Cоздадим целевые таблицы dds:

CREATE TABLE dds.kdz\_7\_flights (

year int NULL,

quarter int NULL,

month int NULL,

flight\_scheduled\_date date NULL,

flight\_actual\_date date NULL,

flight\_dep\_scheduled\_ts timestamp NOT NULL,

flight\_dep\_actual\_ts timestamp NULL,

report\_airline varchar(10) NOT NULL,

tail\_number varchar(10) NOT NULL,

flight\_number\_reporting\_airline varchar(15) NOT NULL,

airport\_origin\_dk int NULL, --постоянный ключ аэропорта. нужно взять из таблицы аэропортов

origin\_code varchar(5) null,

airport\_dest\_dk int NULL, --постоянный ключ аэропорта. нужно взять из таблицы аэропортов

dest\_code varchar(5) null,

dep\_delay\_minutes float NULL,

cancelled int NOT NULL,

cancellation\_code char(1) NULL,

weather\_delay float NULL,

air\_time float NULL,

distance float NULL,

loaded\_ts timestamp default(now()),

CONSTRAINT lights\_pk PRIMARY KEY (flight\_dep\_scheduled\_ts, flight\_number\_reporting\_airline, origin\_code, dest\_code)

);

CREATE TABLE dds.kdz\_7\_airport\_weather (

airport\_dk int NOT NULL, -- постоянный ключ аэропорта. нужно взять из таблицы аэропортов

weather\_type\_dk char(6) NOT NULL, -- постоянный ключ типа погоды. заполняется по формуле

cold smallint default(0),

rain smallint default(0),

snow smallint default(0),

thunderstorm smallint default(0),

drizzle smallint default(0),

fog\_mist smallint default(0),

t int NULL,

max\_gws int NULL,

w\_speed int NULL,

date\_start timestamp NOT NULL,

date\_end timestamp NOT NULL default('3000-01-01'::timestamp),

loaded\_ts timestamp default(now()),

PRIMARY KEY (airport\_dk, date\_start)

);

Cоздадим процедуру ETL для заполнения таблицы уровня dds:

drop table if exists kdz\_7\_etl.flight\_dds;

create table if not exists kdz\_7\_etl.flight\_dds as

with cte as ( -- избежали дубликацию джоинов, поэтому в конце джоинили два раза нашу cte, а не все джоины два раза

select airport\_dk, src\_icao\_id, src\_iata\_id

from dds.airport t1

join dwh.id\_airport t2

on t1.icao\_code = t2.src\_icao\_id)

select

flight\_year as year

,flight\_quarter as quarter

,flight\_month as month

,flight\_date as flight\_scheduled\_date

,case when cancelled = '0'

then flight\_date + crs\_dep\_time

else null

end as flight\_actual\_date

,flight\_date + crs\_dep\_time as flight\_dep\_scheduled\_ts

,case when cancelled = '0'

then flight\_date + crs\_dep\_time

else null

end as flight\_dep\_actual\_ts

,reporting\_airline as report\_airline

,tail\_number as tail\_number

,flight\_number as flight\_number\_reporting\_airline

,t2.airport\_dk as airport\_origin\_dk

,origin as origin\_code

,t3.airport\_dk as airport\_dest\_dk

,destination as dest\_code

,dep\_delay\_min as dep\_delay\_minutes

,cancelled as cancelled

,cancellation\_code as cancellation\_code

,weather\_delay as weather\_delay

,air\_time as air\_time

,distance as distance

,current\_date as loaded\_ts

from kdz\_7\_staging.flights t1

join cte t2

on t1.origin = t2.src\_iata\_id

join cte t3

on t1.destination = t3.src\_iata\_id

where t2.src\_iata\_id = 'ANC' or t3.src\_iata\_id = 'ANC'

;

drop table if exists kdz\_7\_etl.weather\_dds;

create table if not exists kdz\_7\_etl.weather\_dds as

with t1 as ( -- создаем виртуальную таблицу

select

local\_datetime as date\_start

,case when dewpoint\_temp < 0 then 1 else 0 end as cold

,case when lower(phenomena\_observed) like '%rain%' or lower(phenomena\_significant) like '%rain%'

then 1

else 0

end as rain

,case when lower(phenomena\_observed) like '%snow%' or lower(phenomena\_significant) like '%snow%'

then 1

else 0

end as snow

,case when lower(phenomena\_observed) like '%thunderstorm%' or lower(phenomena\_significant) like '%thunderstorm%'

then 1

else 0

end as thunderstorm

,case when lower(phenomena\_observed) like '%drizzle%' or lower(phenomena\_significant) like '%drizzle%'

then 1

else 0

end as drizzle

,case when lower(phenomena\_observed) like '%fog%' or lower(phenomena\_significant) like '%fog%'

or lower(phenomena\_observed) like '%mist%' or lower(phenomena\_significant) like '%mist%'

then 1

else 0

end as fog\_mist

,max\_gust\_speed as max\_gws

,dewpoint\_temp as t

,wind\_speed as w\_speed

,lead(local\_datetime) over (order by local\_datetime) as date\_end

,t4.airport\_dk as airport\_dk -- используем таблицу перекодировки ключей

from kdz\_7\_staging.weather t1

join (select max(loaded\_ts) as loaded\_ts from kdz\_7\_staging.weather) t2

on t1.loaded\_ts=t2.loaded\_ts

join dwh.id\_airport t3

on t1.icao\_code = t3.src\_icao\_id

join dds.airport t4

on t3.src\_icao\_id = t4.icao\_code

where t4.icao\_code = 'PANC'

)

select

airport\_dk

,date\_start

,concat(cold, rain, snow, thunderstorm, drizzle, fog\_mist) wheater\_type\_dk

,cold

,rain

,snow

,thunderstorm

,drizzle

,fog\_mist

,t

,max\_gws

,w\_speed

,date\_end

,current\_date as loaded\_ts

from t1

where date\_end is not null -- обходим одну строчку в данных, date\_end которой не попал период первого квартала 2021

;

Обновим и загрузим данные на уровень dds:

insert into kdz\_7\_dds.flights (year,quarter,month,flight\_scheduled\_date,flight\_actual\_date,

flight\_dep\_scheduled\_ts,flight\_dep\_actual\_ts,report\_airline,tail\_number,flight\_number\_reporting\_airline,airport\_origin\_dk,

origin\_code,airport\_dest\_dk,dest\_code,dep\_delay\_minutes,cancelled,cancellation\_code,weather\_delay,air\_time,distance,loaded\_ts )

select

year

,quarter

,month

,flight\_scheduled\_date

,flight\_actual\_date

,flight\_dep\_scheduled\_ts

,flight\_dep\_actual\_ts

,report\_airline

,tail\_number

,flight\_number\_reporting\_airline

,airport\_origin\_dk,

origin\_code

,airport\_dest\_dk

,dest\_code

,dep\_delay\_minutes

,cancelled

,cancellation\_code

,weather\_delay

,air\_time

,distance

,loaded\_ts

from kdz\_7\_etl.flight\_dds

where tail\_number is not null --обходим сломанную строчку в исходных данных;

insert into kdz\_7\_dds.airport\_weather (airport\_dk, weather\_type\_dk, cold, rain, snow, thunderstorm, drizzle, fog\_mist, t, max\_gws,w\_speed, date\_start, date\_end, loaded\_ts)

select

airport\_dk

,wheater\_type\_dk

,cold

,rain

,snow

,thunderstorm

,drizzle

,fog\_mist

,t

,max\_gws

,w\_speed

,date\_start

,date\_end

,loaded\_ts

from kdz\_7\_etl.weather\_dds;