

# Airports\_Fligths

FINISHED

```
--создание и наполнение объектов

set hive.stats.autogather=true;
set hive.vectorized.execution.enabled=true;
set hive.vectorized.execution.reduce.enabled=true;
set hive.auto.convert.join=true;
set hive.auto.convert.join.noconditionaltask=true;
set hive.exec.compress.intermediate=true;

create database if not exists airline_db;
use airline_db;

create external table if not exists raw_airports (
  airport_id int,
  city string,
  state string,
  name string
)
comment 'Аэропорты'
row format delimited
fields terminated by ','
stored as textfile
location '/user/maria_dev/data/airports'
tblproperties ("skip.header.line.count"="1");

create table if not exists airports (
  airport_id int not null comment 'id аэропорта',
  city string comment 'город аэропорта',
  state string comment 'штат аэропорта',
  name string comment 'название аэропорта'
)
comment 'Аэропорты'
clustered by (airport_id)
sorted by (airport_id)
into 4 buckets
stored as orc
tblproperties (
  "orc.bloom.filter.columns"="airport_id",
  "orc.compress"="SNAPPY"
);

insert overwrite table airports
select * from raw_airports;

create external table raw_flights (
  day_of_month int,
  day_of_week string,
  carrier string,
  origin_airport_id int,
  dest_airport_id int,
  dep_delay int,
  arr_delay int
)
row format delimited
fields terminated by ','
stored as textfile
location '/user/maria_dev/data/flights'
tblproperties ("skip.header.line.count"="1");

create table if not exists flights (
  day_of_month int not null comment 'день месяца',
  day_of_week string not null comment 'день недели',
  origin_airport_id int not null comment 'аэропорт вылета',
  dest_airport_id int not null comment 'аэропорт прилета',
  dep_delay int comment 'задержка вылета',
  arr_delay int comment 'задержка прибытия'
)
comment 'Перелеты'
partitioned by (carrier string comment 'перевозчик')
clustered by (origin_airport_id)
sorted by (day_of_month)
into 16 buckets
stored as orc
tblproperties (
  "orc.compress"="SNAPPY",
  "orc.bloom.filter.columns"="origin_airport_id,dest_airport_id"
);

set hive.exec.dynamic.partition = true;
set hive.exec.dynamic.partition.mode = nonstrict;

insert into flights partition(carrier)
select
  day_of_month,
  day_of_week,
  origin_airport_id,
  dest_airport_id,
  dep_delay,
  arr_delay,
  carrier -- д. быть последней
from raw_flights;

create materialized view if not exists v_airports_flights
comment 'всего аэропортов и перелетов'
stored as parquet
as
  select 'all quantity airports' as src, count(*) as qty from airports
  union
  select 'all quantity flighths' as src, count(*) as qty from flights;

create materialized view if not exists v_dayofweek_flights
comment 'количество перелетов по дням недели в разрезе перевозчиков'
```

