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));

ALTER TABLE "DatesDim"

ADD CONSTRAINT "PK\_DATESDIM"

PRIMARY KEY ("DATE\_ID") USING INDEX;

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(7),

"LAST\_NAME" VARCHAR2(50),

"FIRST\_NAME" VARCHAR2(50),

"MIDDLE\_NAME" VARCHAR2(50),

"AGE" VARCHAR2(3),

"GENDER" VARCHAR2(50),

"PERSONAL\_DISCOUNT" NUMBER(2),

"CITY" VARCHAR2(50),

"REGION" VARCHAR2(50));

ALTER TABLE "CustomersDim"

ADD CONSTRAINT "PK\_CUSTOMERSDIM"

PRIMARY KEY ("CUSTOMER\_ID") USING INDEX;

ALTER TABLE "CustomersDim"

ADD CONSTRAINT "UK\_CUSTOMER\_CODE" UNIQUE ("CUSTOMER\_CODE") USING INDEX;

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));

ALTER TABLE "ProductsDim"

ADD CONSTRAINT "PK\_PRODUCTSDIM"

PRIMARY KEY ("PRODUCT\_ID") USING INDEX;

ALTER TABLE "ProductsDim"

ADD CONSTRAINT "UK\_VENDOR\_CODE" UNIQUE ("PRODUCT\_VENDOR\_CODE") USING INDEX;

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

CREATE TABLE "PayDeliveriesDim"

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

"DELIVERY\_NAME" VARCHAR2(50),

"PAYOPTION\_NAME" VARCHAR2(50));

ALTER TABLE "PayDeliveriesDim"

ADD CONSTRAINT "PK\_PAYDELIVERYD\_01"

PRIMARY KEY ("PAYDELIVERY\_ID") USING INDEX;

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));

ALTER TABLE "PickupPointsDim"

ADD CONSTRAINT "PK\_PICKUPPOINTS\_01"

PRIMARY KEY ("PICKUPPOINT\_ID") USING INDEX;

Измерение BrandsDim – информация о брендах товаров.

CREATE TABLE "BrandsDim"

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

"BRAND\_NAME" VARCHAR2(50));

ALTER TABLE "BrandsDim"

ADD CONSTRAINT "PK\_BRANDSDIM"

PRIMARY KEY ("BRAND\_ID") USING INDEX;

## 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 "SalesItemsFacts"

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

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

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

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

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

"PICKUPPOINT\_ID" NUMBER(3),

"BRAND\_ID" NUMBER(8),

"QUANTITY" NUMBER(10),

"ITEM\_SUM" NUMBER(35,2),

"DISCOUNT\_SUM" NUMBER(35,2),

"TOTAL\_SUM" NUMBER(35,2));

ALTER TABLE "SalesItemsFacts"

ADD CONSTRAINT "PK\_SALESITEMSFACTS"

PRIMARY KEY ("CUSTOMER\_ID","PAYDATE\_ID","PRODUCT\_ID","ORDERDATE\_ID","PAYDELIVERY\_ID") USING INDEX;

ALTER TABLE "SalesItemsFacts"

ADD CONSTRAINT "FK\_SALESITEMSFACTS\_BRANDSDIM"

FOREIGN KEY ("BRAND\_ID") REFERENCES "BrandsDim" ("BRAND\_ID");

ALTER TABLE "SalesItemsFacts"

ADD CONSTRAINT "FK\_SALESITEMSFACTS\_CUSTOMERSDIM"

FOREIGN KEY ("CUSTOMER\_ID") REFERENCES "CustomersDim" ("CUSTOMER\_ID");

ALTER TABLE "SalesItemsFacts"

ADD CONSTRAINT "FK\_SALESITEMSFACTS\_DATESDIM"

FOREIGN KEY ("ORDERDATE\_ID") REFERENCES "DatesDim" ("DATE\_ID");

ALTER TABLE "SalesItemsFacts"

ADD CONSTRAINT "FK\_SALESITEMSFACTS\_DATESDIM02"

FOREIGN KEY ("PAYDATE\_ID") REFERENCES "DatesDim" ("DATE\_ID");

ALTER TABLE "SalesItemsFacts"

ADD CONSTRAINT "FK\_SALESITEMSFACTS\_PAYDELIVERYDIM"

FOREIGN KEY ("PAYDELIVERY\_ID") REFERENCES "PayDeliveriesDim" ("PAYDELIVERY\_ID");

ALTER TABLE "SalesItemsFacts"

ADD CONSTRAINT "FK\_SALESITEMSFACTS\_PICKUPPOINTSDIM"

FOREIGN KEY ("PICKUPPOINT\_ID") REFERENCES "PickupPointsDim" ("PICKUPPOINT\_ID");

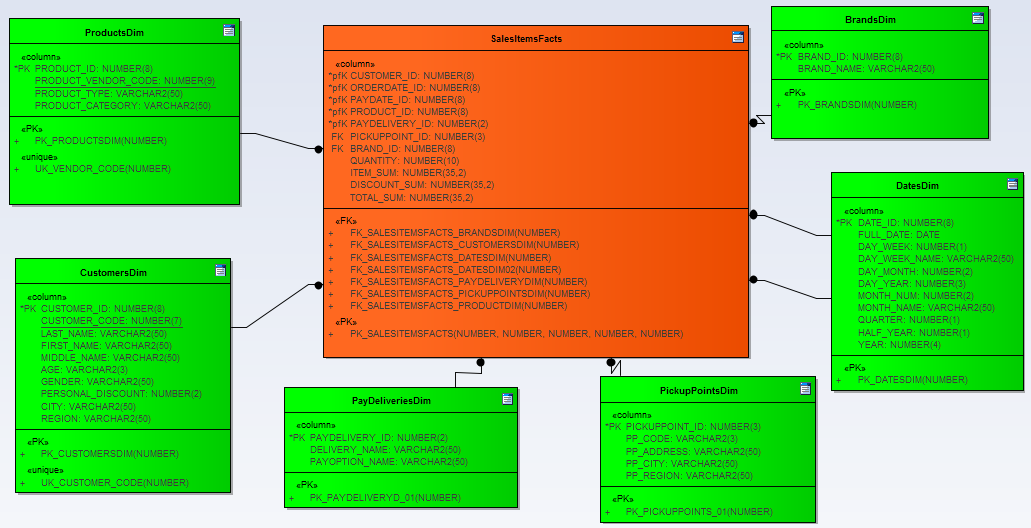
ALTER TABLE "SalesItemsFacts"

ADD CONSTRAINT "FK\_SALESITEMSFACTS\_PRODUCTDIM"

FOREIGN KEY ("PRODUCT\_ID") REFERENCES "ProductsDim" ("PRODUCT\_ID");

# 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.