|  |
| --- |
|  |
| DWH |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| История изменений | | | | | |
| Версия | Изменения | Автор | Дата | Рецензировано | |
| Рецензент | Дата |
| 0.1 | Начальная версия | Olga Hilko | 08-Nov-2017 |  |  |
| 0.2 | Описание бизнес-процесса, контекста, переход на русский язык | Olga Hilko | 08-Nov-2017 |  |  |
| 0.3 | Перевод на русский язык раздела | Olga Hilko | 10-Nov-2017 |  |  |
| 0.4 | Добавление второго зерна (раздел 2.2), описание фактов (раздел 2.4).Разделы Логическая схема, Поток загрузки данных (Data Flow), Стратегия партиционирования фактовых таблиц | Olga Hilko | 15-Nov-2017 |  |  |

Содержание

[1. Описание предметной области 3](#_Toc498532998)

[1.1. Миссия проекта / Бизнесс основа 3](#_Toc498532999)

[1.2. Словарь 3](#_Toc498533000)

[1.3. Недостатки существующих решений 3](#_Toc498533001)

[1.4. Преимущества от внедрения разрабоываемого хранилища данных 3](#_Toc498533002)

[2. Бизнес-контекст 3](#_Toc498533003)

[2.1. Описание бизнес-процесса 3](#_Toc498533004)

[2.2. Выбор зерна 4](#_Toc498533005)

[2.3. Выделение контекста 4](#_Toc498533006)

[2.4. Источники данных 5](#_Toc498533007)

[2.5. Описание факта 5](#_Toc498533008)

[3. Логическая схема 6](#_Toc498533009)

[3.1. Схема слоя BL\_3NF 6](#_Toc498533010)

[4. Поток загрузки данных (Data Flow) 8](#_Toc498533011)

[5. Стратегия партиционирования фактовых таблиц 8](#_Toc498533012)

[6. Стратегия параллельной загрузки 8](#_Toc498533013)

[7. Отчеты 8](#_Toc498533014)

# Описание предметной области

## Миссия проекта / Бизнесс основа

Миссией проекта является проектирование и разработка промышненной базы данных для хранения и последуюшей обработки информации о произведеных платетежах за реализуемое топливо и сопутствующие товары в сети АЗС «Белнефтехим».

## Словарь

|  |  |  |  |
| --- | --- | --- | --- |
| **Термин** | **Перевод** | **Сокра-щение** | **Description** |
| Хранилище данных | Data Warehouse | DWH (ХД) | Разрабатываемое харанилище данных на с использвоанием СУБД Oraсle 12с |
| Автозаправочная станция | Petrol Station / Station | АЗС | Точка продаж топлива и/или сопуствтующих товаров |
| Валюта | Currency |  | Валюта, в которой была произмедена оплата |
| Покупатель | Customer |  | Физическое или юридическое лицо, оплатившее отпущенные товары или топливо |
| Товар | Product |  | Сопуствующий товар, приобратемый покупателем на АЗС |
| Топливо | Fuel |  | Тип топлива, отпустаемого на АЗС |
| Прейскурант | PriceList |  | Исторические данные о цене товаров и/или топлива, действовавшей в течение определенного периода. |

## Недостатки существующих решений

Каждая заправочная станция имеет свою собственную базу данных по продажам, что вызывает некоторые проблемы:  
– низкая скорость обмена данными между отдельными заправочными станциями, головным офисом и бухгалтерией .  
– трудности при сборе информации для ежемесячных, квартальных и годовых отчетов.

## Преимущества от внедрения разрабоываемого хранилища данных

Новая структура DWH позволит:

– сделать обмен данными между подразделениями и управляющими отделами более быстрым, надежным и гибким,

– консодилировать исходные данные из информаионных систем отдельных АЗС в одном хранилище,

– повысить производительность информационных систем, предоставляющих информацию для лиц, принимающих решения,

– ускорить обработку данных,

– повысить качество

Также внедрение DWH будет способствовать:

– улучшению качества обслуживания клиентов.

