**Практика**

Тема: Сервис потоковой передачи музыки

1. Бизнес-процесс: Аналитика потребляемого контента и взаимодействия с ним пользователя.

Цели:

* Оценивать популярность песен
* Составлять актуальные рекомендации
* Оценивать вовлеченность пользователей

1. Grain: Один трек соответствует одному прослушиванию одним пользователем.

Одна строка факта = прослушивание одного трека одним пользователем

|  |
| --- |
| dim\_songs |
| PK: song\_sk |
| source\_song\_id |
| song\_name |
| FK: artist\_sk |
| album\_name |
| duration |
| FK: genre\_sk |

|  |
| --- |
| dim\_date |
| PK: time\_sk |
| source\_time\_id |
| date\_added\_song |
| date\_added\_on\_playlist |
| time\_of\_day |
| time\_of\_week |
| time\_of\_mounth |

1. Таблицы измерений

|  |
| --- |
| dim\_users |
| PK: user\_sk |
| source\_user\_id |
| user\_name |
| user\_age |
| FK: region\_sk |
| FK: subscription\_sk |

|  |
| --- |
| dim\_subscription |
| PK: subscription\_sk |
| source\_subscription\_id |
| subscription\_type |

|  |
| --- |
| dim\_region |
| PK: region\_sk |
| source\_region\_id |
| region\_name |

|  |
| --- |
| dim\_genre |
| PK: genre\_sk |
| source\_genre\_id |
| genre\_name |

|  |
| --- |
| dim\_artist |
| PK: artist\_sk |
| source\_artist\_id |
| artist\_name |
| FK: region\_sk |
| FK: genre\_sk |

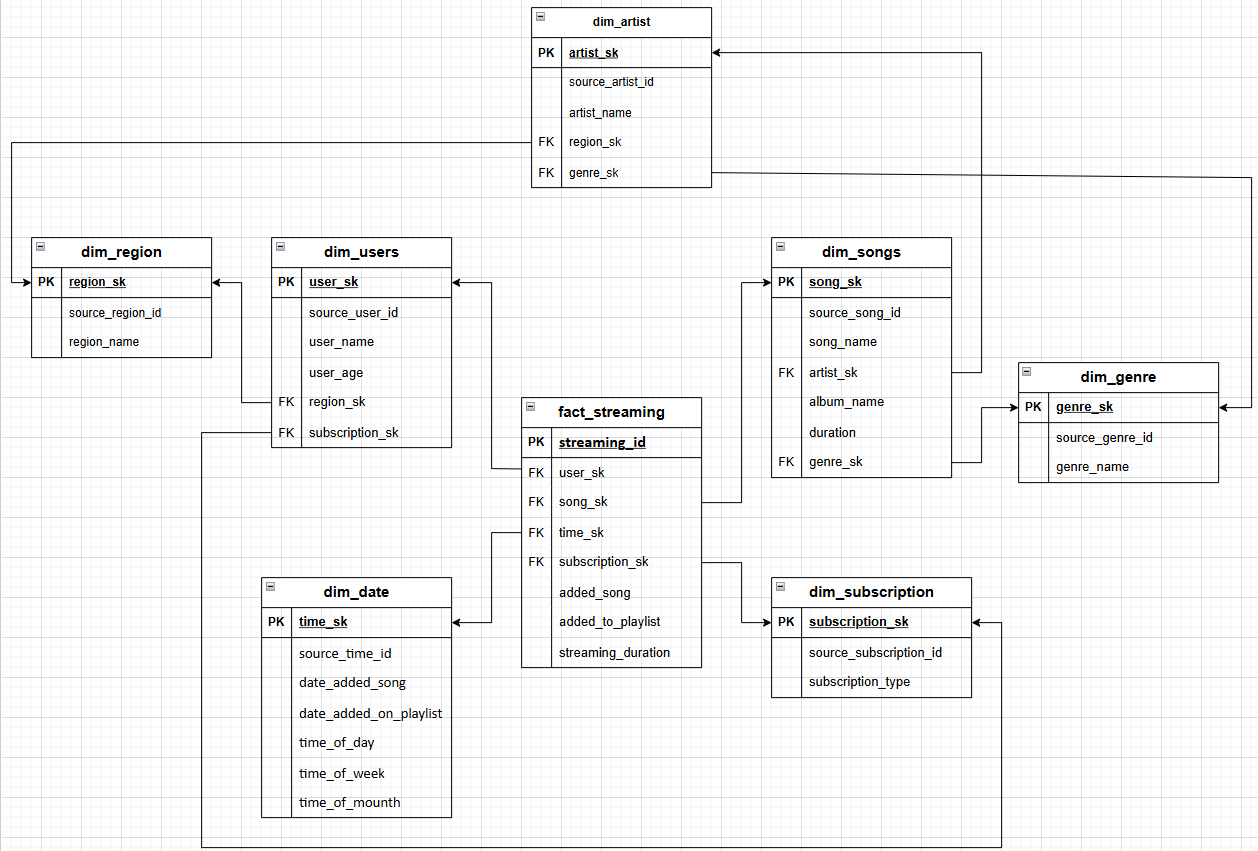
1. Таблица фактов

|  |
| --- |
| fact\_streaming |
| PK: streaming\_id |
| FK: user\_sk |
| FK: song\_sk |
| FK: time\_sk |
| FK: subscription\_sk |
| added\_song |
| added\_to\_playlist |
| streaming\_duration |

Метрики:

* added\_song – добавление песни
* added\_to\_playlist – добавление в плейлист
* streaming\_duration – продолжительность прослушивания

1. Физическая модель: Снежинка. Данная модель лучше всего подходит для OLAP-системы



1. Аналитические запросы

--Средняя длительность прослушивания по жанрам

SELECT g.genre\_name, ROUND(AVG(f.streaming\_duration), 2) AS avg\_duration

FROM fact\_streaming f

JOIN dim\_songs s ON f.song\_sk = s.song\_sk

JOIN dim\_genre g ON s.genre\_sk = g.genre\_sk

GROUP BY g.genre\_name

ORDER BY avg\_duration DESC;

--Сравнение активность пользователей с подпиской и без неё

SELECT s.subscription\_type, COUNT(\*) AS total\_streams

FROM fact\_streaming f

JOIN dim\_subscription s ON f.subscription\_sk = s.subscription\_sk

GROUP BY s.subscription\_type

ORDER BY total\_streams DESC;

--Топ 5 песен по прослушиваниям

SELECT s.song\_name, COUNT(\*) AS play\_count

FROM fact\_streaming f

JOIN dim\_songs s ON f.song\_sk = s.song\_sk

GROUP BY s.song\_name

ORDER BY play\_count DESC

LIMIT 5;