|  |
| --- |
| EPAM Systems, RD Dep. |
| USA Flight Performance |



Contents

[1. Business Description 3](#_Toc412572569)

[1.1. Business background 3](#_Toc412572570)

[1.2. Problems because of poor data management 3](#_Toc412572571)

[1.3. Benefits from implementing a Data Warehouse 3](#_Toc412572572)

[2. Dimensions of a Business 3](#_Toc412572573)

[3. Logical Scheme 3](#_Toc412572574)

[4. Data Flow 3](#_Toc412572575)

[5. Fact Table Partitioning Strategy 3](#_Toc412572576)

[6. Strategy of Parallel Load 3](#_Toc412572577)

[7. Report Layouts 3](#_Toc412572578)

# Business Description

## Business background

* + **Business background**

DWH will provide research department of the all airports and companies schedule with flight performance information (departures, arrivals, etc.).

* + **Problems because of poor data management**

Today it’s hard to know/analyze data about flight performance.

* + **Benefits from implementing a Data Warehouse**

As a result the data will have performance statistics that can be analyzed to improve companies and airports schedule and gain more passengers by static time flights.

# Dimensions of a Business

1. Airlines Dimension
2. Pilot Dimension
3. Plane Dimension
4. Location Dimension
5. Flight performance as fact table

# 3NF model

# 

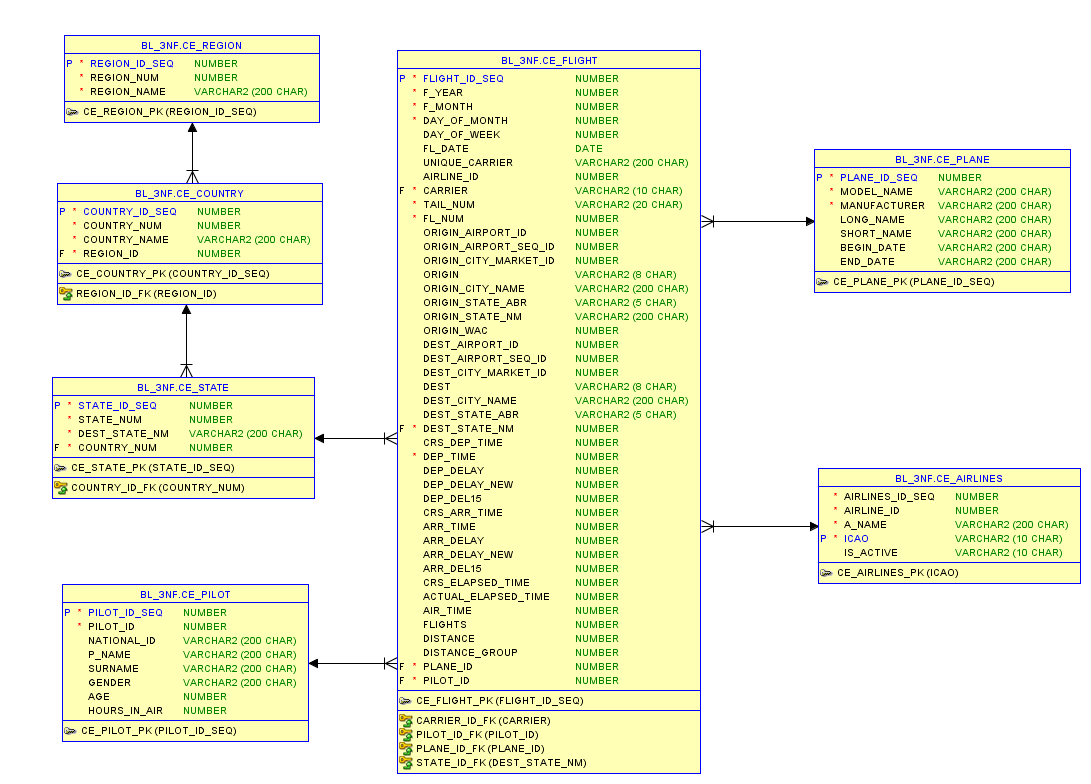


Figure 1 3NF schema

# Logical Scheme

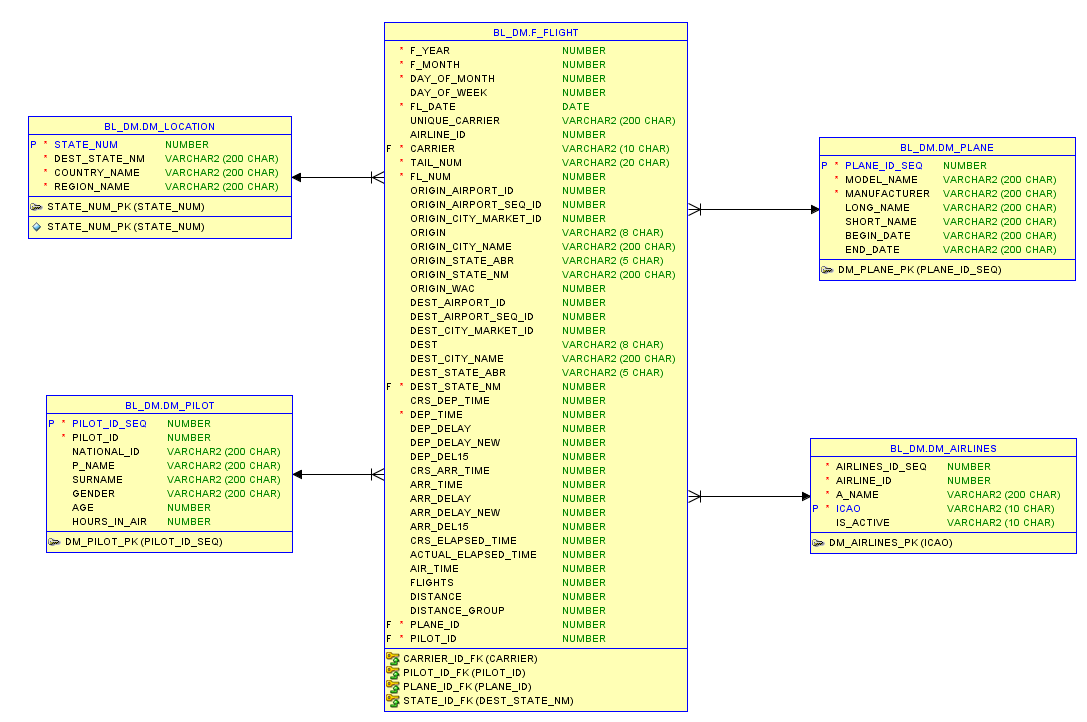
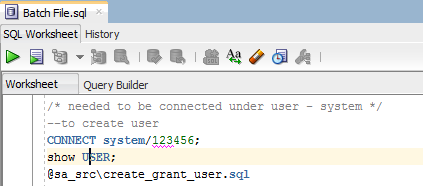


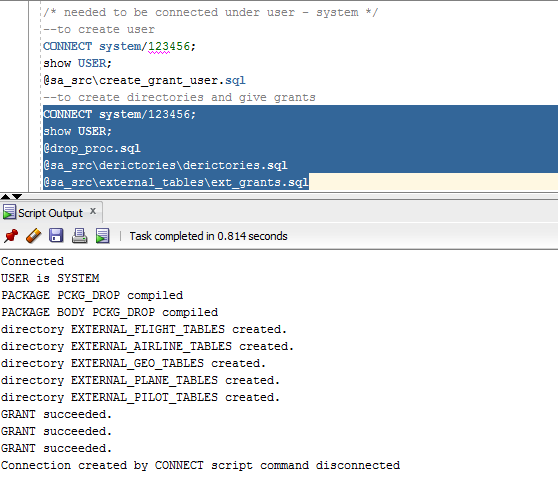
Figure 2 Dimension schema

# Data Flow

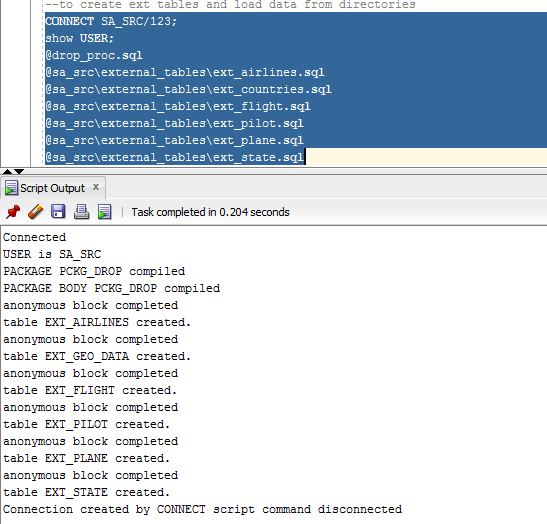
To load data some settings must be done:



Creating directories and giving grants:



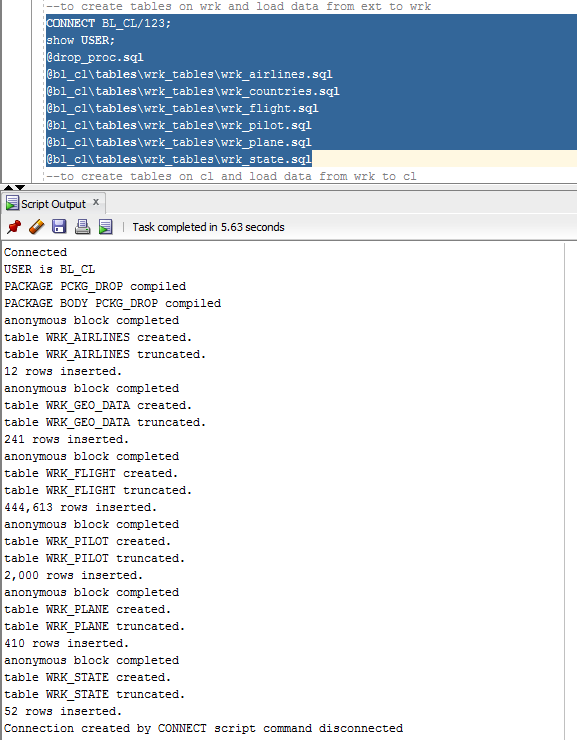
sa\_src layer load:



