**Documentation**

[**1.** **Business Description** 2](#_Toc498560228)

[**2.** **Dimensional layer of your business process** 3](#_Toc498560229)

[**3.** **Logical Scheme** 7](#_Toc498560230)

[**4.** **Data flow** 8](#_Toc498560231)

[**5.** **Fact Table Partitioning Strategy** 9](#_Toc498560232)

[**6.** **Report Layouts** 10](#_Toc498560233)

# **Business Description**

**Business background**

My chosen business is airline ticket sales service. The main goal of this business is selling of tickets and receiving the profit.

Key business processes are looking for a suitable ticket for client, making reservation of seats in the airplane and getting payment.

**Problems because of poor data management**

The problems of poor data management causes the impossibility of an in-depth analysis of the demand for air tickets, profit analysis, having full data about clients and their orders.

**Benefits from implementing a Data Warehouse**

Data Warehouse will let business to store all data about ticket sales, aggregate it, do visualization, analyze data and make predictions for the future.

# **Dimensional layer of your business process**

Процесс дизайна измерений:

Шаг 1. Выбор бизнес-процесса:

Для постраения хранилища данных был выбран процесс продажи авиабилетов.

Шаг 2. Определение зерна

Зерном будет являться продажа одного билета.

Шаг 3. Определение измерений:

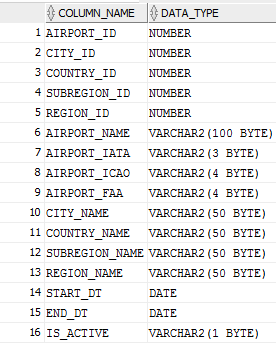
Измерения могут быть следующими - дата покупки билета, пассажиры, аэропорты вылета, аэропорты прибытия, рейс, уровень сервиса.

Шаг 4. Определение фактов:

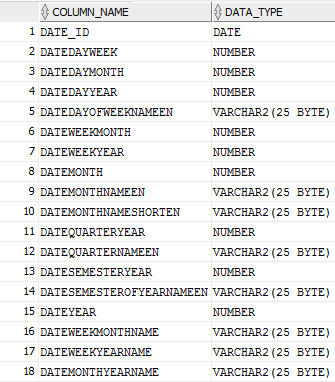
Фактом будет являться – стоимость билета.

