

|  |
| --- |
| Анализ продаж  Анализ продаж авиабилетов |
|  |

Contents

[1 Описание предметной области 3](#_Toc500070872)

[2 Дизайн измерений 4](#_Toc500070873)

[2.1 Измерение dim\_date 4](#_Toc500070874)

[2.2 Измерение dim\_airports 5](#_Toc500070875)

[2.3 Измерение dim\_flights 5](#_Toc500070876)

[2.4 Измерение dim\_airlines 6](#_Toc500070877)

[2.5 Измерение dim\_passengers 6](#_Toc500070878)

[2.6 Измерение dim\_service\_classes\_scd 7](#_Toc500070879)

[2.7 фактовая таблица fact\_purchaces 7](#_Toc500070880)

[2.8 Модель звезда 8](#_Toc500070881)

[3 ЛОгическая схема (3нф) 9](#_Toc500070882)

[4 Поток данных 10](#_Toc500070883)

[5 Партиционирование фактовой таблицы 11](#_Toc500070884)

[6 Параллельная загрузка данных 12](#_Toc500070885)

[7 Макеты отчётов 13](#_Toc500070886)

# Описание предметной области

**Описание:**

Выбранной предметной областью является сервис онлайн продаж авиабилетов. Основная цель бизнеса – продажа билетов и получение прибыли.

К ключевым процессам относятся: поиск подходящего билета для клиента, осуществление бронирования мест и получениео оплаты.

**Проблемы:**

Проблемы, вызванные скудным управлением данными – невозможность глубокого анализа спроса на авиабилеты и анализа динамики изменения прибыли и продаж.

**Плюсы создания хранилица данных:**

Преимущества, которые даёт хранилище данных – хранение данных о продажах, предоставление возможностей для анализа данных и составления прогнозов на будущее.

# Дизайн измерений

Для построения хранилища данных были пройдены четыре шага дизайна.

Шаг 1. Выбор бизнес-процесса:

Для постраения хранилища данных был выбран процесс продажи авиабилетов.

Шаг 2. Определение зерна

Зерном будет являться продажа одного билета.

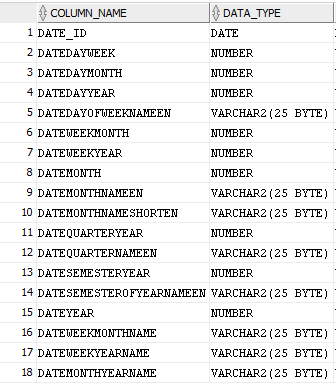
Шаг 3. Определение измерений:

Измерения - дата покупки билета, пассажир, аэропорт вылета, аэропорт прибытия, рейс, уровень сервиса, авиалиния.

Шаг 4. Определение фактов:

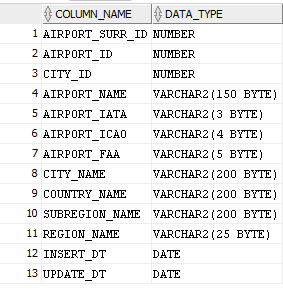
Фактом будет являться – стоимость билета.

## Измерение dim\_date



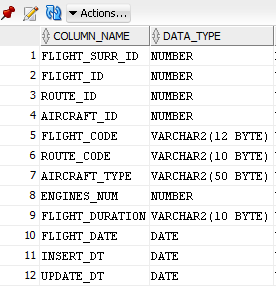
Измерение dim\_date содержит информацию о датах, начиная с 2014 года и по 2018. Данное измерение удобно тем, что хранит информацию не только в формате даты, но также содержит дополнительную информацию о неделях, месяцах, кварталах, семестрах и годах.

## Измерение dim\_airports



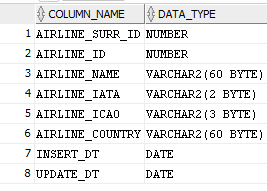
В измерении dim\_airports хранится полная информация об аэропортах: название, география местоположения, международные коды.

## Измерение dim\_flights



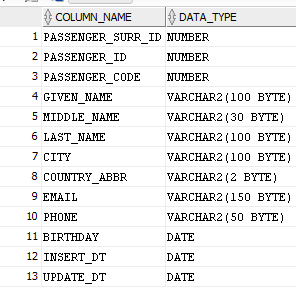
Измерение dim\_flights хранит данные о рейсах. В нём можно найти данные о дате рейса, самолёте осуществляющем рейс, продолжительность полёта.