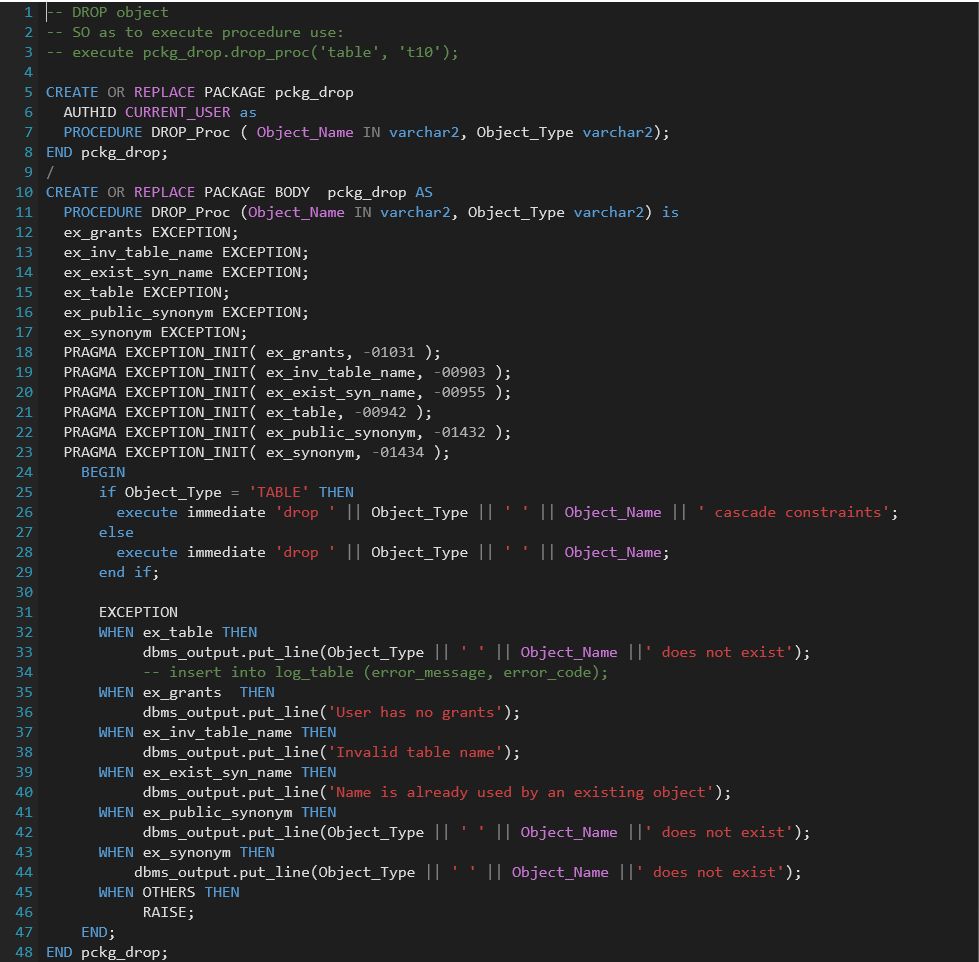
Script example of ext\_table:



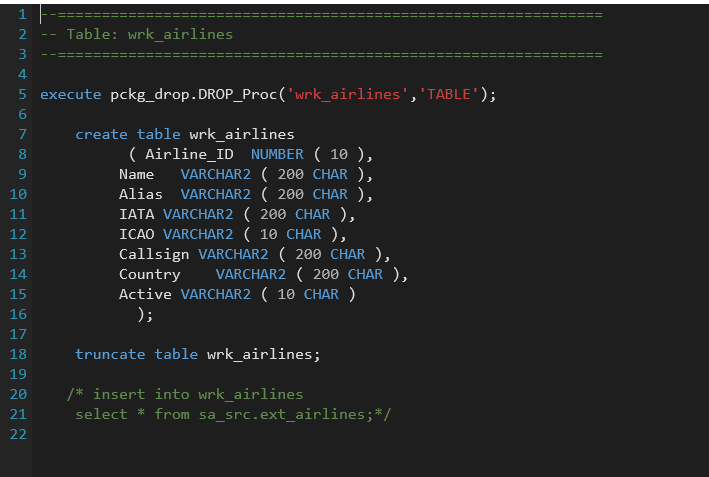
Bl\_cl layer – loading data to wrk tables:



