**Отчет по построению Data Warehouse для ООО "Техна"**

**1. Введение**

***1.1. Общая информация***

ООО "Техна" - динамично развивающаяся розничная сеть по продаже электроники с годовым оборотом 2+ млрд рублей. Компания объединяет:

- 15 розничных магазинов в 5 регионах РФ

- Интернет-магазин с долей 30% от общего объема продаж

- 3 распределительных склада (Москва, Екатеринбург, Новосибирск)

***1.2. Проблемы текущей аналитики***

1. Разрозненность данных:

- Финансовые показатели хранятся в SAP ERP

- Онлайн-продажи и CRM - в 1С Битрикс

- Логистические данные поступают в Excel-отчетах от 3PL-операторов

2. Отсутствие единой отчетности:

- Время формирования консолидированного отчета - 3-5 рабочих дней

- Расхождения в данных между системами достигают 15%

3. Ограниченные аналитические возможности:

- Невозможность кросс-анализа оффлайн и онлайн продаж

- Отсутствие исторических данных для трендового анализа

***1.3. Цель проекта***

Создание единого аналитического хранилища данных, которое:

- Консолидирует информацию из всех бизнес-систем

- Обеспечит "единую версию правды" для руководства

- Позволит сократить время подготовки отчетов с дней до часов

- Откроет возможности для предиктивной аналитики

**2. Бизнес-требования**

***2.1. Стратегические цели***

1. Повышение операционной эффективности:

- Снижение логистических издержек на 7-12% через оптимизацию маршрутов

- Уменьшение складских запасов на 15-20% при сохранении уровня сервиса

2. Рост прибыльности:

- Увеличение среднего чека на 8-10% за счет персонализированных предложений

- Снижение маркетинговых затрат на 20% через точную сегментацию

3. Улучшение клиентского опыта:

- Сокращение времени доставки на 30%

- Повышение NPS (индекс лояльности) с 65 до 75 пунктов

***2.2. Тактические требования***

*2.2.1. Функциональные требования*

1. Консолидация данных:

- Ежедневное обновление показателей с задержкой не более 4 часов

- Хранение истории изменений ключевых параметров (цены, остатки)

2. Аналитические возможности:

- Кросс-канальный анализ продаж (онлайн/оффлайн)

- ABC-XYZ анализ товарного ассортимента

- Расчет LTV (пожизненной ценности) клиента

3. Отчетность:

- 15+ стандартных отчетов (ежедневных, еженедельных, месячных)

- Возможность drill-down до уровня отдельного чека

*2.2.2. Технические требования*

1. Производительность:

- Загрузка данных: не более 2 часов для полного цикла ETL

- Формирование отчетов: до 30 секунд для стандартных дэшбордов

2. Надежность:

- Время восстановления после сбоя - не более 1 часа

- Резервное копирование - ежедневное с хранением 30 дней

3. Безопасность:

- Разграничение доступа по ролям (менеджеры, аналитики, руководство)

- Шифрование персональных данных клиентов

**3. Принципы построения слоев DWH**

***3.1. Подробная архитектура***

*3.1.1. Stage Layer (Сырой слой)*

Назначение: "Приемный пункт" для необработанных данных

Принципы:

1. Полнота:

- Сохранение всех полей из источника без преобразований

- Фиксация метаданных (источник, время загрузки, версия)

2. Историчность:

- Каждая загрузка создает новую версию данных

- Хранение: 7 дней для часто обновляемых данных, 1 год - для редко меняющихся

3. Таблицы слоя

|  |  |  |
| --- | --- | --- |
| **Таблица** | **Источник** | **Пример полей** |
| stage.sap\_sales | SAP ERP | order\_num, material\_id, plant\_code, posting\_date, amount\_net |
| stage.sap\_inventory | SAP ERP | warehouse\_id, material\_id, batch, qty\_available, last\_movement |
| stage.bitrix\_orders | 1С Битрикс | order\_id, date\_insert, user\_id, pay\_system, delivery\_id |
| stage.bitrix\_users | 1С Битрикс | user\_id, email, phone, registration\_date, utm\_source |
| stage.carrier\_a | Excel | waybill\_no, departure\_date, arrival\_date, transport\_cost |
| stage.carrier\_b | Excel | shipment\_id, route\_code, planned\_days, actual\_days |