# Бизнес-контекст

## Описание бизнес-процесса

Сохраняемая в проектируемом хранилище информация описывает процесс приобретения и оплаты топлива и сопутствующих товаров на заправочных станциях "Беларусьнефть". Основной задачей проекта является структуризация и консолидация данных о проведенных платежах.

Платеж (Payment) совершается клиентом (Customer) за один или несколько видов топлива (Fuel) и/или сопутствующих товаров (Product), относящихся к определенной категории (Category) по существующему прейскуранту (PriceList) на определенной заправочной станции (Station). Оплата производится в белорусских рублях, однако ранее имелась возможность оплаты в одной из трех валют (Currency): доллар США, евро, российский рубль. Отгрузка (отпуск) товаров и топлива осуществляется в момент оплаты, характеризуемый заданной датой (Date).

Клиент может быть как физическим лицом, так и юридическим.

## Выбор зерна

Так как основным бизнес-процессом является прием платежа и сохранение информации о нем, то в роли базового зерна выступает факт оплаты отпущенных товаров и/или топлива на определенную дату на определенной запарвочной станции.

Однако должна быть сохранена возможность анализа данных об оплаченных товарах и их количестве, для чего добавляется второе зерно, ядром которого является оплата определенного количества отпущенного товара.

## Выделение контекста

Дальнейшее использование данных о проведенных платежах предполагает их анализ в следующих разрезах:

– временном,

– по покупателям и их типам,

– по заправочным станциям,

– по категориям и видам реализованных товаров,

– по видам топлива.

Так как для проведения платежа необходимо иметь информацию о действующих на момент оплаты прейскурантах цен, то эту информацию также необходимо сохранить в DWH в качестве измерения.

Измерение **Station**:

– ID (суррогатный первичный ключ),

– номер,

– адрес,

– географические координаты,

– номер(а) автодорог(и),

– перечень реализуемых видов топлива,

– флаг Current (является ли значение текущим),

– дата появления информации в системе,

– дата изменения (для текущих значений с флагом Current = true дата равна 01-01-9999).

Измерение **Customer**:

– ID (суррогатный первичный ключ)

– тип (физическое или юридическое лицо):

для физического лица: имя, фамилия, отчество, дата рождения, пол,

для юридического лица: наименование юридического лица, УНП

– адрес (страна, индекс, область/регион, город, улица, дом/строение/корпус, квартира/офис),

– дата появления информации в системе (только дата начала, или две даты + флаг Current).

Измерение **Date**:

– значение даты (естественный первичный ключ),

– год, квартал (номер в году, номер вмете с годом), месяц (номер и название на разных языках), номер недели в году, день в месяце,

– день недели (номер и название на разных языках).

Измерение **Product**:

– ID (суррогатный первичный ключ),

– название товара,

– категория товара (возможно: подкатегория),

– флаг Current (является ли значение текущим),

– дата появления информации в системе,

– дата изменения (для текущих значений с флагом Current = true дата равна 01-01-9999)

Измерение **Fuel**:

– ID (суррогатный первичный ключ),

– код вида топлива (натуральный ключ),

– название вида топлива.

Измерение **PriceList**

– ID (суррогатный первичный ключ),

– период действия цен (даты с и по ),

– цена и валюта,

– тип топлива или товар,

– единица, за которую установлена цена,

– дата появления информации в системе.

Измерение **Currency**:

– ID (суррогатный первичный ключ),

– код валюты (натуральный альтернативный ключ),

– название валюты.

## Источники данных

|  |  |  |
| --- | --- | --- |
| **Сущность** | **Тип источника** | **Исходное расположение** |
| Petrol Station / Station | Excel | http://www.belorusneft.by/beloil-map/?lang=ru |
| **Currency** | Текстовый документ | 4 типа, будет заполняться вручную согласно **PriceList** |
| **Customer** |  | Будут генерироваться в зависимости от типа сторонними источниками |
| **Product** |  | Пока нет мыслей |
| **Fuel** | Текстовый документ | только 6 видов, будет заполняться вручную согласно **PriceList** |
| **PriceList** | Excel | http://www.belorusneft.by/sitebeloil/ru/center/azs/center/fuelandService/price/# |

## Описание факта

В БД содержится две фактовые таблицы, одна из которых (Payments) выступает в качестве Conformed Dimension для второй (Product\_Payments). Для обеспечение этой связи вводится суррогатный первичный ключ для платежей (ID\_Payment).

