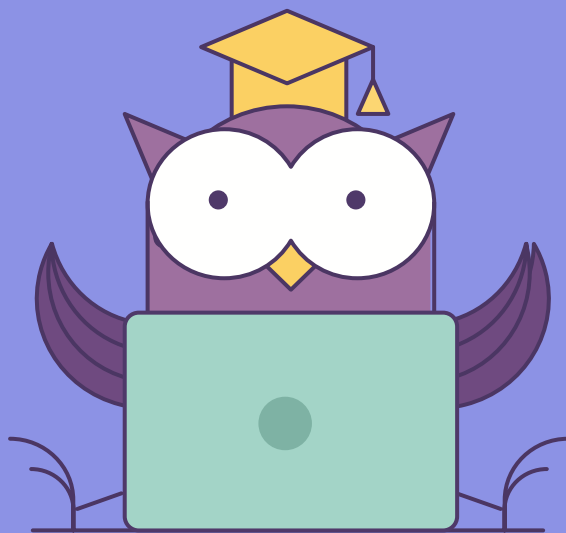




ОНЛАЙН-ОБРАЗОВАНИЕ

# Меня хорошо слышно && видно?



Напишите в чат, если есть проблемы!

Ставьте  если все хорошо

# Инструменты выгрузки данных из сторонних систем

Часть 2 - RDBMS & CDC



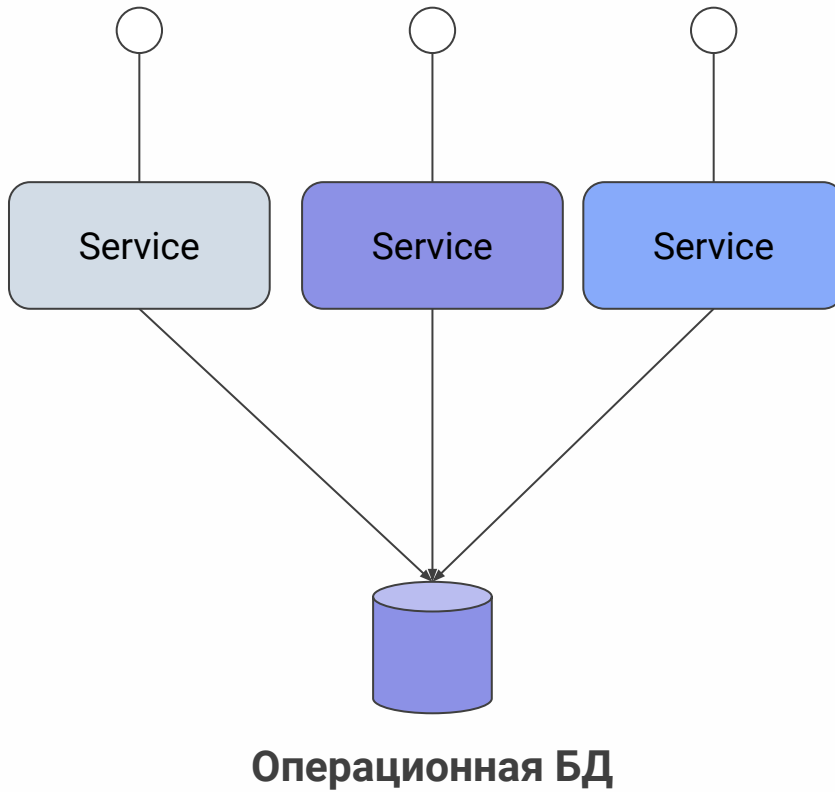
