

АННОТАЦИЯ рабочей программы дисциплины Б1.О.40 «Технологии обработки больших данных»

Объем трудоемкости: 3 зач.ед. (108 часов)

Цели и задачи изучения дисциплины (модуля)

Цель освоения дисциплины

Цель дисциплины - Изучение принципов обработки больших данных, технологий распределенных вычислений, облачных платформ и инструментов анализа данных.

1.2 Задачи дисциплины

- Изучение архитектурных решений для работы с Big Data.
- Освоение методов обработки структурированных и неструктурированных данных.
- Применение распределенных вычислений (Hadoop, Spark).
- Разработка алгоритмов анализа данных в распределенных средах.
- Использование облачных платформ для обработки больших данных.

Требования к знаниям и навыкам:

- Умение работать с распределенными системами (Hadoop, Spark).
- Опыт обработки данных в облачных средах (Yandex Cloud).
- Навыки анализа данных с помощью Python (Pandas, PySpark, Dask).
- Понимание архитектуры Big Data-решений.

Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Технологии обработки больших данных» относится к Блок 1 Дисциплины, Обязательная часть учебного плана.

Входными знаниями (пререквизиты) для освоения данной дисциплины являются знания, умения и опыт, накопленный студентами в процессе изучения дисциплины «Базы данных», «Технологии управления данными в NoSQL», «Машинное обучение», «Обработка данных на Python».

Постреквизиты: DevOps / MLOps, Глубокое обучение, Современные методы фронтирных исследований ИИ, Методы обучения с подкреплением, Технологии обработки языка, звуковых данных, включая распознавание и синтез речи, Генеративный искусственный интеллект.

Смежные: Облачные технологии и бэкэндразработка, Технологии управления данными NoSQL.

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и содержание компетенции	Общий индикатор	Уровни освоения индикаторов компетенции
1	2	3
ОПК-2 Способен применять современный математический аппарат, связанный с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности	ОПК-2.1 Способен применять системный подход к анализу предметной (проблемной) области, выявлению требований к ИС	Знать: методы системного анализа, этапы выявления требований к информационным системам (ИС), современный математический аппарат для проектирования программных продуктов. Уметь: применять системный подход для анализа предметной области, формализовать требования к ИС, выбирать адекватные математические модели для оценки качества программных продуктов.

Код и содержание компетенции	Общий индикатор	Уровни освоения индикаторов компетенции
1	2	3
		Владеть: навыками применения математического аппарата для разработки и оценки качества программных продуктов и комплексов.
BD-3 Способен организовывать хранения данных выбирая адекватные технологические решения	BD-3.2 Разрабатывает, отлаживает и тестирует прикладные решения с элементами ИИ с применением различных технологий хранения неструктурированных данных, оценивает качество.	Знать: основные типы NoSQL-хранилищ (Ключ-Значение, Документные, Колоночные, Графовые); базовые команды для работы с данными в таких хранилищах. Уметь: создавать базы данных в указанных хранилищах, выполнять CRUD-операции, использовать индексы; разрабатывать, отлаживать и тестировать прикладные решения с элементами ИИ. Владеть: технологиями хранения неструктурированных данных на уровне применения наиболее известных подходов каждого класса хранилищ; оценивать качество решений.
BD-4 Способен применять различные модели и (или) технологии обработки данных	BD-4.1 Осуществляет выбор технологий обработки больших данных, приемлемых для создания прикладной системы ИИ с заданными требованиями	Знать: современные технологии распределённого хранения и параллельной обработки (Hadoop, Spark); архитектурные принципы Big Data-решений. Уметь: организовывать распределённое хранилище и параллельную обработку на базе Hadoop/Spark; осуществлять выбор технологий под заданные требования. Владеть: навыками развёртывания и настройки кластеров (в т.ч. в облачных средах), оптимизации вычислительных процессов.
	BD-4.2 Разрабатывает и отлаживает прикладные решения с элементами ИИ с применением технологий обработки данных	Знать: методы и алгоритмы обработки данных в распределённых средах (ETL, агрегация, анализ). Уметь: разрабатывать и отлаживать прикладные решения с элементами ИИ с использованием технологий обработки данных (Spark, SQL, потоковая обработка). Владеть: инструментарием для реализации полных циклов обработки данных (от загрузки до визуализации).
BD-5 Способен применять технологии организации инфраструктуры БД	BD-5.1 Осуществляет выбор направления вспомогательных технологических решений для формирования единого стека работы с большими данными для решения поставленной задачи	Знать: основные компоненты инфраструктуры для работы с большими данными (хранилища, очереди, оркестрация). Уметь: выполнять отдельные функции в проектах по созданию инфраструктуры БД; выбирать и

