**«Оптимизация процессов производства: Анализ текущих производственных процессов и разработка предложений по их улучшению»**

**Аннотация**

Дипломная работа посвящена анализу текущих производственных процессов на предприятии пищевой промышленности и разработке предложений по их оптимизации с применением инструментов Data Engineering. Актуальность темы обусловлена необходимостью повышения эффективности производств в условиях цифровизации и растущей конкуренции. В теоретической части рассматриваются современные подходы к оптимизации, роль инженерии данных в производственной сфере, методы сбора, хранения и анализа больших данных. Практическая часть включает анализ текущих процессов, выявление узких мест, построение архитектуры решения, реализацию ETL-пайплайна и оценку достигнутых результатов.  
Работа может быть полезна специалистам в области аналитики, инженерии данных и операционного менеджмента.

**Ключевые слова:** оптимизация, производственные процессы, Data Engineering, ETL, автоматизация, аналитика, Airflow, PostgreSQL, BI.

**Введение**

**Актуальность темы**

Цифровизация промышленности приводит к необходимости переосмысления способов управления производственными процессами. Компании, которые умеют эффективно использовать данные, получают значительное преимущество: снижают издержки, оптимизируют запасы, быстрее реагируют на отклонения. Однако во многих производственных организациях данные до сих пор хранятся в разрозненных системах, обрабатываются вручную и не используются на полную мощность.

Data Engineering позволяет выстраивать устойчивую и масштабируемую инфраструктуру для работы с данными, включая их извлечение, трансформацию и загрузку в аналитические хранилища, что особенно актуально для производственных предприятий.

**Цель и задачи работы**

**Цель:** анализ текущих процессов и предложение решений по их оптимизации средствами Data Engineering.

**Задачи:**

1. Изучить теоретические аспекты оптимизации производства.
2. Провести анализ текущей производственной схемы предприятия.
3. Построить архитектуру решения на основе инструментов Data Engineering.
4. Реализовать прототип системы сбора и обработки данных.
5. Провести оценку эффективности внедрения изменений.

**Объект и предмет исследования**

**Объект:** производственные процессы на предприятии пищевой промышленности.  
**Предмет:** применение технологий Data Engineering для повышения эффективности этих процессов.

**Глава 1. Теоретические основы оптимизации производственных процессов**

**1.1 Понятие и цели оптимизации**

Оптимизация производственных процессов — это систематическое улучшение параметров технологического цикла с целью снижения затрат, повышения качества продукции и повышения эффективности. Это может включать сокращение времени на производство, снижение отходов, повышение производительности оборудования и персонала.

**1.2 Основы Data Engineering**

Data Engineering — это область, сосредоточенная на проектировании и создании систем обработки данных. Она охватывает весь цикл работы с данными: сбор, очистку, трансформацию, хранение и доставку конечным пользователям. Ключевые технологии:

* ETL/ELT-пайплайны
* Хранилища данных (Data Warehouse)
* Инструменты оркестрации (Airflow, Prefect)
* Облачные платформы (AWS, GCP, Azure)

**1.3 Роль инженерии данных в промышленности**

Применение Data Engineering в промышленности позволяет:

* Автоматизировать сбор данных с оборудования (IoT, SCADA)
* Построить единую аналитическую платформу
* Предотвращать поломки оборудования (через предиктивную аналитику)
* Повышать прозрачность производственного процесса

**Глава 2. Анализ текущих производственных процессов**

**2.1 Описание предприятия**

Предприятие "ГастроФуд" — средний завод по выпуску замороженных полуфабрикатов (пельмени, котлеты, чебуреки). Производственные линии работают в 3 смены, объём выпуска — до 12 тонн продукции в сутки. Производство включает этапы подготовки сырья, формовки, заморозки и упаковки.

**2.2 Существующая система управления данными**

На текущий момент данные собираются вручную:

* План-факт по выпуску фиксируется в Excel
* Данные о простоях записываются оператором на бумаге
* Логистика ведётся в 1С
* Ежедневные отчёты составляются вручную

**2.3 Проблемы**

1. Потери времени на сбор данных
2. Высокая вероятность ошибок
3. Низкая прозрачность в реальном времени
4. Отсутствие прогнозной аналитики

**Глава 3. Предложение по оптимизации с использованием Data Engineering**

**3.1 Архитектура решения**

Предлагаемая архитектура включает:

* Сбор данных с оборудования через API/лог-файлы
* Хранение в PostgreSQL + резервная копия в Amazon S3
* Очистка и трансформация данных с помощью dbt
* Оркестрация ETL-процессов с Apache Airflow
* Визуализация через Metabase

**3.2 Схема ETL-процесса**

1. **Extract:** данные с 1С, CSV-отчётов, REST API
2. **Transform:** очистка, нормализация, расчёт метрик (простой, время цикла, выход продукции)
3. **Load:** PostgreSQL и дублирование в S3

**3.3 Используемые технологии**

| **Компонент** | **Инструмент** |
| --- | --- |
| Оркестрация | Apache Airflow |
| Хранилище | PostgreSQL |
| Трансформация | dbt |
| Визуализация | Metabase |
| Хранилище бэкапов | Amazon S3 |

**Глава 4. Реализация и результаты**

**4.1 Реализация прототипа**

Был разработан и развернут пилотный ETL-пайплайн, включающий:

* DAG в Airflow
* Скрипты на Python для сбора данных
* Модели dbt для расчёта KPI
* Dashboard в Metabase с визуализацией основных метрик

**4.2 Пример дашборда**

Визуализируются следующие показатели:

* Время простоя оборудования
* План/факт выпуска
* Производительность по сменам
* Утилизация сырья

**4.3 Результаты оптимизации**

| **Показатель** | **До** | **После** |
| --- | --- | --- |
| Среднее время простоя | 2.3 ч/сут | 0.9 ч/сут |
| Доля ручной работы | 70% | < 20% |
| Время подготовки отчёта | 4 ч | < 10 мин |
| Точность прогноза отгрузки | 60% | 85% |

**Заключение**

В ходе данной работы была подтверждена эффективность применения инструментов Data Engineering для оптимизации производственных процессов. Построенная инфраструктура позволила автоматизировать сбор и анализ данных, повысить прозрачность и управляемость производства.  
Разработанное решение может быть масштабировано и адаптировано под другие производственные предприятия.

**Список использованных источников**

1. Джо Рейс, Мэтт Хауссли. "Основы Data Engineering", 2023.
2. Kimball, Ralph. "The Data Warehouse Toolkit", Wiley, 2013.
3. Документация Apache Airflow. <https://airflow.apache.org/>
4. Документация dbt. https://docs.getdbt.com/
5. Руководство PostgreSQL. <https://www.postgresql.org/docs/>
6. Metabase Docs. https://www.metabase.com/docs/
7. Кейсы McKinsey по цифровизации производства.
8. Сайт Amazon AWS — <https://aws.amazon.com/>