*3.1.2. ODS Layer (Операционный слой)*

Назначение: "Единая версия правды" в бизнес-терминах

Принципы:

1. Стандартизация:

- Приведение к единым справочникам (товары, клиенты)

- Унификация форматов дат, валют, единиц измерения

2. Качество данных:

- Валидация по бизнес-правилам (цена > 0, дата в допустимом диапазоне)

- Очистка от тестовых транзакций

3. Таблицы слоя

| **Таблица** | **Назначение** | **Ключевые поля** |
| --- | --- | --- |
| ods.sales | Все продажи | sales\_id, doc\_date, customer\_key, product\_key, qty, net\_amount |
| ods.customers | Клиенты | customer\_key, name, phone\_hash, region\_code, loyalty\_tier |
| ods.products | Товары | product\_key, sku, category\_id, base\_uom, min\_price |
| ods.inventory | Остатки | inventory\_id, warehouse\_key, product\_key, stock\_qty, reserved\_qty |
| ods.deliveries | Доставки | delivery\_key, order\_id, carrier\_id, promised\_date, actual\_date |

*3.1.3. DDS Layer (Аналитический слой)*

Назначение: Оптимизированная модель для анализа

Принципы:

1. Денормализация:

- Предварительные агрегаты для ускорения отчетов

- Иерархии в измерениях (регион → город → магазин)

2. Историчность:

- Type 2 SCD (Slowly Changing Dimensions) для ключевых атрибутов

- Хранение фактов с привязкой к временным периодам

3. Таблицы слоя

**Измерения (Dimensions)**

| **Таблица** | **Описание** | **Пример SCD-полей** |
| --- | --- | --- |
| dim\_product | Товары | product\_key, product\_ver, valid\_from, valid\_to, current\_flag, ean\_code |
| dim\_customer | Клиенты | customer\_key, segment\_ver, valid\_from, valid\_to, is\_active, cohort\_month |
| dim\_warehouse | Склады | warehouse\_key, location\_key, capacity\_sqm, temp\_zone |
| dim\_geo | География | geo\_key, city, region, federal\_district, timezone |
| dim\_time | Время | time\_key, full\_date, day\_of\_week, is\_holiday, quarter |

**Факты (Facts)**

| **Таблица** | **Гранулярность** | **Метрики** |
| --- | --- | --- |
| fact\_sales | Чек/позиция | sales\_key, product\_key, customer\_key, store\_key, qty, amount, discount |
| fact\_inventory | Склад/день | inventory\_key, warehouse\_key, product\_key, opening\_balance, received |
| fact\_delivery | Заказ | delivery\_key, order\_key, carrier\_key, planned\_hours, delay\_hours |
| fact\_promotions | Акция/товар | promo\_key, product\_key, lift\_effect, redemption\_rate |

*3.1.4. Data Mart Layer (Витрины данных)*

Назначение: Готовые решения для бизнес-задач

Принципы:

1. Предметная ориентация:

- Отдельные витрины для финансов, логистики, маркетинга

- Поддержка различных уровней агрегации (детализированные и сводные)

2. Производительность:

- Материализованные представления для часто используемых отчетов

- Оптимизированные индексы под типовые запросы

3. Таблицы слоя

| **Витрина** | **Назначение** | **Ключевые показатели** |
| --- | --- | --- |
| dm.sales\_funnel | Воронка продаж | visits, carts, checkouts, conversion\_rate |
| dm.customer\_lifetime | LTV анализ | first\_purchase\_date, last\_purchase\_date, avg\_order\_value |
| dm.warehouse\_turnover | Оборот складов | turnover\_days, stock\_coverage, out\_of\_stock\_rate |
| dm.promo\_effectiveness | Эффективность акций | incremental\_sales, roi, halo\_effect |

***3.2. Жизненный цикл данных***

1. Поступление:

- SAP ERP: ежечасные выгрузки через SAP OData

- 1С Битрикс: REST API каждые 15 минут

- Excel: автоматизированный парсинг при поступлении файлов

2. Очистка:

- Стандартизация SKU товаров через ETL-сопоставление

- Обогащение геоданных (регион → федеральный округ)

3. Трансформация:

- Расчет производных показателей (наценка, маржа)

- Выявление аномалий (возвраты > 30% от суммы)

4. Публикация:

- Автоматическое обновление Power BI datasets

- Нотификации о критических изменениях KPI