DWH – Task 8

Contents

[1. Data Modeling Task 2](#_Toc498197760)

[2. Analytical task 3](#_Toc498197761)

# Data Modeling Task

Создание структуры dimension DatesDim. При пересоздании необходимо откомментировать и выполнить команду drop.

-- drop table datesdim;

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;

Скрипт DatesDim\_GenData вначале очищает все строки в таблице DatesDim, а затем заполняет их данными, начиная с 1 января 2010 года и до 17 мая 2037 года.

truncate table datesdim;

insert into datesdim

select 10000+level as date\_id,

to\_date('31/12/2009','dd.mm.yyyy') + numtodsinterval(level,'day') as full\_date,

to\_number(to\_char(to\_date('31.12.09','dd.mm.yyyy') + numtodsinterval(level,'day'), 'd')) as day\_week,

to\_char(to\_date('31.12.09','dd.mm.yyyy')+ numtodsinterval(level,'day'),'day', 'nls\_date\_language=russian') as day\_week\_name,

to\_number(to\_char(to\_date('31/12/2009','dd.mm.yyyy') + numtodsinterval(level,'day'), 'dd')) as day\_month,

to\_number(to\_char(to\_date('31/12/2009','dd.mm.yyyy') + numtodsinterval(level,'day'),'ddd')) as day\_year,

to\_number(to\_char(to\_date('31/12/2009','dd.mm.yyyy') + numtodsinterval(level,'day'),'mm')) as month\_num,

to\_char(to\_date('31/12/2013','dd.mm.yyyy') + numtodsinterval(level,'day'), 'month', 'nls\_date\_language=russian') as month\_name,

to\_number(to\_char(to\_date('31/12/2009','dd.mm.yyyy') + numtodsinterval(level,'day'), 'Q')) as quarter,

case

when to\_char(to\_date('31/12/2009','dd.mm.yyyy') + numtodsinterval(level,'day'),'MM')<'07'

then 1

else 2

end as half\_year,

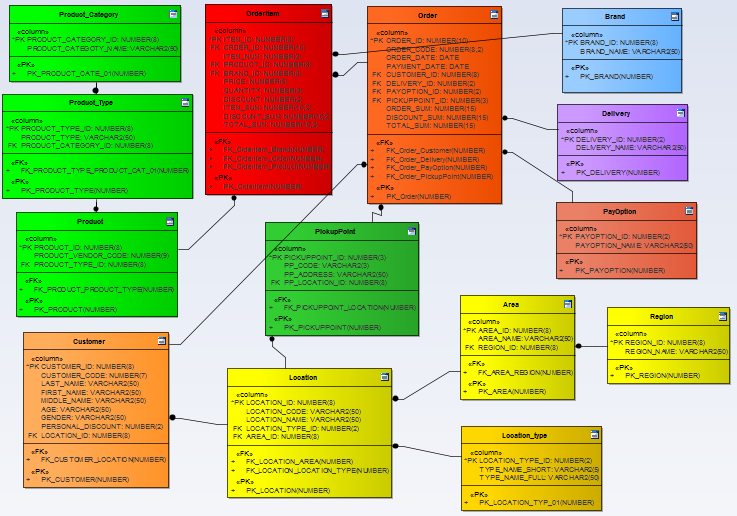
to\_number(to\_char(to\_date('31/12/2009','dd.mm.yyyy') + numtodsinterval(level,'day'),'YYYY')) as year

from dual connect by level < 10000

;

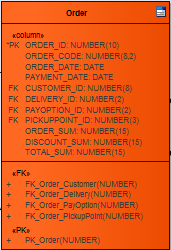
commit;

# Analytical task



Одна из основных таблиц Order содержит в себе информацию о заказе, его номер, дата заказа, дата оплаты, идентификатор клиента, совершившего данный заказ, идентификатор типа оплаты, идентификатор типа доставки и идентификатор пункта самовывоза, а также сумму заказа, сумму предоставленной скидки и общая сумма, оплаченная клиентом.

Один заказ Order\_Code может быть сделан одним клиентом Customer\_Id в один календарный день Order\_Date. Он может быть оплачен с помощью одного из вариантов PayOption\_id и может быть доставлен курьером, почтой или получен в пункте самовывоза Delivery\_Id , в третьем случае указывается какой именно пункт PickupPoint\_Id. Фиксируются две даты – дата заказа Order\_Date и дата оплаты Payment\_Date.



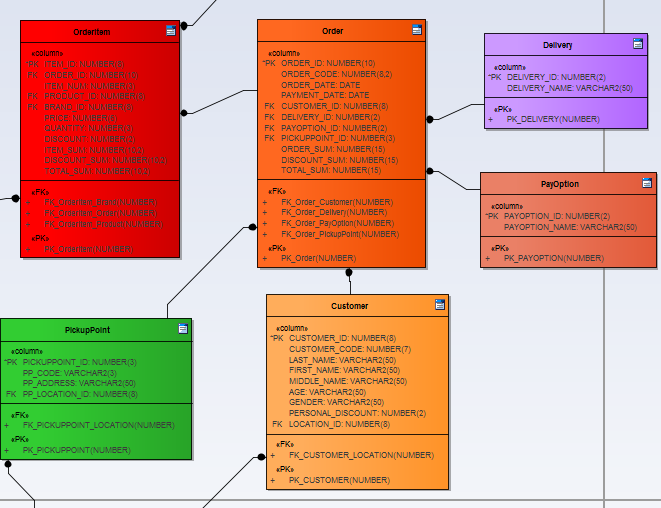
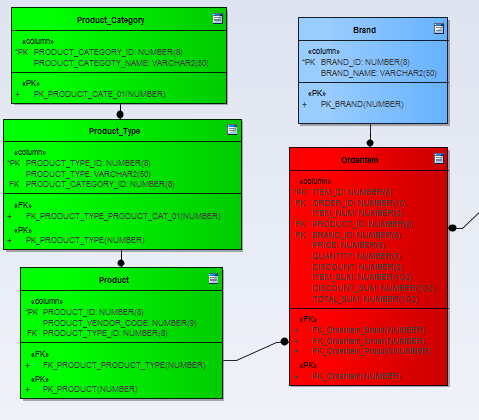


Таблица OrderItem содержит в себе информацию о пунктах заказа, идентификатор заказа, идентификатор товара, идентификатор бренда товара, а также цену товара, его количество, процент скидки (суммируются скидка клиента и разовые скидки), сумму за указанный товара, сумму скидки и итоговую сумму к оплате.

Один пункт заказа Item\_Num относится только к одному заказу Order\_Id. Один заказ может содержать N пунктов. Пункт относится исключительно к одному товару Product\_Id и соответственно только к 1 бренду.



Для понимания приведу пример данных для ветки товаров: *Product* Платье Bracegirdle, артикул: 2120550 🡪 *Product\_Type* [Платья-миди](https://www.wildberries.by/catalog/zhenshchinam/odezhda/platya-midi) 🡪 *Product\_Category* Платья



Пункты самовывоза и клиенты имеют location\_id, который связывает их с таблицей location. Поскольку доставка производится по Беларуси, то это отражается в формировании ветки локаций. Пример для понимания:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Location\_Code | Location\_Name | Area\_Name | Region\_Name | Type\_name\_short |
| 6240833091 | Нарочь | Мядельский | Минская область | аг. |