В фактовую таблицу **Payments** включаются внешние ключи, соответствующие связанным измерениям :

– время платежа,

– тип платежа,

– валюта платежа,

– покупатель (если известен),

– заправочная станция.

В качестве меры используется уплаченная сумма.

Дополнительны атрибутом, по которому на будет проводиться анализ, но данные которого необходимо сохранить, является время проведения оплаты.

Кроме Conformed Dimension (Payments) в фактовую таблицу **Product\_Payments** включаются внешние ключи, соответствующие связанным измерениям:

– единица измерения отпущенного товара/топлива,

– продукт,

– валюта платежа,

В качестве мер используются :

– количество отпущенного товара/топлива,

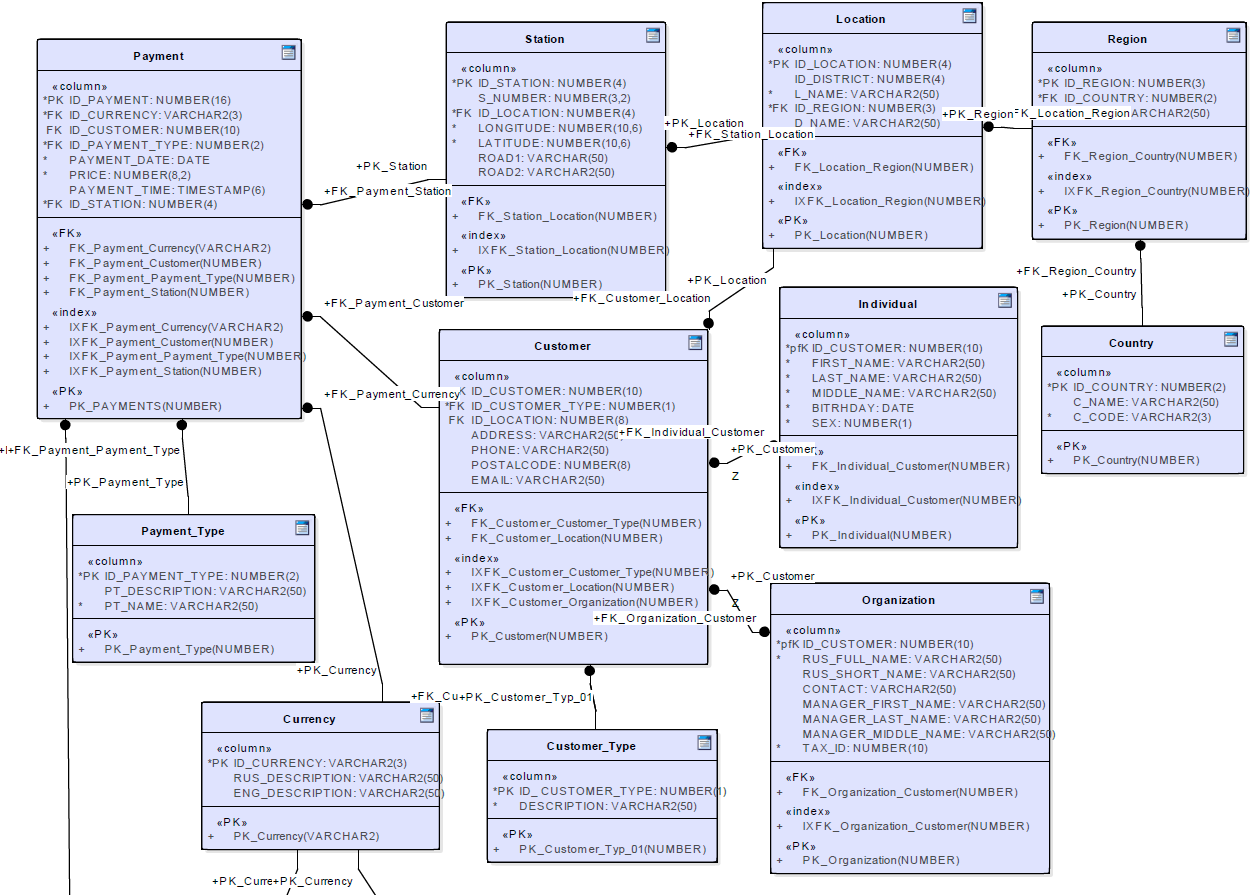
– цена за единицу отпущенного товара/топлива на момент оплаты,

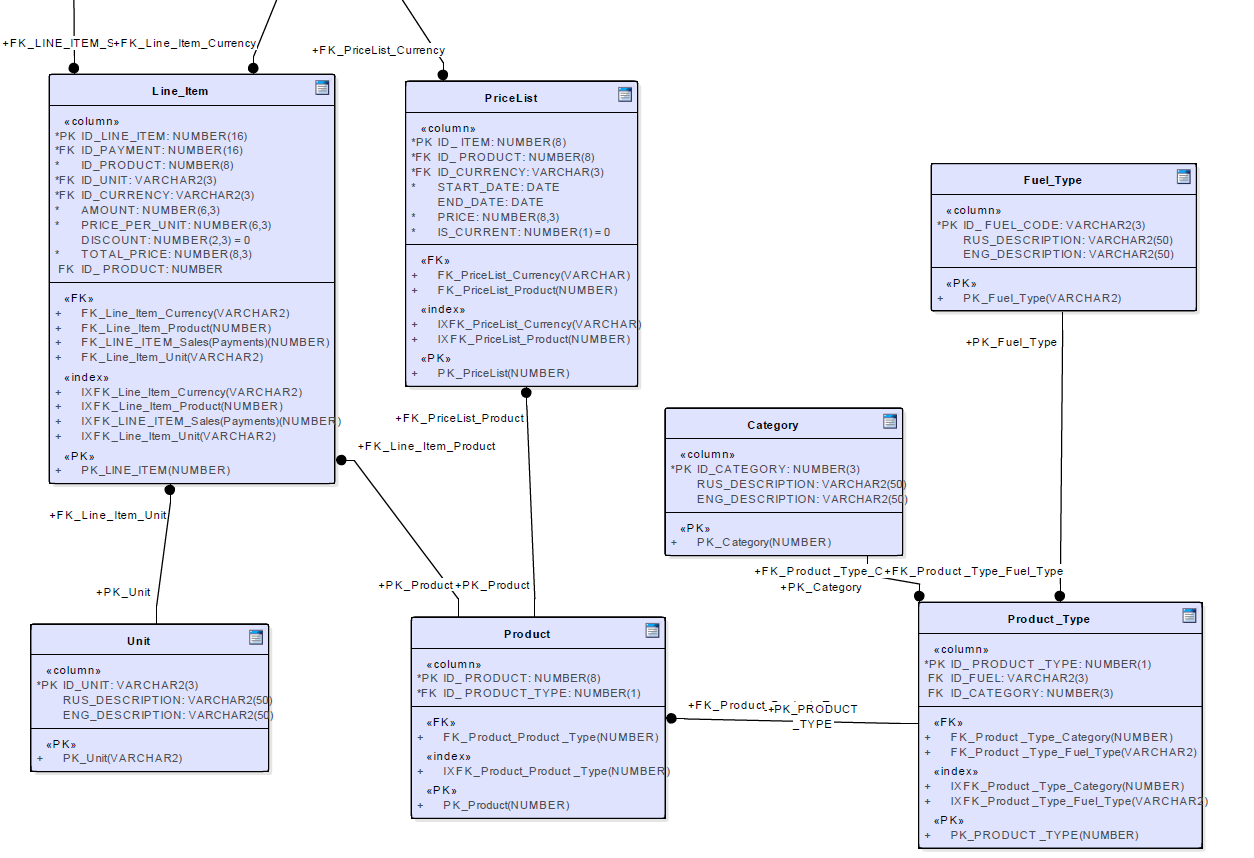
– предоставляемая скидка,

– итоговая стоимость заданного количества товара.