**2.1 Измерения**

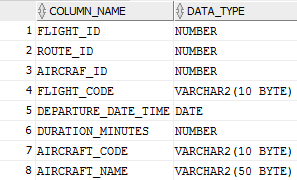
**Dim\_airport**



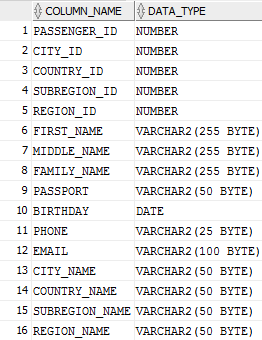
**Dim\_date**



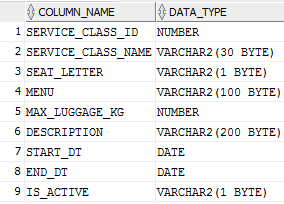
**Dim\_flight**



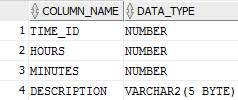
**Dim\_passenger**



**Dim\_service\_classes**

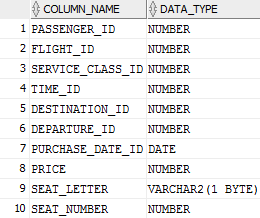


**Dim\_time**

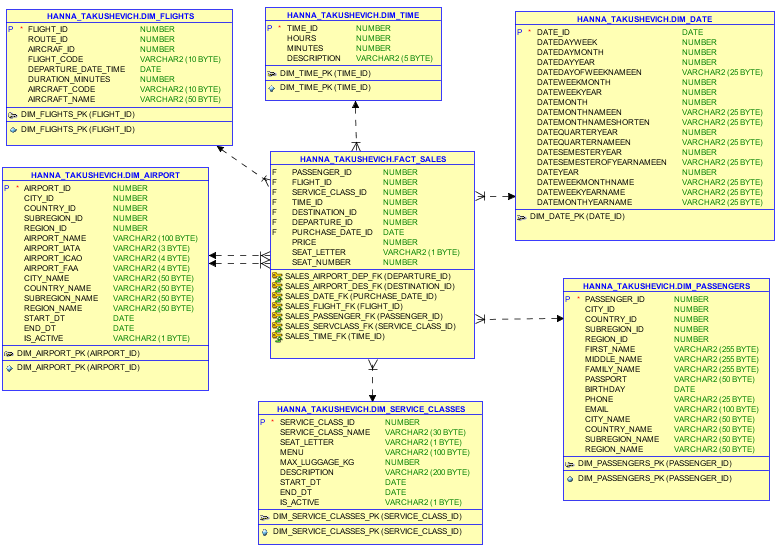


**2.2 Фактовая таблица**

**Fact\_sales**

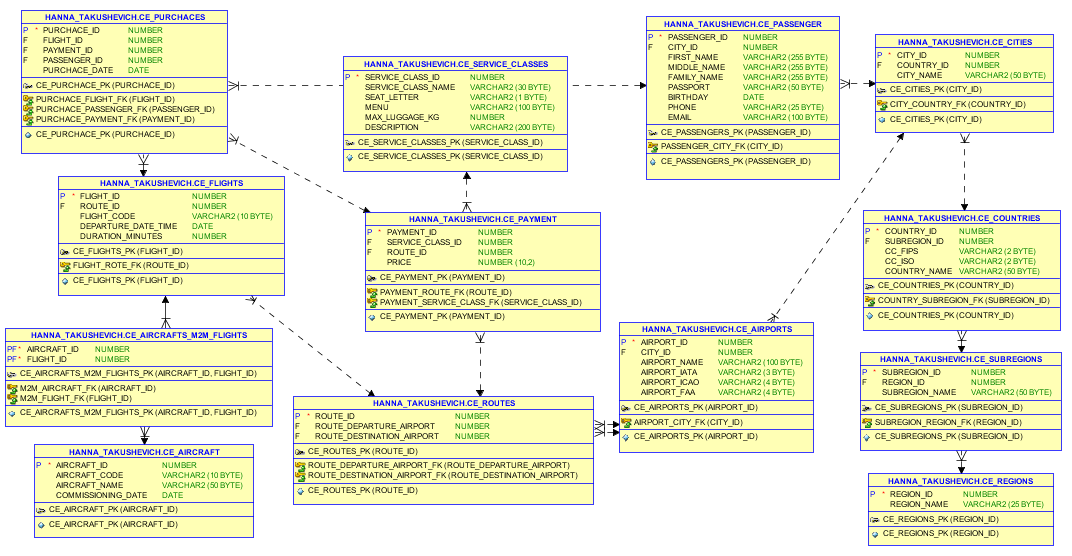


**2.3 Звезда**



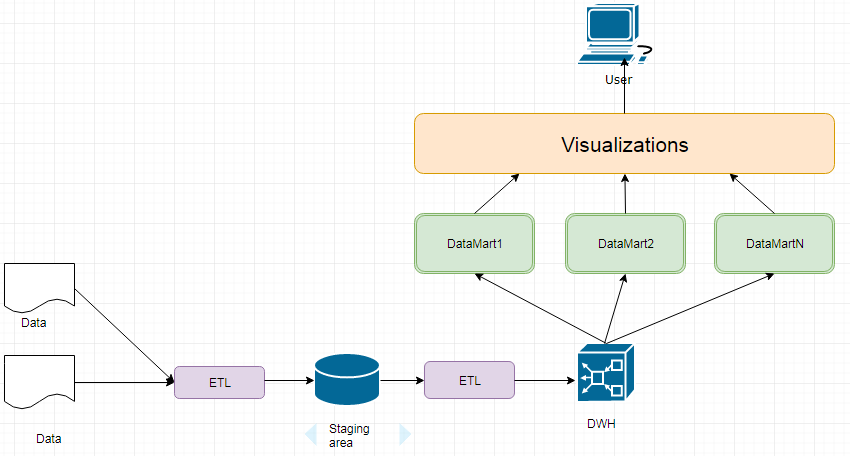
# **Logical Scheme**

Логическая схема имеет следующий вид:



В схеме 13 таблиц. Основной таблицей является таблица ce\_purchaces. Она содержит информацию о рейсе, стоимости билета, пассажире и дате покупки билета. Остальные таблицы содержат более подробные данные, такие как данные о пассажирах, аэропортах, самолётах, полётах, маршрутах, классах обслуживания и географических данных.

# **Data flow**



Изначально сырые данные хранятся в источниках различных форматов. С помощью ETL эти данные подгружаются в базу данных, чтобы далее при помощи другого ETL эти данные можно было загрузить в наше хранилище данных. На основе хранилища будут построены витрины данных, которые в последствие будут использоваться для визуализации.

# **Fact Table Partitioning Strategy**

Для анализа данных о покупках авиабилетов чаще всего будут необходимы данные, полученные не позднее десяти лет на момент проведения анализа. Чтобы максимально эффективно извлекать необходимые данные по датам, будут созданые партиции на аттрибуте «Дата покупки билета». Партиции будут содержать в себе данные для таких временных промежутков, как (А) данные полученные более чем десять лет назад, (Б) данные полученные от пяти до десяти лет назад и (В) данные полученные не позднее пяти лет назад. Вид партиционирования – range.

# **Report Layouts**

Я хочу проанализировать спрос клиентов на определённые маршруты, классы обслуживания и динамику общего интерес к авиакомпанни.

Для построения репортов были использованы следующие фиктивные данные:

|  |  |
| --- | --- |
| Год | Количество проданных билетов |
| 2008 | 856 |
| 2009 | 986 |
| 2010 | 945 |
| 2011 | 1387 |
| 2012 | 1904 |
| 2013 | 2504 |
| 2014 | 2100 |
| 2015 | 2732 |
| 2016 | 2726 |
| 2017 | 3123 |

|  |  |
| --- | --- |
| Класс обслуживания | Процент |
| Бизнес-класс | 20.78% |
| Эконом-класс | 61.80% |
| Вип-класс | 5.60% |
| Другие | 11.82% |

|  |  |
| --- | --- |
| Маршрут | Количество купленных билетов |
| Минск-Москва | 356 |
| Варшава - Питер | 123 |
| Киев - Рим | 341 |
| Москва - Токио | 589 |
| Минск - Пекин | 12 |

Дашборд:

