

Разработване на системи с интензивни данни - Резюме

Data-Intensive Systems: С-ми, които обработват и анализират големи обеми данни, използвайки различни методологии и технологии

Основни научни области - AI, ML и Data Science: Изкуствен интелект, машинно обучение и наука за данните са основни компоненти в разработването на системи с интензивни данни.

Процеси: Включват различни етапи като събиране, съхранение, почистване и трансформация на данни.

Методологии- ориентирани към процеси; ориент. към управлен.

Process-oriented: Фокус върху автоматиз. и оптимиз. на процесите.

Management-oriented: Фокус върху управл. и координ. на екипите.

DataOps: Методология за управление на данни, която включва автоматизация и оптимизация на процесите за обработка на данни.

DataOps Manifesto: Осн. принципи и ценности, вкл. важността на взаимод. м/у хората, работещите анализи и колаборация с клиентите

MLOps: Методология за управление на машинното обучение, която включва непрекъснатата интеграция и доставка на модели.

AIOps: Използване на изк. интелект за оптимизация на IT операции

Жизнен цикъл на Data Science

CRISP-DM: Стандартен процес за разработка на проекти в областта на науката за данните.

TDSP (Team Data Science Process): Процес, който включва роли като Project Manager, Data Engineer, Data Scientist, Application Developer и Project Lead.

DevOps: Практики и технологии за автоматизация, които подобряват скоростта на доставяне и итерация на приложения.

Continuous Integration (CI): Автоматично изграждане, тестване и анализ на софтуерни промени.

Continuous Deployment (CD): Автоматично тестване и внедряване на софтуерни промени в продукционни среди.

Continuous Delivery (CDE): Осигуряване на готовност за доставка на софтуерни промени чрез тестване в среди, подобни на продукционните.

Model Training: Процесът на обучение на модели за машинно обуч.

Model Inference: Процесът на използване на обучените модели за предсказания и анализи.

DataOps Manifesto

Ценности: Включват важността на взаимодействията между хората, работещите анализи, колаборацията с клиентите и експериментацията.

Принципи: Включват удовлетворяване на клиентите, работа с аналитични данни, приемане на промени, ежедневни взаимодействия и самостоятелна организация на екипите.

Компоненти на DataOps

Оркестрация: Управление на данни, инструменти, код, среди и работа на аналитичните екипи.

Възпроизводимост: Версионизиране на данни, конфигурации и код за осигуряване на възпроизводими резултати.

Качество: Автоматично откриване на аномалии и проблеми със сигурността в кода, конфигурациите и данните.