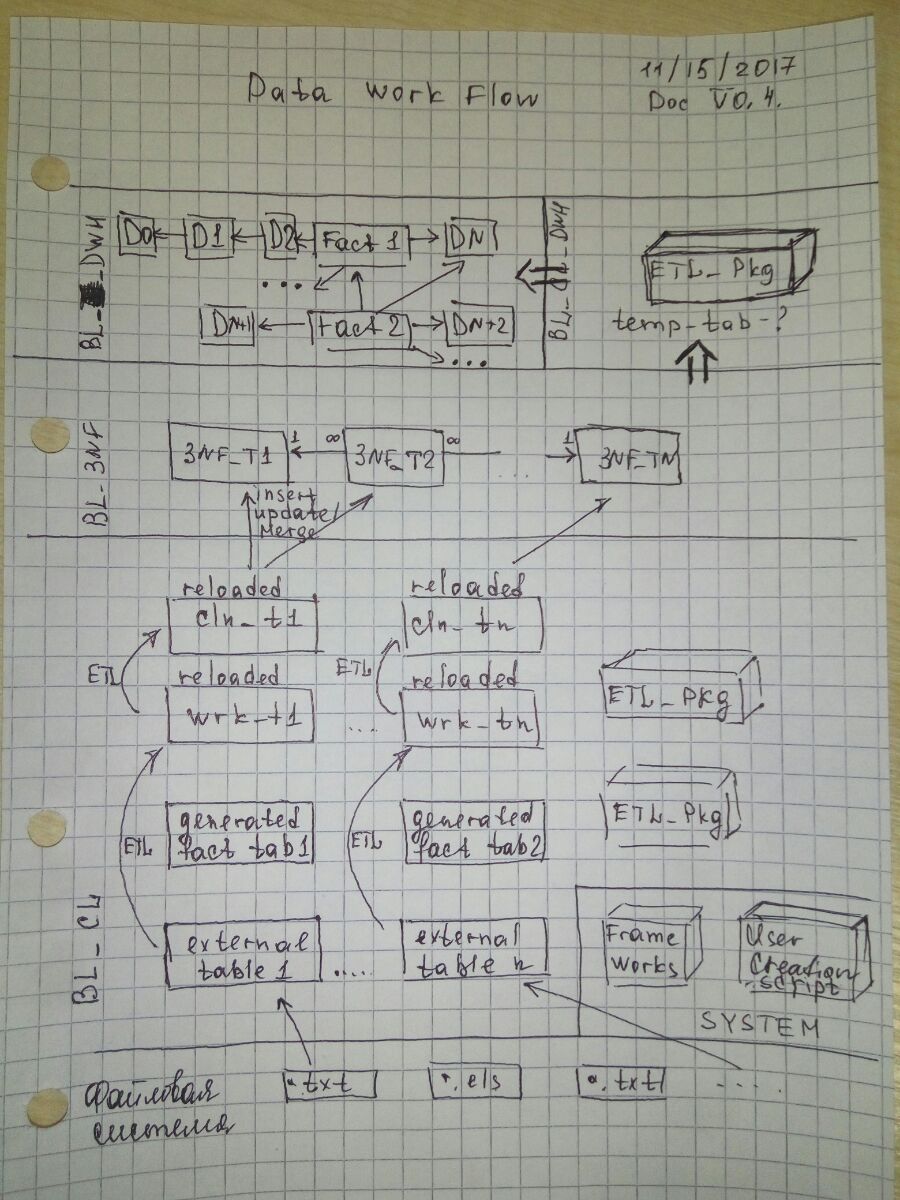
# Логическая схема

## Схема слоя BL\_3NF





# Поток загрузки данных (Data Flow)



# Стратегия партиционирования фактовых таблиц

Партиционирование фактовых таблицы целесообразно производить по дате оплаты, причем:

– в первой партиции целесообразно размещать данные по продажам за все прошедшие годы, кроме текущего календарного и прошлого (на настоящим момент – до 01-01-2016)

– во второй – данные за прошедший год (с 01-01-2016 по 01-01-2017),

– в третьей – данные за текущий год (с 01-01-2017 по sysdate),

– в четвертой – данные будущих периодов (заготовка для переходного периода при наступлении нового календарного года).

# Стратегия параллельной загрузки

В разработке

# Отчеты

В разработке