## Измерение dim\_airlines



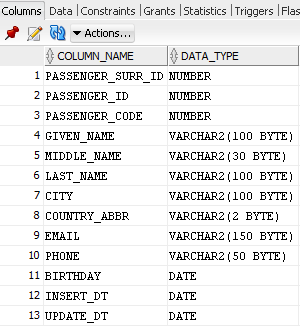
Измерение dim\_airlines содержит такую информацию об авиалиния, как название, международные коды и страну.

## Измерение dim\_passengers



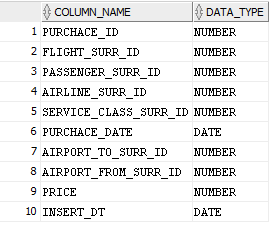
Измерение хранит личные данные людей, покупавших авиабилеты.

## Измерение dim\_service\_classes\_scd



Данное измерение является SCD2 измереним. В нём хранится информация о классах облуживания.

## фактовая таблица fact\_purchaces



Фактовая таблица содержит ссылки на измерения, а также цену покупки авиабилета.

## Модель звезда

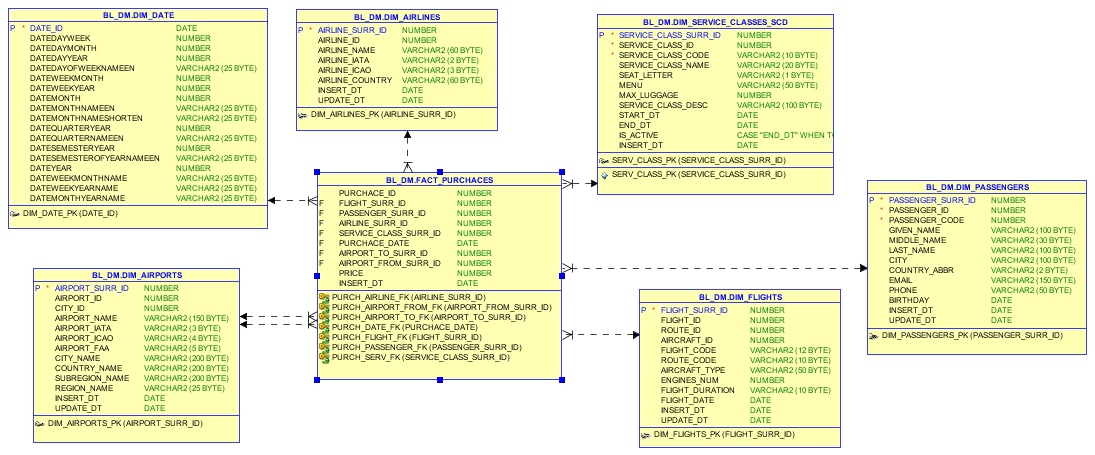
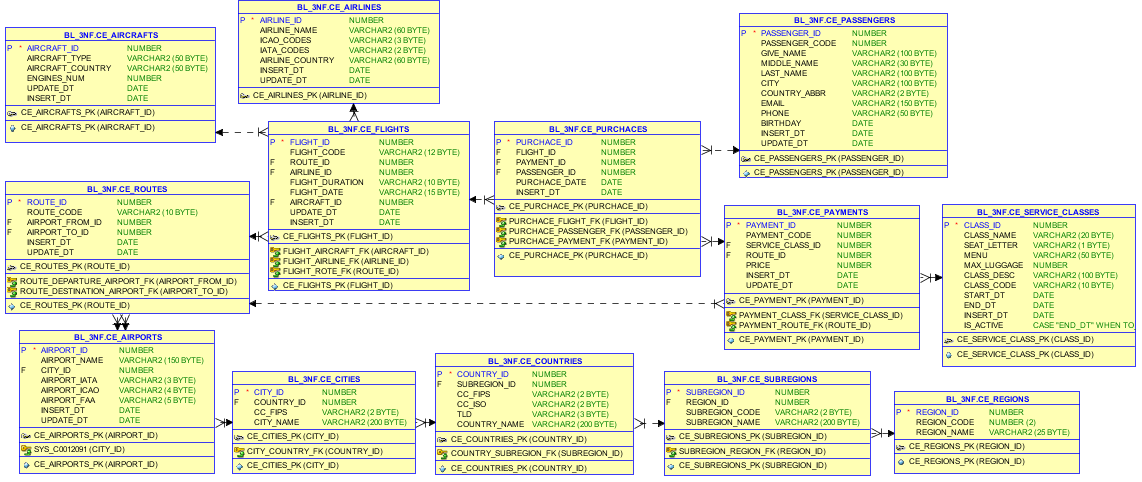


