

## Д32. Построение архитектуры хранилища данных

### 1 Бизнес-требования и цели создания Data warehouse

#### 1.1. Цель создания data warehouse

Создать единое аналитическое хранилище данных, объединяющее информацию о продажах, клиентах, продуктах и складских остатках из всех каналов бизнеса, для повышения прозрачности, ускорения принятия решений и оптимизации ключевых бизнес-процессов.

#### 1.2. Бизнес-цели

№	Бизнес - цель	Ожидаемый результат
1	Повысить точность и скорость отчетности для руководства	Сократить время подготовки отчетов различной степени с 3 дней до <3 часов
2	Оптимизировать управление запасами и снизить задержки	Снизить избыточные остатки на 10%, сократить дефицит на 15%
3	Улучшить маржинальность ассортимента	Исключить 20% товаров с маржой <5% на основе анализа за 12 мес.
4	Обеспечить единую версию по KPI	Устранить расхождения в данных между отделами (достичь >98% согласованности)
5	Масштабирование бизнеса (новые регионы)	Обеспечить подключение нового источника данных за <5 рабочих дней

#### 1.3. Бизнес – требования

Бизнес – требования сгруппированы по категориям:

##### 1.3.1. Требования к данным

ID	Категория	Требование
BR-DATA-01	Продажи	DWH должен содержать все заказы (онлайн, офлайн, маркетплейсы) с детализацией: дата, время, магазин, товар, цена, скидка, клиент, статус, способ оплаты
BR-DATA-02	Клиенты	DWH должен содержать унифицированный профиль клиента: ФИО, email, телефон, история покупок
BR-DATA-03	Продукты	DWH должен содержать справочник товаров: наименование, категория, бренд, цена закупки, цена продажи, маржа, вес, габариты
BR-DATA-04	Склады	DWH должен содержать данные по остаткам: склад, товар, количество, резерв, дата последнего поступления, срок годности (если есть)
BR-DATA-05	История	Данные должны храниться не менее 5 лет с возможностью восстановления состояния на любую дату

##### 1.3.2. Требования к качеству данных

ID	Категория	Требование
BR-DQ-01	Актуальность	Данные должны обновляться ежедневно (основной цикл); ключевые метрики (продажи, остатки) — ежечасно
BR-DQ-02	Полнота	Не допускается >2% пропущенных обязательных полей
BR-DQ-03	Согласованность	Данные по одному клиенту/товару/заказу должны быть идентичны во всех источниках после загрузки в DWH (>98% соответствия)

BR-DQ-04	Достоверность	Должны быть реализованы правила валидации: формат email, положительная цена, логические связи (например, заказ → клиент → email)
BR-DQ-05	Прослеживаемость	Должна быть возможность отследить источник и дату загрузки любого значения в DWH

### 1.3.3. Требования к функциональности и пользователям

ID	Категория	Требование
BR-FUNC-01	Отчётность	Возможность формировать стандартные отчёты: «Продажи по дням», «ТОП-100 товаров», «Динамика клиентов» — за < 15 минут
BR-FUNC-02	Аналитика	Возможность строить различные запросы
BR-FUNC-03	Экспорт	Возможность экспорта данных в Excel/CSV с сохранением структуры и фильтров
BR-FUNC-04	Дашборды	Интеграция с Power BI/Tableau: преднастроенные дашборды для каждой роли (финансы, маркетинг, операции)
BR-FUNC-05	Оповещения	Настройка автоматических оповещений: «остаток товара < порога», «падение продаж в регионе >20%»

### 1.3.4. Требования к безопасности и доступу

ID	Категория	Требование
BR-SEC-01	RBAC	Разграничение доступа по ролям: маркетолог — только клиенты и кампании; финансист — только финансы; магазин — только свой регион
BR-SEC-02	Аудит	Логирование всех действий: кто, когда, какой отчёт/запрос выполнил, какие данные экспортировал
BR-SEC-03	PD	Хранение и обработка персональных данных в соответствии с ФЗ-152
BR-SEC-04	Экспорт	Экспорт данных за пределы DWH — только после двухфакторной аутентификации и логирования
BR-SEC-05	Маскировка	Для тестовых сред — маскировка email, телефонов, ФИО клиентов

