

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Петренко Вячеслав Иванович
Должность: и.о. директора Института цифрового развития
Дата подписания: 06.06.2024 09:42:20
Уникальный программный ключ:
b47f96eea65a16e179194dbf7f00b10000e7ad4d

Наименование дисциплины	Инжиниринг данных
Содержание	Введение в дисциплину. Данные. Подходы и определения. Метаданные Определения данных. Философский подход. Юридический подход. Жизненный цикл данных. Создание данных (Data Generation/Data Capture) . Обслуживание данных (Data Maintenance). Синтез данных (Data Synthesis). Использование данных (Data Usage). Публикация данных (Data Publication). Архивация данных (Data Archival). Уничтожение данных (Data Purging). Понятие метаданных. Жизненный цикл метаданных. Оценка требований и анализ контента. Спецификация системных требований. Система метаданных. Сервис и оценка. Большие данные. Системы управления большими данными Распределенные файловые системы. Распределенные фреймворки. Бенчмаркинг. Серверное программирование. Планирование. Системы развертывания. Интеграция данных. Информационная безопасность. Машинное обучение. Базы данных NoSQL и новые SQL базы данных. Архитектура системы обработки Больших данных. Хранилища данных Прием данных (Data Ingestion). Сбор данных (Data Staging). Анализ данных (Analysis Layer). Представление результатов (Consumption Layer). Параллельные алгоритмы для работы с данными. Операторы Map и Reduce. Оператор Reduce (свертка). Оператор Map. Лямбда-архитектура. Программные платформы и системы для Больших данных Системы управления потоками данных. Системы хранения Больших данных. Платформы Больших данных. Обработка данных в реальном времени. Системы управления Большими данными. Аналитические платформы. GreenPlum. Оборудование для обработки Больших данных. Центры обработки Больших данных
Результаты освоения дисциплины	Формулирует задачу для решения методами дата инженера, проводит ее декомпозицию, выделяет проблемы. Разрабатывает план действия извлечения, обработки и загрузки данных исходя из программно-аппаратных ресурсов хранилища. Формирует запрос в соответствии со сроками предоставления данных для аналитики. Знает принципы работы программных решений и специальных инструментов для представления разнородной информации в едином виде, принцип работы аналитических баз данных. Владеет современными транзакционными и аналитическими СУБД. Использует SQL запросы при решении задач дата инжиниринга, очищает данные: убирает дубли и устраняет ошибки. Структурирует данные так, чтобы они стали пригодны для аналитики. Обновляет используемые инструменты и ищет способы улучшения и автоматизации существующих процессов. Знает как собрать, обработать и хранить данные из различных источников. Использует методы для сбора, обработки и загрузки данных из различных источников. Использует программные решения и специальные инструменты для

[illegible]

	IPR BOOKS. - ISBN 978-5-9795-1743-8, экземпляров неограничено
--	--