Схема звезда была выбрана по следующим причинам:

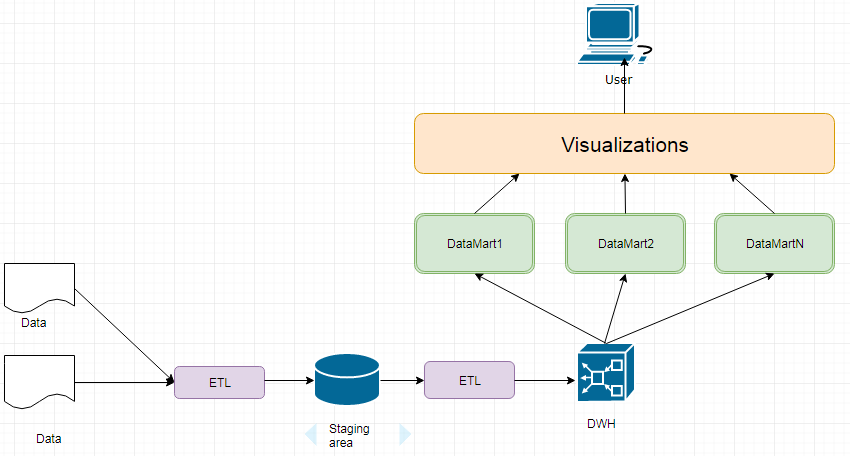
* Простая структура.
* Высокая производительность за счёт отсутствия множественных связей между таблицами.

# ЛОгическая схема (3нф)



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.

# Поток данных



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.

# Партиционирование фактовой таблицы

Для анализа данных о покупках авиабилетов чаще всего будут необходимы данные для конкретного временного промежутка. Чтобы максимально эффективно извлекать необходимые данные по датам, будут созданые партиции на аттрибуте «Дата покупки билета». Вид партиционирования – range.

CREATE TABLE fact\_purchaces

(

...

)PARTITION BY RANGE (purchace\_date)

(PARTITION t1p1 VALUES LESS THAN (TO\_DATE('01-jan-2015', 'dd-mon-yy')),

PARTITION t1p2 VALUES LESS THAN (TO\_DATE('01-jan-2016', 'dd-mon-yy')),

PARTITION t1p3 VALUES LESS THAN (TO\_DATE('01-jan-2017', 'dd-mon-yy')),

PARTITION t1p4 VALUES LESS THAN (TO\_DATE('01-jan-2018', 'dd-mon-yy'))

);

# Параллельная загрузка данных

При вставке данных в фактовую таблицу используется селект с большим значение cost. Для ускорения выборки было принято решение использовать хинт /\*+ parallel(cfp 4) \*/

MERGE INTO fact\_purchaces r USING

( SELECT /\*+ parallel(cfp 4) \*/ purchace\_id,

(select flight\_surr\_id from dim\_flights where flight\_id=cfp.flight\_id) as flight\_surr\_id,

(select passenger\_surr\_id from dim\_passengers where passenger\_id=cfp.passenger\_id) as passenger\_surr\_id,

(select airline\_surr\_id from dim\_airlines where airline\_id=cfp.airline\_id) airline\_surr\_id,

(select service\_class\_surr\_id from dim\_service\_classes\_scd where service\_class\_id=cfp.service\_class\_id) as service\_class\_surr\_id, purchace\_date,

(select airport\_surr\_id from dim\_airports where airport\_id=cfp.airport\_to\_id) as airport\_to\_surr\_id,

(select airport\_surr\_id from dim\_airports where airport\_id=cfp.airport\_from\_id) as airport\_from\_surr\_id, price FROM cls\_fact\_purchaces cfp

MINUS

SELECT /\*+ parallel( fp 4) \*/ purchace\_id, flight\_surr\_id, passenger\_surr\_id, airline\_surr\_id, service\_class\_surr\_id, purchace\_date, airport\_to\_surr\_id, airport\_from\_surr\_id,price FROM fact\_purchaces fp

) cls ON ( cls.purchace\_id = r.purchace\_id )

WHEN MATCHED THEN ...

# Макеты отчётов

Предположим, что нам необходимо проанализировать спрос клиентов на определённые маршруты, классы обслуживания и динамику общего интерес к авиакомпанни. Тогда можно построить дашборд следующего вида:

