

АННОТАЦИЯ рабочей программы дисциплины «Технологии обработки больших данных»

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

Цели дисциплины:

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

Задачи дисциплины:

в соответствии с поставленной целью состоят в следующем:

- 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.

Место дисциплины в структуре ООП ВО

Дисциплина «Технологии обработки больших данных» относится к Блоку 1 (Дисциплины), обязательная часть учебного плана по направлению подготовки 02.03.01 "Математика и компьютерные науки", направленность "Искусственный интеллект и аналитика данных". Освоение дисциплины базируется на знаниях, полученных при изучении дисциплин «Базы данных», «Технологии управления данными в NoSQL», «Машинное обучение», «Обработка данных на Python», «Аналитика данных».

Требования к уровню освоения дисциплины

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

Код и содержание компетенции	Общий индикатор	Индикатор уровня (Б – базовый, П – продвинутый, Э – экспертный)
BD-3 Способен организовывать хранения данных выбирая адекватные технологические решения	Разрабатывает, отлаживает и тестирует прикладные решения с элементами ИИ с применением различных технологий хранения неструктурированных данных, оценивает качество	BD-3.2 П. Пишет аналитические запросы к данным и анализирует план запроса. Умеет создавать представления, хранимые процедуры, функции и триггеры.
BD-4 Способен применять различные модели и (или) технологии обработки данных	Осуществляет выбор технологий обработки больших данных, приемлемых для создания прикладной системы ИИ с заданными требованиями	BD-4.1 П. Способен организовывать распределенное хранилище и параллельную обработку на базе современных технологий (Hadoop, Spark) больших данных
BD-5 Способен применять технологии организации инфраструктуры БД	Осуществляет выбор направления вспомогательных технологических решений для формирования единого стека работы с большими данными для решения поставленной задачи	BD-5.1 П. Выполняет отдельные функции в проектах по созданию инфраструктуры БД
ML-1 Способен применять знания об истории развития и трендах современного ИИ для формулирования корректных постановок задач и поиска	Определяет тенденции развития, оценивает новизну и практическую значимость своих решений с точки зрения современного искусственного интеллекта	ML-1.2 П. Объясняет причины появления концепции больших данных (БД), разницу определений. Выявляет различные категории проблем больших данных с примерами. Анализирует динамику

перспективных способов решения проблем с помощью ИИ		появления новых технологий, сопоставляет собственные решения с современными исследованиями и индустриальными стандартами
PL-1 Способен применять язык программирования Python для решения задач в области ИИ	Разрабатывает и поддерживает системы обработки больших данных различной степени сложности	PL-1.3 П. Умеет использовать инструменты для распределённых вычислений (Dusk, Ray) с обоснованием выбора конкретных технологий для различных ситуаций

Распределение трудоёмкости дисциплины по видам работ

Вид учебной работы	Всего часов	Семестры (часы)
		6
Контактная работа, в том числе:	68,3	68,3
Аудиторные занятия (всего):	64	64
Занятия лекционного типа	32	32
Лабораторные занятия	32	32
Иная контактная работа:	4,3	4,3
Контроль самостоятельной работы (КСР)	4	4
Промежуточная аттестация (ИКР)	0,3	0,3
Самостоятельная работа, в том числе:	4	4
Подготовка к текущему контролю	4	4
Контроль:	35,7	35,7
Подготовка к экзамену	35,7	35,7
Общая трудоёмкость	час.	108
	в том числе контактная работа	68,3
	зач. ед	3

Структура дисциплины:

Распределение видов учебной работы и их трудоёмкости по разделам дисциплины.

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