Код и содержание компетенции	Общий индикатор	Уровни освоения индикаторов компетенции
1	2	3
		обосновывать вспомогательные технологические решения. Владеть: практическими навыками настройки, мониторинга и интеграции элементов инфраструктуры (кластеры, облачные сервисы).
	BD-5.2 Разрабатывает и отлаживает прикладные решения с элементами ИИ с применением технологий организации БД и инфраструктуры	Знать: различные технологии организации баз данных и инфраструктуры (NoSQL, объектные хранилища, репликация, шардинг). Уметь: участвовать в разработке решений с элементами ИИ, применяя разные технологии организации инфраструктуры БД. Владеть: методами интеграции прикладных решений с элементами ИИ в существующую инфраструктуру обработки данных.
LC-8 Способен применять знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска перспективных способов решения проблем с помощью ИИ	LC-8.2 Определяет тенденции развития, оценивает новизну и практическую значимость своих решений с точки зрения современного искусственного интеллекта	Знать: причины появления концепции больших данных, разницу определений; категории проблем больших данных; историю и динамику появления новых технологий. Уметь: выявлять различные категории проблем больших данных с примерами; анализировать динамику появления технологий; сопоставлять собственные решения с современными исследованиями и промышленными стандартами; определять тенденции развития. Владеть: навыками критической оценки новизны и практической значимости своих решений в контексте современного ИИ.
PL-1 Способен применять язык программирования Python для решения задач в области ИИ	PL-1.3 Разрабатывает и поддерживает системы обработки больших данных различной степени сложности	Знать: основные функции фреймворка PySpark, возможности Airflow для оркестрации. Уметь: применять PySpark для обработки данных, использовать Airflow для построения процессов обработки больших данных. Владеть: навыками разработки и поддержки систем обработки больших данных различной степени сложности.

Структура дисциплины

№	Наименование разделов (тем)	Количество часов			
		Всего	Аудиторная работа		Внеаудиторная работа
			Л	ЛР	СРС
1	2	3	4	6	7
1.	Введение в Big Data	4	2	2	

№	Наименование разделов (тем)	Количество часов			
		Всего	Аудиторная работа		Внеаудиторная работа
			Л	ЛР	СРС
1	2	3	4	6	7
2.	Архитектуры обработки больших данных	4	2	2	
3.	Hadoop и экосистема	4	2	2	
4.	Apache Spark	4	2	2	
5.	Обработка потоковых данных	6	4	2	
6.	NoSQL для больших данных	4	2	2	
7.	Облачные технологии для Big Data	6	4	2	
8.	Методы анализа больших данных	6	2	4	
9.	Машинное обучение на больших данных	6	2	4	
10.	Визуализация больших данных	4	2	2	
11.	Оптимизация и масштабирование	4	2	2	
12.	Безопасность данных	6	2	2	2
13.	Реальные кейсы использования Big Data	6	2	2	2
14.	Этика и законодательство в Big Data	4	2	2	
ИТОГО по разделам дисциплины		68	32	32	4
Контроль самостоятельной работы (КСР)		35,7			
Промежуточная аттестация (ИКР)		0,3			
Общая трудоемкость по дисциплине		108			

Курсовые работы: не предусмотрены

Форма проведения аттестации по дисциплине: экзамен.

Автор: Калайдин Е.Н. – доктор физ.-мат. наук, профессор, профессор кафедры прикладной математики КубГУ