1. Бизнес‑процесс

Аналитика прослушиваний треков: сбор событий «прослушивание трека» (stream events) от пользователей/устройств с целью анализа вовлечённости, популярности треков/исполнителей/плейлистов, удержания и монетизации (рекламные/подписочные метрики).

1. Уровень детализации (grain)

Grain факта: одна запись = одно завершённое прослушивание трека одним пользователем (stream\_event). Если слушание начинается и длится < X секунд, может считаться «skip» — это атрибут. Полная запись содержит время начала, длительность воспроизведения, флаг завершено/прервано.

1. Таблицы измерений

* dim\_time — календарная/временная размерность (дата, час, день недели, квартал и т.д.).
* dim\_user — информация о пользователе (id, регион, страна, возрастная группа, подписка\_type и signup\_date).
* dim\_track — метаданные трека (track\_id, title, duration\_seconds, album\_id, artist\_id, genre, release\_date).

1. Таблица фактов

fact\_stream\_event (star-схема вокруг фактов прослушиваний).

1. **Суррогатные ключи (SK)**:
   * playback\_sk — первичный ключ.
   * date\_sk — ссылка на измерение даты (заменяет event\_time и time\_id).
   * user\_sk, track\_sk, artist\_sk, device\_sk, playlist\_sk — ссылки на соответствующие измерения.
2. **Метрики**:
   * played\_seconds — длительность воспроизведения.
   * skip\_flag и completed\_flag — флаги пропуска и завершения трека.
   * revenue\_amount — доход от события.
3. **Денормализованные поля**:
   * track\_duration\_seconds, quality, country — скопированы из измерений для упрощения запросов.
4. **Вычисляемые поля**:
   * percent\_played — процент прослушанного трека (генерируется автоматически).
5. Модель: Star schema (факт в центре и dimension tables вокруг). Snowflake не нужен для учебной задачи; отдельные таблицы artist/album могут быть нормализованы, но оставим как отдельные dims для читаемости.
6. Диаграмма (ER) — ниже.