## Airports Flights

```

stored as orc
tblproperties ("orc.compress"="SNAPPY")
as
select
  carrier,
  sum(case when day_of_week = 1 then 1 else 0 end) as mon_qty,
  sum(case when day_of_week = 2 then 1 else 0 end) as tue_qty,
  sum(case when day_of_week = 3 then 1 else 0 end) as wed_qty,
  sum(case when day_of_week = 4 then 1 else 0 end) as thu_qty,
  sum(case when day_of_week = 5 then 1 else 0 end) as fri_qty,
  sum(case when day_of_week = 6 then 1 else 0 end) as sat_qty,
  sum(case when day_of_week = 7 then 1 else 0 end) as sun_qty
from flights
group by carrier;

alter table v_dayofweek_flights change column carrier carrier string comment 'перевозчик';
alter table v_dayofweek_flights change column mon_qty mon_qty bigint comment 'количество перелетов в понедельник';
alter table v_dayofweek_flights change column tue_qty tue_qty bigint comment 'количество перелетов в вторник';
alter table v_dayofweek_flights change column wed_qty wed_qty bigint comment 'количество перелетов в среду';
alter table v_dayofweek_flights change column thu_qty thu_qty bigint comment 'количество перелетов в четверг';
alter table v_dayofweek_flights change column fri_qty fri_qty bigint comment 'количество перелетов в пятницу';
alter table v_dayofweek_flights change column sat_qty sat_qty bigint comment 'количество перелетов в субботу';
alter table v_dayofweek_flights change column sun_qty sun_qty bigint comment 'количество перелетов в воскресенье';

create materialized view if not exists v_popularity_destination
comment 'количество перелетов по направлениям'
stored as orc
tblproperties ("orc.compress"="SNAPPY")
as
with route_stats as (
  select
    origin_airport_id,
    dest_airport_id,
    count(*) as qty
  from flights
  group by
    origin_airport_id,
    dest_airport_id
)
select
  t.qty,
  t.origin_airport_id as departure_id,
  dep.city as departure_city,
  dep.state as departure_state,
  dep.name as departure_air,
  t.dest_airport_id as arrival_id,
  arr.city as arrival_city,
  arr.state as arrival_state,
  arr.name as arrival_air
from route_stats t
join airports dep on dep.airport_id = t.origin_airport_id
join airports arr on arr.airport_id = t.dest_airport_id;

alter table v_popularity_destination change column qty qty bigint comment 'количество перелетов';
alter table v_popularity_destination change column departure_id departure_id string comment 'id аэропорта вылета';
alter table v_popularity_destination change column departure_city departure_city string comment 'город аэропорта вылета';
alter table v_popularity_destination change column departure_state departure_state string comment 'штат аэропорта вылета';
alter table v_popularity_destination change column departure_air departure_air string comment 'название аэропорта вылета';

alter table v_popularity_destination change column arrival_id arrival_id string comment 'id аэропорта прибытия';
alter table v_popularity_destination change column arrival_city arrival_city string comment 'город аэропорта прибытия';
alter table v_popularity_destination change column arrival_state arrival_state string comment 'штат аэропорта прибытия';
alter table v_popularity_destination change column arrival_air arrival_air string comment 'название аэропорта прибытия';

create materialized view if not exists v_airport_congestion
comment 'загруженность аэропортов перелетами'
stored as orc
tblproperties ("orc.compress"="SNAPPY")
as
with depart_stats as (
  select
    origin_airport_id as airport_id,
    count(*) as cnt
  from flights
  group by origin_airport_id
),
arrival_stats as (
  select
    dest_airport_id as airport_id,
    count(*) as cnt
  from flights
  group by dest_airport_id
),
total_traffic as (
  select
    airport_id,
    sum(cnt) as qty
  from (
    select airport_id, cnt from depart_stats
    union all
    select airport_id, cnt from arrival_stats
  ) t
  group by airport_id
  having (qty > 0)
)
select
  traf.airport_id,
  air.city,
  air.state,
  air.name,
  traf.qty
from total_traffic traf
join airports air on air.airport_id = traf.airport_id;

alter table v_airport_congestion change column airport_id airport_id int comment 'id аэропорта';
alter table v_airport_congestion change column city city string comment 'город аэропорта';
alter table v_airport_congestion change column state state string comment 'штат аэропорта';
alter table v_airport_congestion change column name name string comment 'название аэропорта';
alter table v_airport_congestion change column qty qty bigint comment 'количество перелетов аэропорта';

```

## Airports\_Fligths

```
create materialized view if not exists v_carrier_delay_stats
comment 'перевозчики которые чаще других задерживают вылеты и прилеты(средневзвешенно)'
stored as orc
tblproperties ("orc.compress"="SNAPPY")
select
  carrier,
  count(*) as total_flights,
  round((sum(case when dep_delay > 15 then 1 else 0 end) / count(*)) * 100, 2) as pct_delayed_15min,
  round(avg(case when dep_delay > 0 then dep_delay else 0 end), 2) as avg_dep_delay_weighted,
  round(avg(case when arr_delay > 0 then arr_delay else 0 end), 2) as avg_arr_delay_weighted
from flights
group by carrier;

alter table v_carrier_delay_stats change column carrier carrier string comment 'перевозчик';
alter table v_carrier_delay_stats change column total_flights total_flights string comment 'всего полетов';

alter table v_carrier_delay_stats change column pct_delayed_15min pct_delayed_15min double
comment 'процент рейсов, которые опоздали больше чем на 15 минут';

alter table v_carrier_delay_stats change column avg_dep_delay_weighted avg_dep_delay_weighted double
comment 'среднее время, на которое самолеты опаздывают с вылетом';

alter table v_carrier_delay_stats change column avg_arr_delay_weighted avg_arr_delay_weighted double
comment 'среднее время опоздания по прилету';
```

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Query executed successfully. Affected rows : -1

Airports\_Flighths

Took 1 min 21 sec. Last updated by maria\_dev at November 30 2025, 5:11:45 PM.

```
%jdbc(hive)

use airline_db;

--данные и описания метаданных

--внешние таблицы
select * from raw_flights limit 5;
select * from raw_airports limit 5;
--таблицы
select * from flights limit 5;
select * from airports limit 5;

--материализованные представления
select * from v_airports_flights limit 5;
select * from v_dayofweek_flights limit 5;
select * from v_carrier_delay_stats order by avg_dep_delay_weighted desc limit 5;
select * from v_airport_congestion order by qty desc limit 5;
select * from v_popularity_destination order by qty desc limit 5;

--описание объектов
describe formatted raw_flights;
describe formatted raw_airports;

describe formatted flights;
describe formatted airports;

describe formatted v_airport_congestion;
describe formatted v_airports_flights;
describe formatted v_carrier_delay_stats;
describe formatted v_dayofweek_flights;
describe formatted v_popularity_destination;
```

