**Разработка Data Warehouse для компании по продаже электроники**

**1. Бизнес-требования и цели создания Data Warehouse**

**Основные цели проекта:**

* Централизация данных из различных источников
* Обеспечение быстрого доступа к аналитическим данным
* Создание единой системы отчетности
* Улучшение качества данных
* Поддержка принятия управленческих решений

**Бизнес-требования:**

* Хранение данных о продажах за последние 3 года
* Возможность детализации данных до уровня отдельной транзакции
* Поддержка многомерного анализа данных
* Обеспечение целостности и непротиворечивости данных
* Регулярное обновление данных в режиме реального времени

**2. Системы-источники данных**

**CRM-система**  
Формат: JSON  
Данные: информация о клиентах, история покупок, контактные данные  
Пример записи:

{

"customer\_id": 123,

"name": "Иванов И.И.",

"phone": "+7(999)123-45-67",

"email": "ivanov@mail.ru"

}

**Система управления складом**  
Формат: XML  
Данные: остатки товаров, перемещения, поставки  
Пример записи:

<stock>

<product\_id>456</product\_id>

<warehouse\_id>1</warehouse\_id>

<quantity>100</quantity>

<date>2025-06-03</date>

</stock>

**POS-терминалы**  
Формат: CSV  
Данные: информация о продажах  
Пример записи:

transaction\_id,product\_id,quantity,price,date

1001,456,2,15000,2025-06-03

**3. Модель данных (DDS слой)**

**Основные сущности:**

1. **Измерение времени (Dim\_Date)**

* date\_key
* date\_full
* day\_of\_week
* month
* year

1. **Измерение продукта (Dim\_Product)**

* product\_key
* product\_id
* product\_name
* category\_id
* manufacturer
* price

1. **Измерение клиента (Dim\_Customer)**

* customer\_key
* customer\_id
* first\_name
* last\_name
* phone
* email
* registration\_date

1. **Измерение склада (Dim\_Warehouse)**

* warehouse\_key
* warehouse\_id
* warehouse\_name
* address
* capacity

1. **Факт продаж (Fact\_Sales)**

* sales\_key
* date\_key
* product\_key
* customer\_key
* warehouse\_key
* quantity
* total\_amount

**Связи между сущностями:**

* Fact\_Sales → Dim\_Date (1:M)
* Fact\_Sales → Dim\_Product (1:M)
* Fact\_Sales → Dim\_Customer (1:M)
* Fact\_Sales → Dim\_Warehouse (1:M)

**4. Разложение данных по слоям**

**Stage слой**

* Хранит сырые данные из источников
* Сохраняет исходные форматы данных
* Не производится очистка и трансформация

**ODS слой**

* Хранит очищенные данные из источников
* Производится базовая валидация
* Удаляются дубликаты
* Сохраняется история изменений

**DDS слой**

* Хранит структурированные данные
* Создаются измерения и факты
* Применяются бизнес-правила
* Обеспечивается целостность данных

**Datamart слой**

* Хранит готовые витрины данных
* Оптимизирован для аналитики
* Содержит агрегированные данные
* Предназначен для конечных пользователей

**Структура отчета**

1. Введение

* Описание компании
* Цели и задачи проекта

1. Анализ источников данных

* Описание каждой системы-источника
* Примеры форматов данных

1. Проектирование модели данных

* Описание всех сущностей
* Схема связей между таблицами
* Примеры DDL-скриптов

1. Реализация ETL-процессов

* Описание преобразования данных
* Схема движения данных между слоями

1. Заключение

* Результаты работы
* Рекомендации по улучшению
* Возможные риски

**Рекомендации по реализации**

1. Использовать **Snowflake** схему для измерений
2. Применить **Slowly Changing Dimensions** для отслеживания изменений
3. Реализовать **инкрементальную загрузку** данных
4. Внедрить систему **мониторинга качества данных**
5. Использовать **хранилище на основе СУБД** (например, PostgreSQL)

При разработке необходимо уделить особое внимание качеству данных, производительности запросов и обеспечению безопасности информации.