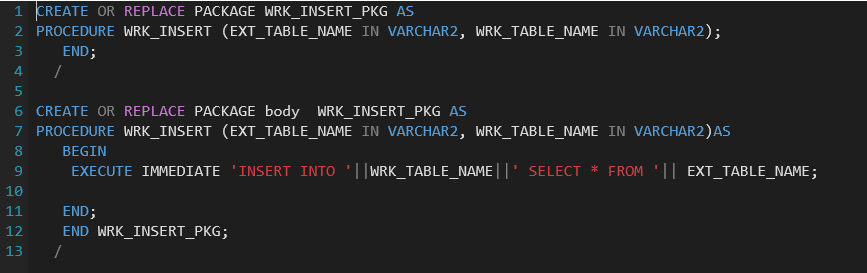
Script example of drop\_proc:



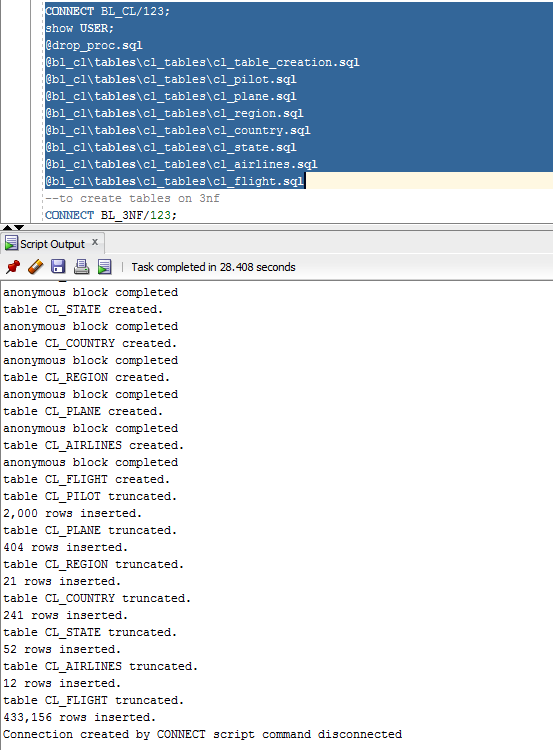
Script example of wrk\_table:



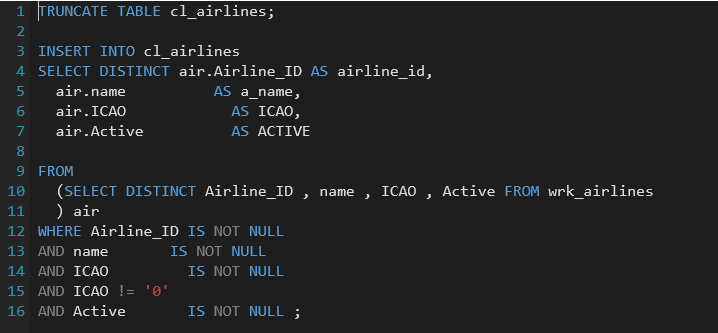
wrk\_insert\_pkg.sql was created:



To load everything to cleaning layer (cl\_tables):



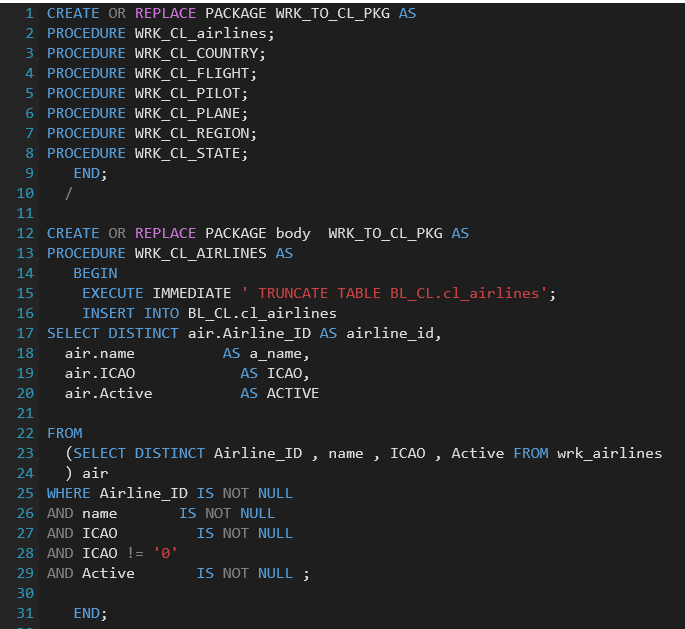
Script example pf cl\_load:



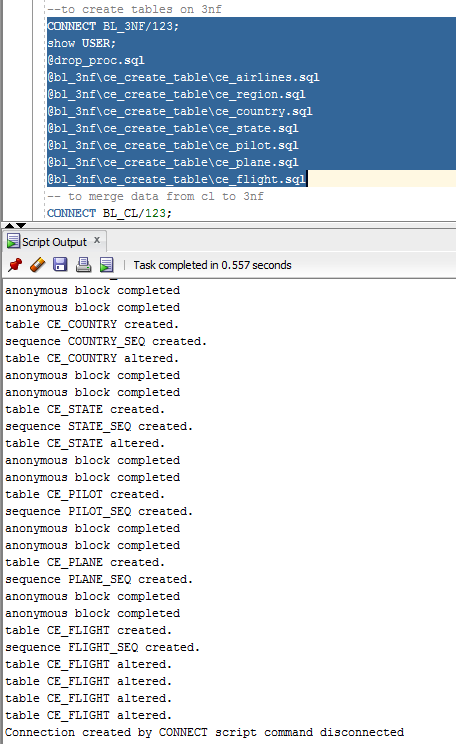
Script example of creating cl\_tables:



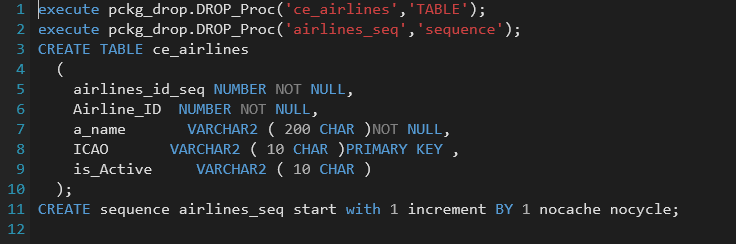
Another way to load is cl\_insert\_pkg.sql package:



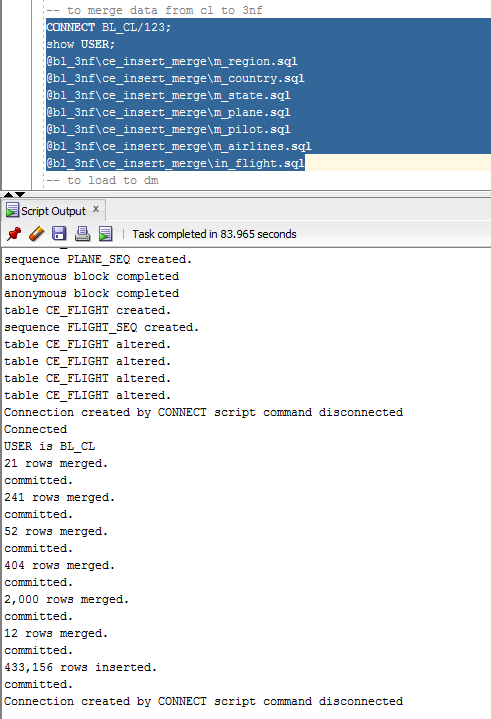
To create 3nf tables:



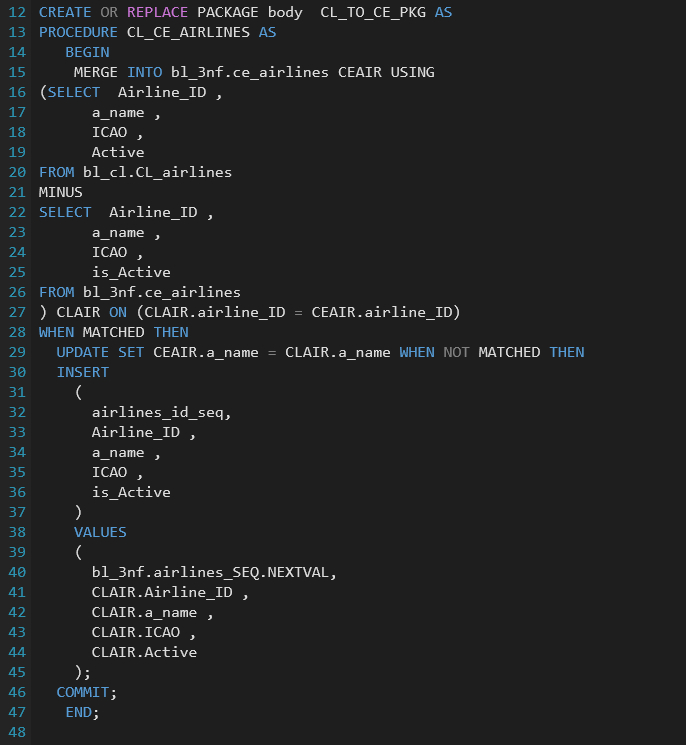
Script example of creating 3nf\_tables:



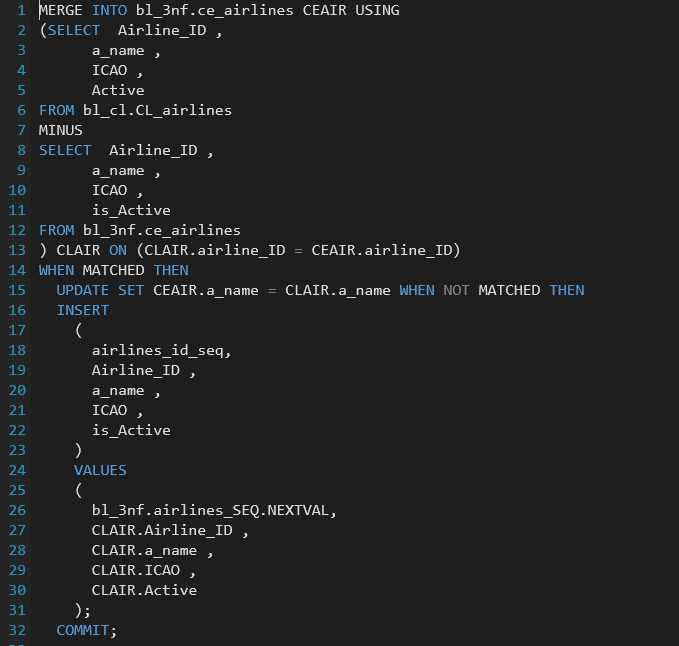
To merge data from cl tables to 3 nf tables:



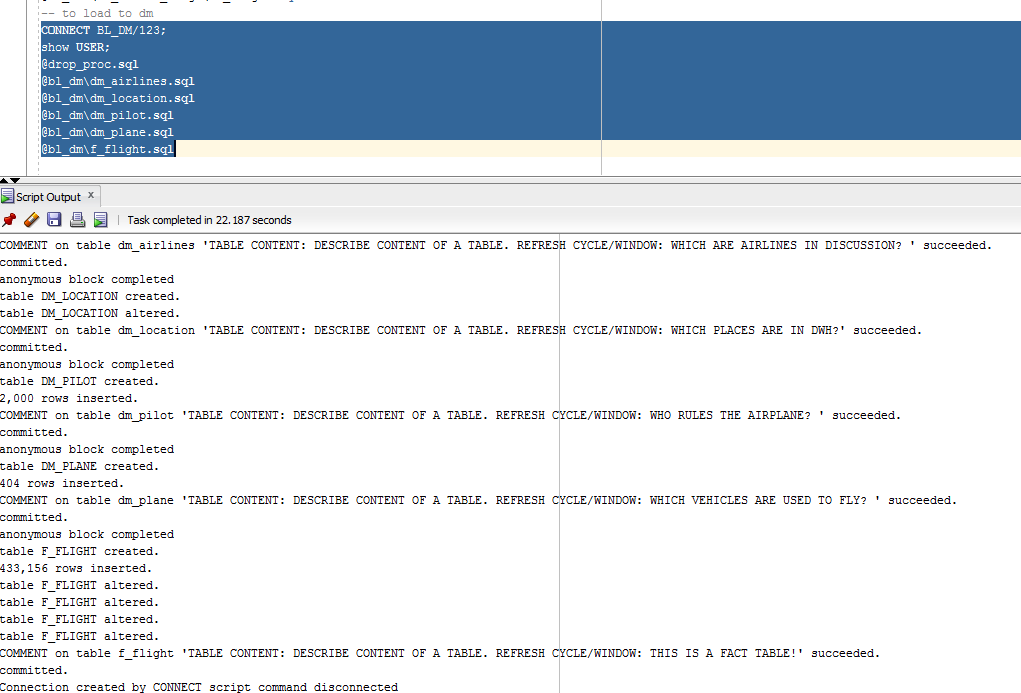
Script example of loading to 3nf with the help of cl\_to\_ce\_pkg.sql package:



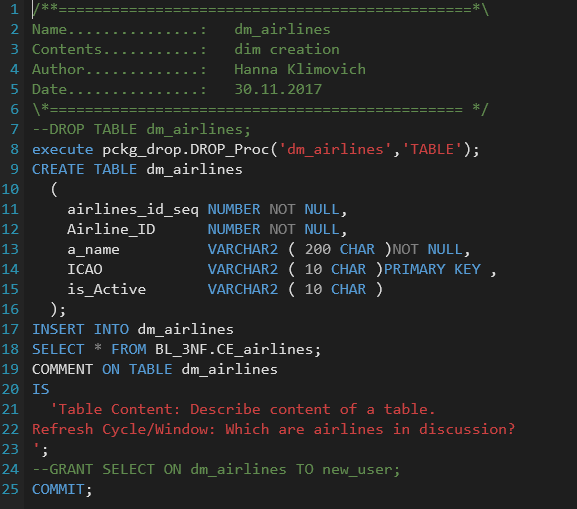
Another way to load is by each sql file:



To create tables and load data to dimensions:



Script example of creation and loading data:



Schemas of my bl\_cl layer:

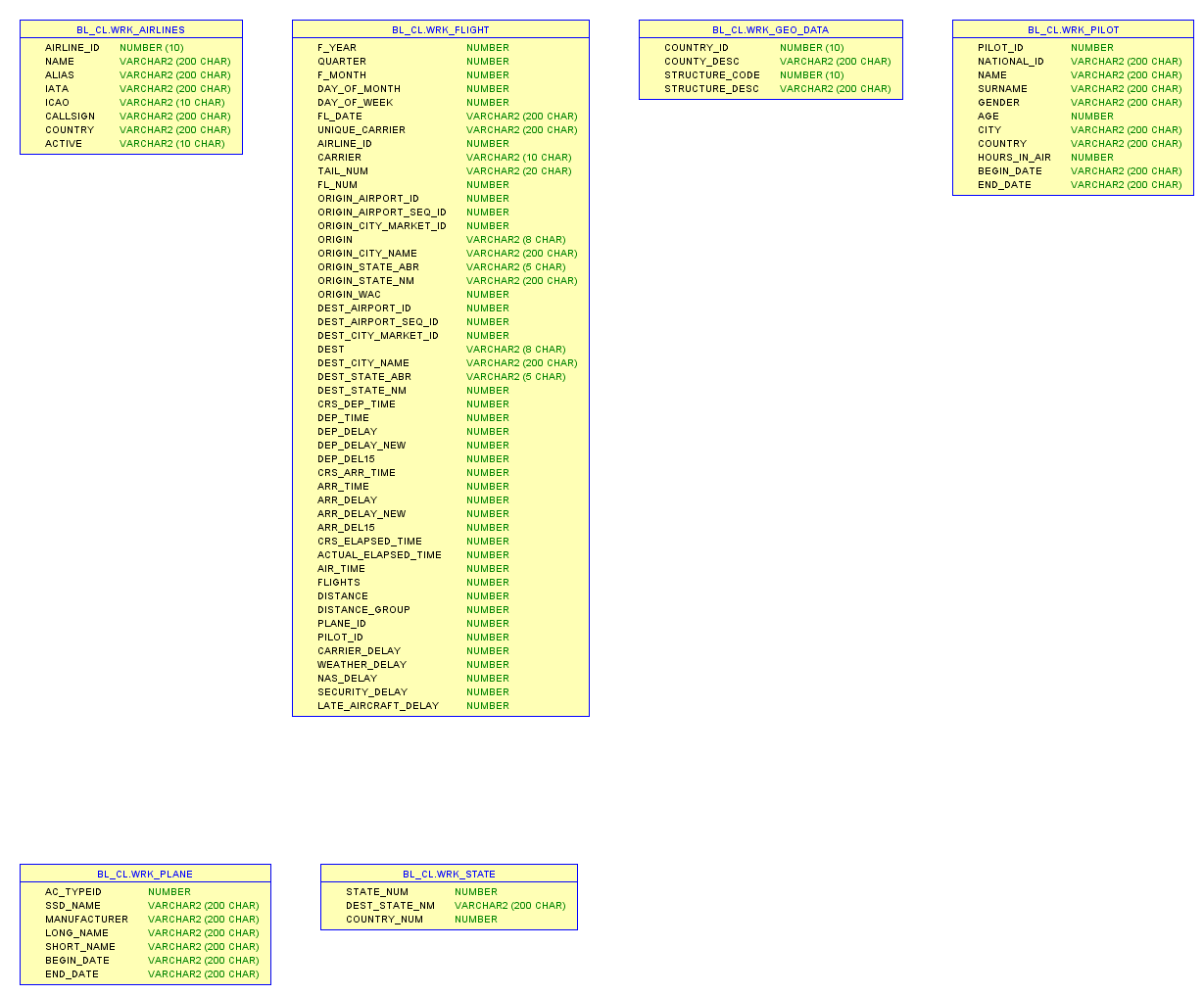


Figure 3 Sources tables

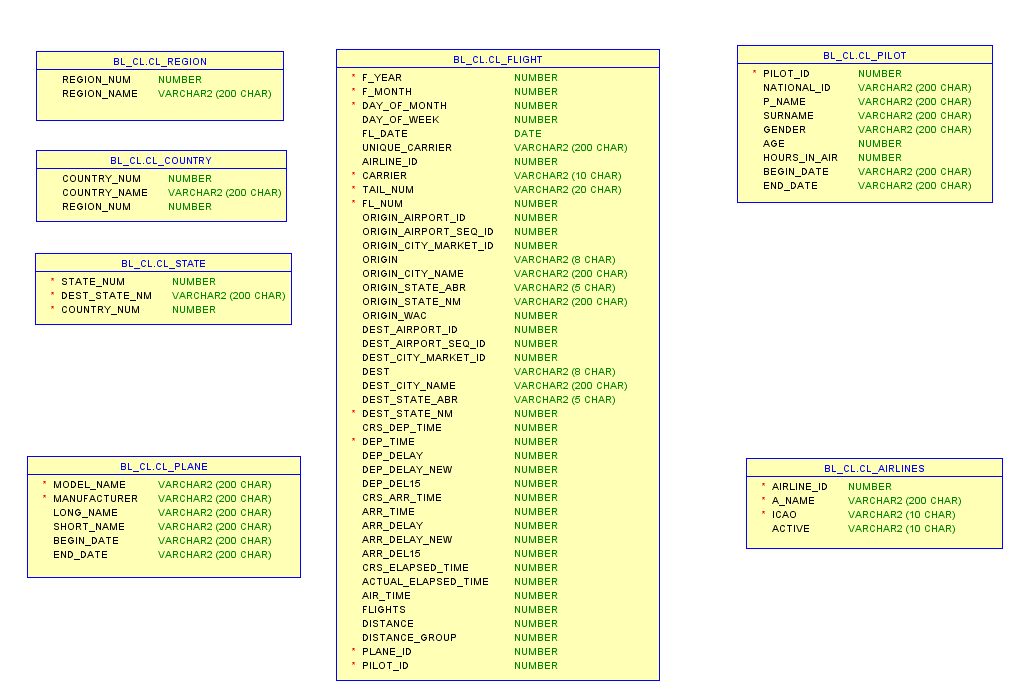


Figure 4 Cleansing tables

# Fact Table Partitioning Strategy

This fact table will be partitioned by airlines. In the future, when data will be stored in the time period of different months and years it’s better to partition by months.