Query executed successfully. Affected rows : -1

settings ▲

Table Options

Name	Value
useFilter ⓘ	<input type="checkbox"/>
showPagination ⓘ	<input type="checkbox"/>
showAggregationFooter ⓘ	<input type="checkbox"/>

raw_flights.day_of_month	raw_flights.day_of_week	raw_flights.carrier	raw_flights.origin_airport_id
19	5	DL	11433
19	5	DL	14869
19	5	DL	14057
19	5	DL	15016
19	5	DL	11193

settings

raw_airports.airport_id	raw_airports.city	raw_airports.state	raw_airports.name		
10105	Adak Island	AK	Adak		
10299	Anchorage	AK	Ted Stevens Anchorage International		
10304	Aniak	AK	Aniak Airport		
10754	Barrow	AK	Wiley Post/Will Rogers Memorial		
10551	Bethel	AK	Bethel Airport		

settings

flights.day_of_month	flights.day_of_week	flights.origin_airport_id	flights.dest_airport_id	flight	
1	1	10821	14492	7	
1	1	10821	14492	7	
1	1	15304	14492	8	
1	2	10821	14492	7	
1	2	10821	11193	7	

settings

airports.airport_id	airports.city	airports.state	airports.name		
10141	Aberdeen	SD	Aberdeen Regional		
10158	Atlantic City	NJ	Atlantic City International		
10245	King Salmon	AK	King Salmon Airport		
10279	Amarillo	TX	Rick Husband Amarillo International		
10304	Aniak	AK	Aniak Airport		

settings

v_airports_flights.src	v_airports_flights.qty		
all quantity airports	365		
all quantity flights	2702218		

settings

v_dayofweek_flights.carrier	v_dayofweek_flights.mon_qty	v_dayofweek_flights.tue_qty	v_dayofweek_flights.wed_qty		
9E	12419	12225	12503		
AA	43168	42336	42618		

Airports\_Flighths

AS	10164	9954	9907
B6	18137	17461	17377
	57067	55348	56972

settings

v_carrier_delay_stats.carrier	v_carrier_delay_stats.total_flights	v_carrier_delay_stats.pct_delayed_15i	
MQ	113212	24.69	
EV	157928	22.9	
VX	34739	19.8	
B6	121906	21.98	
UA	286418	20.51	

settings

v_airport_congestion.airport_id	v_airport_congestion.city	v_airport_congestion.state	v_airport_c	
10397	Atlanta	GA	Hartsfield Internatio	
13930	Chicago	IL	Chicago (	
12892	Los Angeles	CA	Los Ange	
11298	Dallas/Fort Worth	TX	Dallas/Fo	
11292	Denver	CO	Denver In	

settings

v_popularity_destination.qty	v_popularity_destination.departure_id	v_popularity_destination.departure_i	
9643	14771	San Francisco	
9583	12892	Los Angeles	
6864	12892	Los Angeles	
6835	12889	Las Vegas	
6643	12478	New York	

settings

col_name	data_type	comment	
# col_name	data_type	comment	
day_of_month	int		
day_of_week	string		
carrier	string		
origin_airport_id	int		
dest_airport_id	int		
dep_delay	int		
arr_delay	int		

settings ▾

## Airports\_Fligths

col_name	data_type	comment	
# col_name	data_type	comment	
airport_id	int		
city	string		
state	string		
name	string		
	null	null	
# Detailed Table Information	null	null	
Database:	airline_db	null	

 settings ▾

col_name	data_type	comment	
# col_name	data_type	comment	
day_of_month	int	день месяца	
day_of_week	string	день недели	
origin_airport_id	int	аэропорт вылета	
dest_airport_id	int	аэропорт прилета	
dep_delay	int	задержка вылета	
arr_delay	int	задержка прибытия	
	null	null	

 settings ▾

col_name	data_type	comment	
# col_name	data_type	comment	
airport_id	int	id аэропорта	
city	string	город аэропорта	
state	string	штат аэропорта	
name	string	название аэропорта	
	null	null	
# Detailed Table Information	null	null	
Database:	airline_db	null	

 settings ▾

col_name	data_type	comment	
# col_name	data_type	comment	
airport_id	int	id аэропорта	
city	string	город аэропорта	
state	string	штат аэропорта	
name	string	название аэропорта	
qty	bigint	количество перелетов аэропорта	
	null	null	
# Detailed Table Information	null	null	

 settings ▾

col_name	data_type	comment	
# col_name	data_type	comment	
src	string		

Airports\_Flghths

qty	bigint	
	null	null
	null	null
Database:	airline_db	null
OwnerType:	USER	null
Owner:	hive	null

settings ▾

settings ▾

settings ▾