и отслеживает расположение блоков данных.
7. Какой тип NoSQL-базы подходит для хранения JSON-документов?
Ответ: документо-ориентированная (например, MongoDB).
8. Какой формат данных оптимизирован для колоночного хранения и сжатия в Big Data?
Ответ: Parquet.
9. В чём ключевое отличие Data Lake от Data Warehouse?
Ответ: Data Lake хранит необработанные данные любого формата, Data Warehouse — структурированные данные для аналитики.
10. Для чего используется Apache Hive?
Ответ: для выполнения SQL-подобных запросов к данным в HDFS.
11. Что такое RDD в Apache Spark?
Ответ: Resilient Distributed Dataset — отказоустойчивый распределённый набор данных.
12. Какой компонент Spark отвечает за распределение задач между исполнителями?
Ответ: драйвер (Driver Program).
13. Для чего применяется кэширование в Spark?
Ответ: для сохранения промежуточных результатов в памяти и ускорения последующих операций.
14. Какой инструмент Spark используется для обработки потоковых данных?
Ответ: Spark Streaming.
15. Что такое ленивые вычисления в Spark?
Ответ: операции (трансформации) не выполняются сразу, а планируются и запускаются только при вызове действия (action).

8 СЕМЕСТР

ПК-4

1. Какой инструмент позволяет распределённо обучать ML-модели на Spark?
Ответ: Spark MLlib.
2. Какой метод отбора признаков подходит для больших наборов данных из-за низкой вычислительной сложности?
Ответ: фильтрационные методы (например, корреляция с целевой переменной).
3. Какая метрика качества чаще всего используется для оценки регрессионных моделей на больших данных?
Ответ: RMSE (Root Mean Squared Error).
4. Что такое онлайн-обучение в контексте Big Data?
Ответ: подход, при котором модель обновляется последовательно по мере поступления новых данных, без полного переобучения.
5. Какой фреймворк подходит для распределённого обучения нейронных сетей на больших данных?
Ответ: TensorFlow с распределённым режимом или PyTorch DDP.
6. Какой кейс иллюстрирует применение Big Data в e-commerce?
Ответ: построение рекомендательной системы на основе истории покупок и поведения пользователей.
7. Какой метод анонимизации данных добавляет статистический шум для защиты приватности?
Ответ: differential privacy.
8. Какой нормативный акт регулирует обработку персональных данных в РФ?
Ответ: Федеральный закон № 152-ФЗ «О персональных данных».
9. Что такое федеративное обучение?
Ответ: метод обучения моделей на распределённых данных без их централизации, используется для сохранения приватности.
10. Какой инструмент подходит для мониторинга распределённой Big Data-системы в реальном времени?
Ответ: Prometheus + Grafana.
11. Какой подход используется для прогнозирования нагрузки на серверы по логам?
Ответ: анализ временных рядов с помощью ML-моделей.
12. Что такое sentiment analysis в контексте Big Data?
Ответ: анализ тональности текстовых данных (положительных/отрицательных отзывов) на больших наборах.
13. Какой протокол обеспечивает шифрование данных при передаче в Big Data-пайплайне?
Ответ: TLS/SSL.
14. Какой метод контроля доступа ограничивает права пользователей на основе ролей?
Ответ: RBAC (Role-Based Access Control).
15. Какой инструмент используется для сбора и передачи потоковых данных между компонентами системы?
Ответ: Apache Kafka.