# Strategy of Parallel Load

It’s better to load all dimensional in parallel with fact table.

# Report Layouts

# Examples are stored down below.

# 

Figure 5 Report example 1

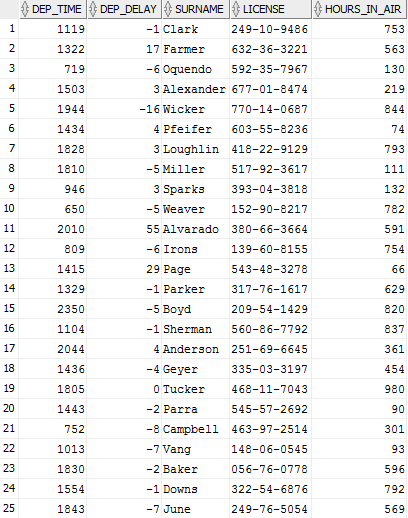


Figure 6 Report example 2

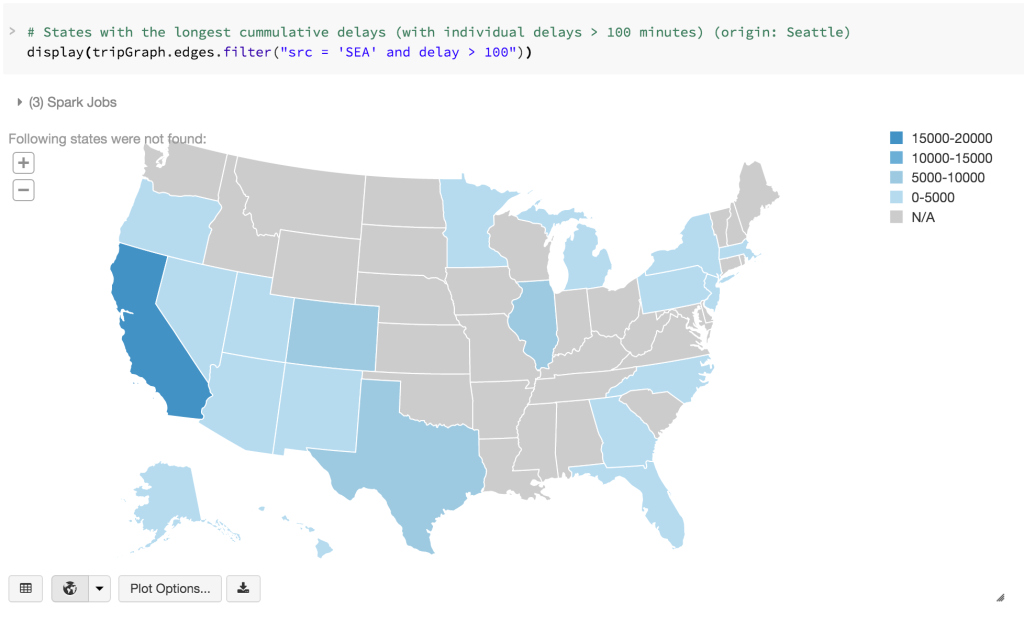


Figure 7 Report example 3

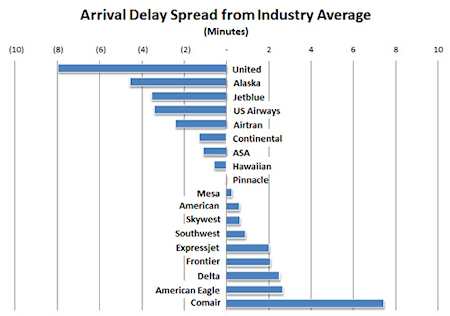


Figure 8 Report example 4

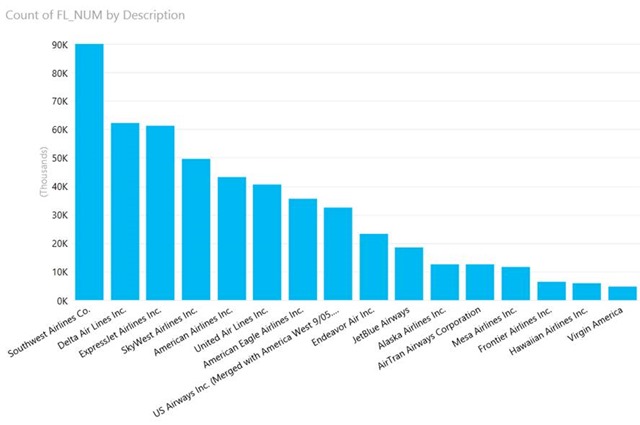


Figure 9 Report example 5