

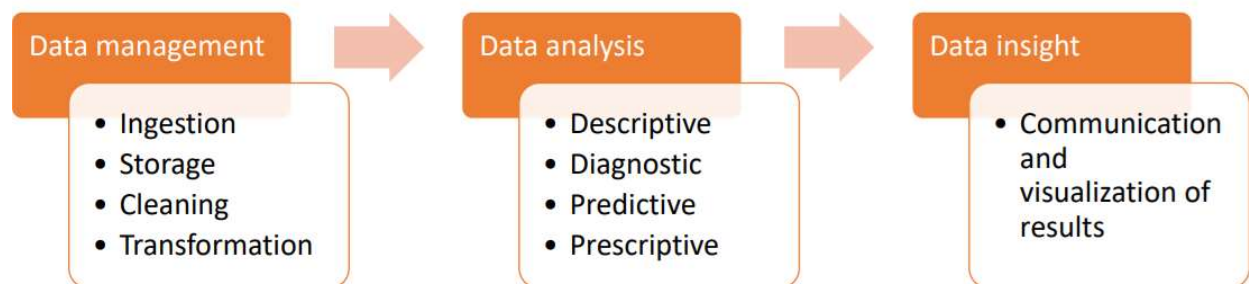
8.Разработване на системи с интензивни данни – Резюме

Developing Data-Intensive Systems: Methodologies

Data Science се занимава с отговаряне на въпроси чрез данни

- **Data science** включва:
 - **Формиране на количествени въпроси**
 - **Идентифициране на данните**, които могат да бъдат използвани за отговор на тези въпроси
 - **Почистване на данните**
 - **Форматиране и обработка**
 - **Анализиране на данните**
 - **комуникиране на резултатите** към останалите.

Big Data Value Chain



DevOps е набор от практики и автоматизационни технологии, които подобряват скоростта на доставяне и итерация на приложения. **DevOps** позволява на **Facebook** да пуска десетки до стотици актуализации на всеки няколко часа.

DevOps Practices:

- **Continuous Integration (CI)**- разработчиците **автоматично** изграждат, тестват и анализират промените в софтуера при всяка нова версия в **source repository**.
- **Continuous Deployment (CD)** - инкременталните промени в софтуера се автоматично тестват, преглеждат и разгръщат в production среди.
- **Continuous Delivery (CDE)**-гарантира, че софтуерната промяна е готова за доставка и използване от клиентите, като се тества в production-подобни среди.

AIOps (Artificial Intelligence for Operations) означава **изкуствен интелект за ИТ операции**. Използва данни и машинно обучение, за да подобри **IT Operations**, като предоставя **непрекъсната аналитична информация**.

DataOps Manifesto - Values

- **Individuals and interactions over processes and tools**
- **Working analytics over comprehensive documentation**
- **Customer collaboration over contract negotiation**
- **Experimentation, iteration, and feedback over extensive upfront design**
- **Cross-functional ownership of operations over siloed responsibilities**

DataOps Manifesto - Principles

- **Continually satisfy your customer** – Най-важният приоритет е **непрекъснатата доставка на аналитични инсайти** в интервал от **няколко минути до седмици**.

- **Value working analytics** – Основната мярка за производителност е степента, в която аналитичните инсайти са точни, надеждни и използвани.
- **Embrace change** – Адаптиране към променящите се клиентски нужди за конкурентно предимство.
- **It's a team sport** – Разнообразието от роли, умения и инструменти повишава иновациите и продуктивността.
- **Daily interactions** – Клиенти, аналитични екипи и operations трябва да работят заедно всеки ден по време на проекта.
- **Self-organize** – Най-добрите алгоритми, архитектури и анализи произлизат от самоорганизиращи се екипи.
- **Reduce heroism** – Целта е устойчиви и мащабируеми аналитични процеси, а не разчитане на "герои" в екипа.
- **Reflect** – Екипите трябва редовно да оценяват представянето си чрез обратна връзка от клиенти и анализи.
- **Analytics is code** – Всички аналитични инструменти генерират код и конфигурации, описващи как данните се обработват.
- **Orchestrate** – Координацията на данни, инструменти, код и среди е ключов фактор за успеха на аналитичния процес.
- **Quality is paramount** – Автоматично откриване на грешки и сигурност в кода, конфигурацията и данните.
- **Monitor quality and performance** – Постоянно наблюдение на производителността, сигурността и качеството за откриване на проблеми.
- **Reuse** – Избягване на дублиране на работа, за да се повиши ефективността на аналитичните екипи.

- **Improve cycle times – Оптимизиране на процеса** от клиентска заявка до пускане в производство и рефакторинг.

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.