### 1.3.5. Требования к производительности и доступности

ID	Категория	Требование
BR-SLA-01	Доступность	DWH доступен 24/7, кроме планового окна обновления
BR-SLA-02	Скорость	Время выполнения типового отчёта — <15 сек; сложного ad-hoc запроса — <2 мин
BR-SLA-03	Загрузка	ETL-процесс завершается в течение 2 часов после окончания бизнес-дня

## 1.4. KPI успеха создания DWH

ID	Категория	Цель	Метод измерения
KPI-01	Время подготовки различных отчетов	< 3 часов	Тайминг в BI-системе
KPI-02	% согласованности данных между источниками	> 98%	Аудит выборки
KPI-03	Снижение ручного труда по сбору данных	70%	Опрос, замер времени
KPI-04	Снижение избыточных складских остатков	10%	Сравнение остатков до/после

## 2 Источники данных

Для построения DWH будут использоваться данные из трёх основных систем. Каждая система имеет свой формат хранения и передачи данных.

### 2.1. Интернет-магазин

**Тип:** Транзакционная OLTP-система, основной канал продаж.

**Формат передачи:** SQL-запросы к реляционной базе данных MySQL (экспорт таблиц через ETL-процесс).

**Основные данные:** данные о заказах, данные о позициях заказов.

### 2.2. CRM – система

**Тип:** Система управления взаимоотношениями с клиентами.

**Формат передачи:** REST API, возвращающий данные в формате JSON.

**Основные данные:** контактные данные клиента, потенциальные клиенты.

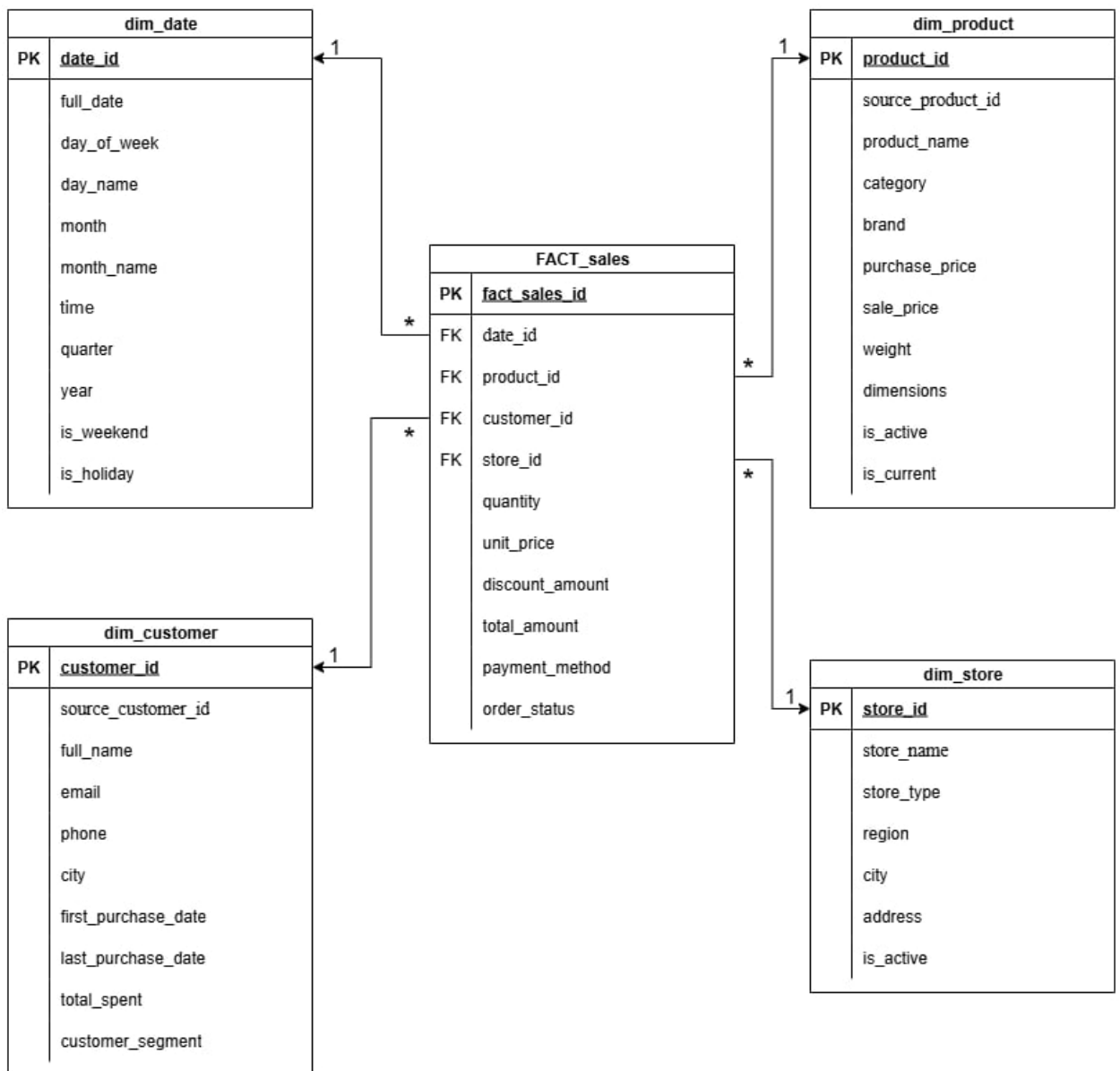
### 2.3. Система управления складом

**Тип:** Локальная ERP – система на базе 1С: Управление торговлей.

**Формат передачи:** Ежедневные CSV – выгрузки по таймеру.

**Основные данные:** данные о складе, данные о товаре.

### 3 Модель данных для Data Warehouse (DDS-слой)



### 4 Принципы разложения данных по слоям stage - ods - dds – datamart

#### 1. Stage

Назначение: Временное хранилище сырых данных из источников, без какой-либо обработки или трансформации. Это буферная зона между источниками и DWH.

Источники данных:

- Интернет-магазин (MySQL)
- CRM (REST API → JSON)
- 1С: Управление торговлей (CSV-выгрузки)

Принципы:

- Данные копируются из источников в их исходном формате и структуре.
