



Академия АйТи  
a Softline Company



## **Data Quality: управление и наблюдаемость данных**

Код курса: DQ

# Data Quality: управление и наблюдаемость данных

Код курса: DQ

|                          |                          |
|--------------------------|--------------------------|
| <b>Длительность</b>      | 40 ак. часов             |
| <b>Формат</b>            |                          |
| <b>Разработчик курса</b> | Академия АйТи            |
| <b>Тип</b>               | Учебный курс             |
| <b>Способ обучения</b>   | Под руководством тренера |

## О курсе

Курс дает комплексный подход к повышению надежности корпоративных данных. Участники курса научатся выявлять проблемы качества данных с помощью профилирования и метрик, освоят методы автоматизированной очистки и стандартизации, внедрят мониторинг и построят процессы своевременного реагирования на инциденты. В программу интегрированы практические инструменты: Great Expectations, SodaCL и OpenMetadata, что позволит быстро перейти от теории к реальному применению полученных знаний. Итоговый проект курса — готовый к внедрению пайплайн управления качеством данных, включающий расчет ROI и разработку политик Data Governance. После обучения участники смогут самостоятельно измерять, контролировать и улучшать качество данных в своих организациях, предотвращать бизнес-риски, связанные с плохими данными, и эффективно обосновывать руководству необходимость инвестиций в качество данных.

## Подробная информация

### Профиль аудитории:

Data-инженеры, BI-аналитики, data-scientists уровня junior-middle, специалисты по ETL/ELT, а также сотрудники, отвечающие за внедрение и развитие процессов управления качеством данных и Data Governance в своих компаниях.

### Предварительные требования:

- Владение SQL на уровне запросов средней сложности.
- Базовое знание Python (для анализа и обработки данных).
- Опыт работы с корпоративными данными (базы данных, хранилища данных, BI-системы).

### По окончании курса слушатели смогут:

- Самостоятельно проводить профилирование данных, выявлять и измерять проблемы их качества с использованием метрик и инструментов (Great Expectations, SodaCL).
- Выполнять очистку и стандартизацию данных, а также автоматизировать этот процесс.
- Настроить постоянный мониторинг качества данных, использовать дашборды и своевременно

реагировать на инциденты.

- Формулировать политики и процессы Data Governance, рассчитывать и обосновывать ROI от мероприятий по повышению качества данных.
- Разработать и внедрить полноценный рабочий DQ-пайплайн, который позволяет превентивно выявлять и устранять проблемы с качеством данных.

## Программа курса

### Введение и Основы

- Ценность данных как актива.
- Качество Данных (DQ) - определения, ключевые понятия.
- Бизнес-последствия плохого качества данных
- Обзор процесса управления DQ

### Модуль 1: Измерения (Dimensions) Качества Данных

- Глубокий разбор ключевых измерений:
  - Точность (Accuracy): Соответствие реальности.
  - Полнота (Completeness): Отсутствие пропусков.
  - Своевременность (Timeliness / Currency): Актуальность на момент использования.
  - Непротиворечивость (Consistency): Согласованность между системами и наборами.
  - Уникальность (Uniqueness): Отсутствие дубликатов сущностей.
  - Валидность (Validity): Соответствие формату и бизнес-правилам.
  - Целостность (Integrity): Соблюдение ссылочной целостности (Foreign Keys).
  - Доступность (Accessibility) и Понятность (Interpretability).
- Относительность измерений: Как приоритеты DQ зависят от контекста использования данных.
- *Практика*: Разбор примеров проблем DQ

### Модуль 2: Профилирование и Оценка Данных

- Профилирование данных (Data Profiling): Задачи, методы, типы анализа
- Выявление аномалий
- Инструменты профилирования
- Метрики и KPI для DQ
- *Практика*: Профилирование предоставленного набора данных. Выявление основных проблем.

### Модуль 3: Очистка (Cleansing) и Стандартизация Данных

- Техники очистки
- Стандартизация данных
- Инструменты очистки
- Разработка правил очистки
- *Практика*: Применение техник очистки

### Модуль 4: Мониторинг и Управление Процессами DQ

- Постоянный мониторинг DQ
- Автоматизация проверок DQ

- Создание и управление правилами DQ
- Рабочие процессы (Workflow) для обработки инцидентов DQ
- *Практика:* Разработка плана мониторинга для критичного набора данных. Создание дашборда с ключевыми метриками DQ.

#### Модуль 5: Организационные Аспекты

- Ключевые роли
- Data Governance и DQ
- Внедрение культуры качества данных
- Разработка политики и стратегии DQ. Дорожная карта улучшения DQ.
- *Практика:* Разработка фрагмента политики DQ.

#### Модуль 6: Продвинутое Темы и Тренды

- Data Observability
- Машинное обучение для DQ
- Streaming DQ
- Регуляторные требования
- *Практика:* Обсуждение влияния регуляторики.

#### Модуль 7: Инструменты и Технологии (Обзор)

- Категории инструментов DQ
- Критерии выбора инструмента

#### Модуль 8: Планирование

- Разработка плана действий
- Измерение ROI от улучшений DQ
- Будущее управления качеством данных
- Резюме ключевых выводов курса.
- *Практика:* Разработка концепции решения DQ

[Посмотреть расписание курса и записаться на обучение](#)

**Обращайтесь по любым вопросам**  
к менеджерам Академии АйТи

**+7 (495) 150 96 00** | academy@academyit.ru