

ДЕПАРТАМЕНТ ОБРАЗОВАНИЯ И НАУКИ ГОРОДА МОСКВЫ

Государственное автономное образовательное учреждение

высшего образования города Москвы

«Московский городской педагогический университет»

(ГАОУ ВО МГПУ)

Институт цифрового образования

Департамент информатики, управления и технологий

Практическая(лабораторная) работа № 3.2

по дисциплине «Платформы Data Engineering»

Выполнил:

студент группы БД-251м

Направление подготовки/Специальность

38.04.05 - Бизнес-информатика

Войт Иван Иванович

(Ф.И.О.)

Проверил:

Доцент департамента информатики, управления и технологий, доктор

экономических наук

(ученая степень, звание)

Босенко Т.М.

(Ф.И.О.)

Москва 2025

Оглавление	
ВВЕДЕНИЕ.....	3
ПРОЦЕСС РАЗРАБОТКИ	3
АНАЛИЗ РЕЗУЛЬТАТОВ.....	4
ВЫВОДЫ.....	4
ЗАКЛЮЧЕНИЕ	6
ССЫЛКИ	6

ВВЕДЕНИЕ

Цель работы: Провести комплексное исследование современных практик автоматизации в области Data Engineering, выявить ключевые тенденции, проблемы и эффективные подходы к построению data-pipeline.

Выбранная тема: "Анализ уровня автоматизации процессов в Data Engineering: инструменты, метрики эффективности и барьеры внедрения"

Целевая аудитория опроса:

- Data Engineers и инженеры данных
- Data Scientists и ML-инженеры
- Analytics Engineers и дата-аналитики
- DevOps/Platform инженеры
- Руководители data-команд
- Специалисты по работе с данными с опытом работы от 1 года

Актуальность темы: В условиях роста объемов данных и сложности аналитических задач, автоматизация процессов становится критическим фактором успеха data-проектов.

ПРОЦЕСС РАЗРАБОТКИ

Этап 1: Проектирование опроса

- Проведен анализ существующих исследований в области Data Engineering
- Определены 5 ключевых блоков для комплексного охвата темы
- Разработана смешанная стратегия вопросов (единичный/множественный выбор, шкалы оценки)
- Этап 2: Импорт данных
- Импорт данных из заполненных анкет
- Этап 3: Загрузка данных в DataLense
- Создание датасете
- Создание вычисляемых полей в датасете
- Создание чартов

- Сбор всех чартов на едином дашборде

АНАЛИЗ РЕЗУЛЬТАТОВ

Визуализация 1: Распределение ролей респондентов

- *Диаграмма:* Круговая диаграмма
- *Интерпретация:* Преобладание Data Engineers (45%), что обеспечивает релевантность данных для исследования автоматизации

Визуализация 2: Уровень автоматизации по процессам

- *Диаграмма:* Тепловая карта с градиентом цветов
- *Интерпретация:* Наиболее автоматизированы процессы сбора данных (4.2/5), наименее - мониторинг качества данных (2.8/5)

Визуализация 3: Используемые инструменты оркестрации

- *Диаграмма:* Столбчатая диаграмма
- *Интерпретация:* Apache Airflow доминирует (65%), cloud-native решения набирают популярность (20%)

Визуализация 4: Влияние автоматизации на рабочие процессы

- *Диаграмма:* Радарная диаграмма
- *Интерпретация:* Наибольшее влияние отмечено на масштабируемость (4.5/5) и надежность (4.3/5)

Визуализация 5: Основные трудности внедрения

- *Диаграмма:* Горизонтальная столбчатая диаграмма
- *Интерпретация:* Лидируют "Нехватка экспертизы" (58%) и "Технический долг" (42%)

ВЫВОДЫ

Вывод 1: Доминирование Apache Airflow как стандарта индустрии 65% респондентов используют Airflow в качестве основного инструмента оркестрации, что свидетельствует о формировании отраслевого стандарта.

Вывод 2: Критический разрыв в автоматизации качества данных
При высокой автоматизации ETL-процессов, мониторинг качества данных остается преимущественно ручным (2.8/5), создавая риски для reliability.

Вывод 3: Недостаточное измерение эффективности автоматизации
Только 25% компаний формально измеряют ROI от автоматизации, что указывает на недостаток data-driven подхода к управлению data-инфраструктурой.

Вывод 4: Кадровый дефицит как основной барьер
58% респондентов отмечают нехватку экспертизы как главное препятствие, что требует инвестиций в обучение и развитие команд.

Вывод 5: Высокая отдача от инвестиций в автоматизацию
Несмотря на барьеры, 85% респондентов отмечают значительное улучшение масштабируемости и надежности процессов после внедрения автоматизации.

ЗАКЛЮЧЕНИЕ

Проделанная работа позволила систематизировать современное состояние автоматизации в Data Engineering и выявить ключевые тенденции:

Достигнутые результаты:

- Разработан комплексный инструмент для оценки зрелости автоматизации
- Выявлены приоритетные направления для улучшения процессов
- Создан шаблон для регулярного мониторинга прогресса в командах

Полученный опыт:

- Получены навыки проектирования исследовательских опросов
- Разработана методика анализа качественных и количественных данных

ССЫЛКИ

Git-репозиторий с полными материалами работы:
<https://github.com/youngvoyt/DEP-MGPU/tree/713c4245cc7218b45134fc7f37618acb61f32d1d/Module03/Lab%203.2>

Дашборд в DataLense: <https://datalens.ru/t4jr7478kzyse>