DWH – Task 7

Contents

[1. 4 steps Dimensional Process 2](#_Toc498176915)

[1.1. Шаг 1. Бизнес-процесс. 2](#_Toc498176916)

[1.2. Шаг 2. Зерно Grain. 2](#_Toc498176917)

[1.3. Шаг 3. Измерения Dimensions. 2](#_Toc498176918)

[1.4. Шаг 4. Факты Facts. 3](#_Toc498176919)

[2. DWH Star schema. 5](#_Toc498176920)

[3. Data source generation 6](#_Toc498176921)

# 4 steps Dimensional Process

## Шаг 1. Бизнес-процесс.

Учет продажи товаров различных марок в интернет-магазине Wildberries с учетом различных показателей.

## Шаг 2. Зерно Grain.

Зерном будет одна продажа конкретного товара одной торговой марки из одной категории одному клиенту в одном пункте самовывоза в пределах одного заказа.

## Шаг 3. Измерения Dimensions.

В моем хранилище будет реализовано 6 dimensions:

Измерение DatesDim – содержит необходимые данные для дальнейшего анализа в разрезе времени – дата, день недели, название дня недели, день месяца, день года, номер месяца, названия месяца, квартал, полугодие, год. Будет применяться как измерение для дат заказов и дат оплат товаров.

CREATE TABLE "DatesDim"

( "DATE\_ID" NUMBER(8) NOT NULL,

"FULL\_DATE" DATE,

"DAY\_WEEK" NUMBER(1),

"DAY\_WEEK\_NAME" VARCHAR2(50),

"DAY\_MONTH" NUMBER(2),

"DAY\_YEAR" NUMBER(3),

"MONTH\_NUM" NUMBER(2),

"MONTH\_NAME" VARCHAR2(50),

"QUARTER" NUMBER(1),

"HALF\_YEAR" NUMBER(1),

"YEAR" NUMBER(4)

);

Измерение CustomersDim – dimension, содержащий информацию о клиентах магазина. Это большой dimension, который включает более 100 000 строк.

CREATE TABLE "CustomersDim"

( "CUSTOMER\_ID" NUMBER(8) NOT NULL,

"CUSTOMER\_CODE" NUMBER(6,2),

"LAST\_NAME" VARCHAR2(50),

"FIRST\_NAME" VARCHAR2(50),

"MIDDLE\_NAME" VARCHAR2(50),

"AGE" VARCHAR2(3),

"GENDER" VARCHAR2(50),

"CITY" VARCHAR2(50),

"REGION" VARCHAR2(50)

);

Измерение ProductsDim – измерение, которое содержит необходимую информацию о товаре, бренде (производителе) и категории товара. Является иерархическим dimension.

CREATE TABLE "ProductsDim"

( "PRODUCT\_ID" NUMBER(8) NOT NULL,

"PRODUCT\_VENDOR\_CODE" NUMBER(9),

"PRODUCT\_TYPE" VARCHAR2(50),

"PRODUCT\_CATEGORY" VARCHAR2(50)

);

Измерение PayDeliveryDim – описание способа оплаты и способа доставки. Идет как Junk dimension.

CREATE TABLE "PayDeliveryDim"

( "PAYDELIVERY\_ID" NUMBER(2) NOT NULL,

"DELIVERY\_METOD" VARCHAR2(50),

"PAYMENT\_METHOD" VARCHAR2(50)

);

Измерение PickupPointsDim – информация о пунктах самовывоза.

CREATE TABLE "PickupPointsDim"

( "PICKUPPOINT\_ID" NUMBER(3) NOT NULL,

"PP\_CODE" VARCHAR2(3),

"PP\_ADDRESS" VARCHAR2(50),

"PP\_CITY" VARCHAR2(50),

"PP\_REGION" VARCHAR2(50)

);

Измерение BrendsDim – информация о брендах товаров и странах производителях.

CREATE TABLE "BrendsDim"

( "BREND\_ID" NUMBER(8) NOT NULL,

"BREND\_NAME" VARCHAR2(50),

"BREND\_COUNTRY" VARCHAR2(50)

);

## Шаг 4. Факты Facts.

Таблица фактов будет содержать количество проданных единиц товара и сумму реализации за месяц в разрезе товара, бренда, категории товара, клиента, способа доставки и оплаты.

CREATE TABLE "SalesFacts"

(

"CUSTOMER\_ID" NUMBER(8) NOT NULL,

"ORDERDATE\_ID" NUMBER(8) NOT NULL,

"PAYDATE\_ID" NUMBER(8) NOT NULL,

"DELIVERY\_ID" NUMBER(8) NOT NULL,

"PRODUCT\_ID" NUMBER(8) NOT NULL,

"PICKUPPOINT\_ID" NUMBER(3),

"PAYDELIVERY\_ID" NUMBER(2) NOT NULL,

"BREND\_ID" NUMBER(8),

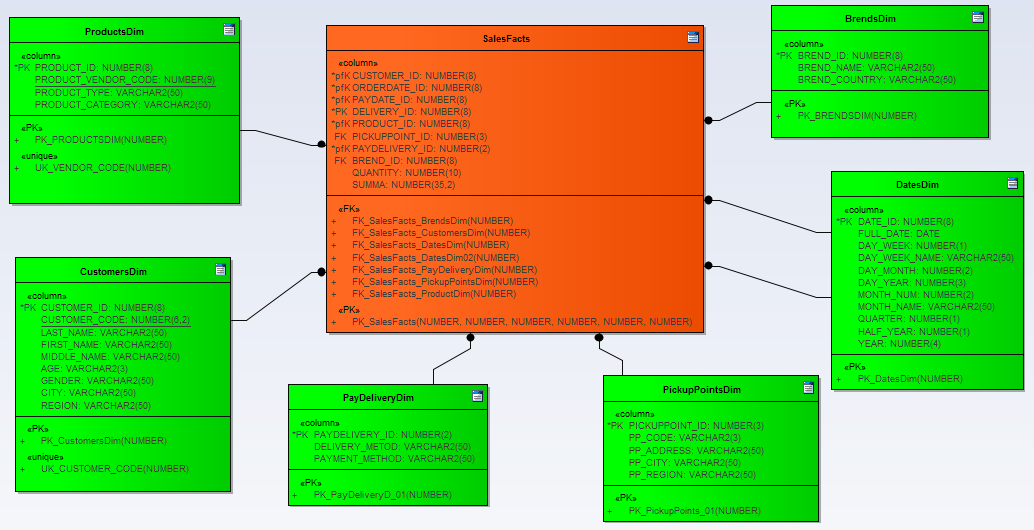
"QUANTITY" NUMBER(10),

"SUMMA" NUMBER(35,2)

);

# DWH Star schema.

Все связи между dimensions и таблицей фактов реализованы как один ко многим.



# Data source generation

Основная информация о товарах, категориях, брендах, способах доставки, оплаты и пунктах самовывоза взята на официальном сайте: <https://www.wildberries.by/>

<https://www.wildberries.by/services/sposoby-oplaty>

<https://www.wildberries.by/services/besplatnaya-dostavka#pickupPoints>

<https://www.wildberries.by/catalog/0/brand.aspx>

Населенные пункты РБ - <https://www.bygeo.ru/materialy/naselenie-belarusi/>

Какая-то часть данных будет взята с сайтов, какая-то сгенерирована c помощью других интернет сайтов, например http://www.fakenamegenerator.com, также будет применена генерация с помощью PLSQL или вставка данных вручную.