- Данные хранятся только на время выполнения ETL-процесса. После успешной загрузки в ODS они могут быть удалены или архивированы.
- Этот слой изолирует источники от DWH. Любые сбои, изменения структуры или ошибки в источниках сначала попадают сюда, не затрагивая ODS и DDS.
- Слой позволяет отследить, какие именно данные пришли от источника.

## 2. ODS

Назначение: Создание единого, интегрированного, но детализированного представления операционных данных из всех источников.

Принципы:

- Данные из разных источников (Stage) объединяются по бизнес-ключам.
- Проводиться очистка и стандартизация, а именно: удаление дубликатов, заполнение пустых значений по умолчанию, приведение к единому формату (даты, деньги, email, номер телефона и др.).
- Сохраняется максимальная детализация, как в исходных системах. Каждая строка в ODS соответствует одной транзакции или записи в источнике.
- Данные обновляются с высокой частотой — ежечасно.
- Структура данных часто нормализована, чтобы избежать избыточности на этом этапе.

## 3. DDS

Назначение: Хранение исторических, очищенных, согласованных и структурированных данных, оптимизированных для аналитики и отчетности.

Принципы:

- Данные хранятся длительное время (минимум 5 лет) с поддержкой истории изменений.
- Применяются строгие правила бизнес-логики и валидации. Данные проходят финальную очистку и преобразование.
- Данные организованы в оптимизированные для чтения структуры.
- Данные группируются вокруг ключевых бизнес-сущностей и процессов: продажи, клиенты, продукты, финансы, склад.
- Обычно обновляется пачками в конце дня или ночью, как требует

## 4. Datamart

Назначение: Предоставление готовых, предварительно агрегированных и структурированных данных для конкретных бизнес-подразделений, ролей или типов отчетности.

- Каждая витрина фокусируется на одной бизнес-области или подразделении.

- Содержит предварительно рассчитанные агрегаты, что обеспечивает мгновенную скорость отчетов.
- Структура и данные адаптированы под нужды конкретной группы пользователей.
- Является основным источником данных для инструментов визуализации.
- На этом уровне легко реализуется разграничение доступа.