# 01

## Выгрузка из реляционных БД

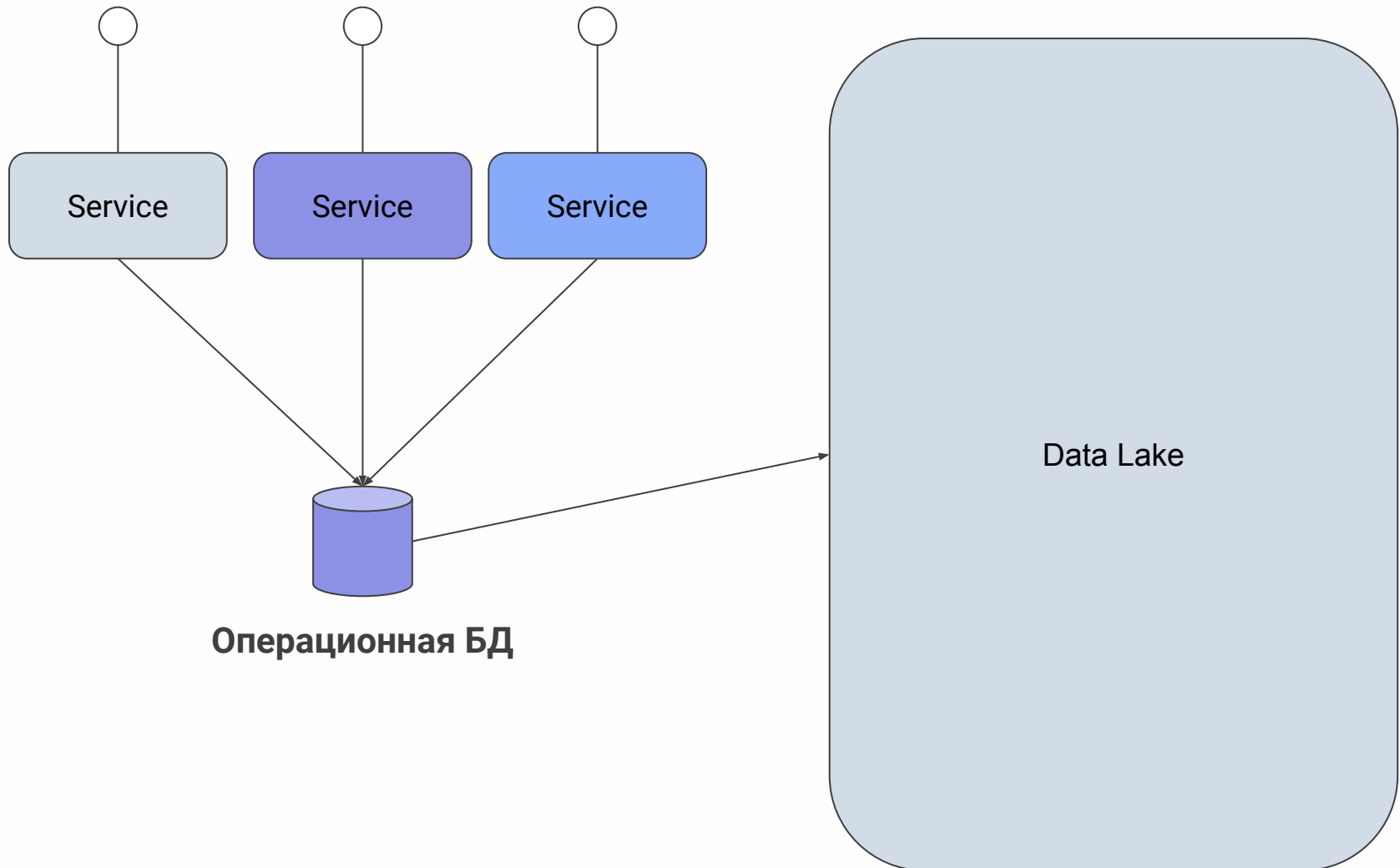
ОСНОВЫ

- Выгрузка из реляционных баз
- Пассивная выгрузка (Snapshot)
- Активная выгрузка (CDC)
- Практический пример

Сервисы активно пишут/читают из операционной БД



BI/ML/DS пользователи хотят анализировать данные - как нам их отдавать?

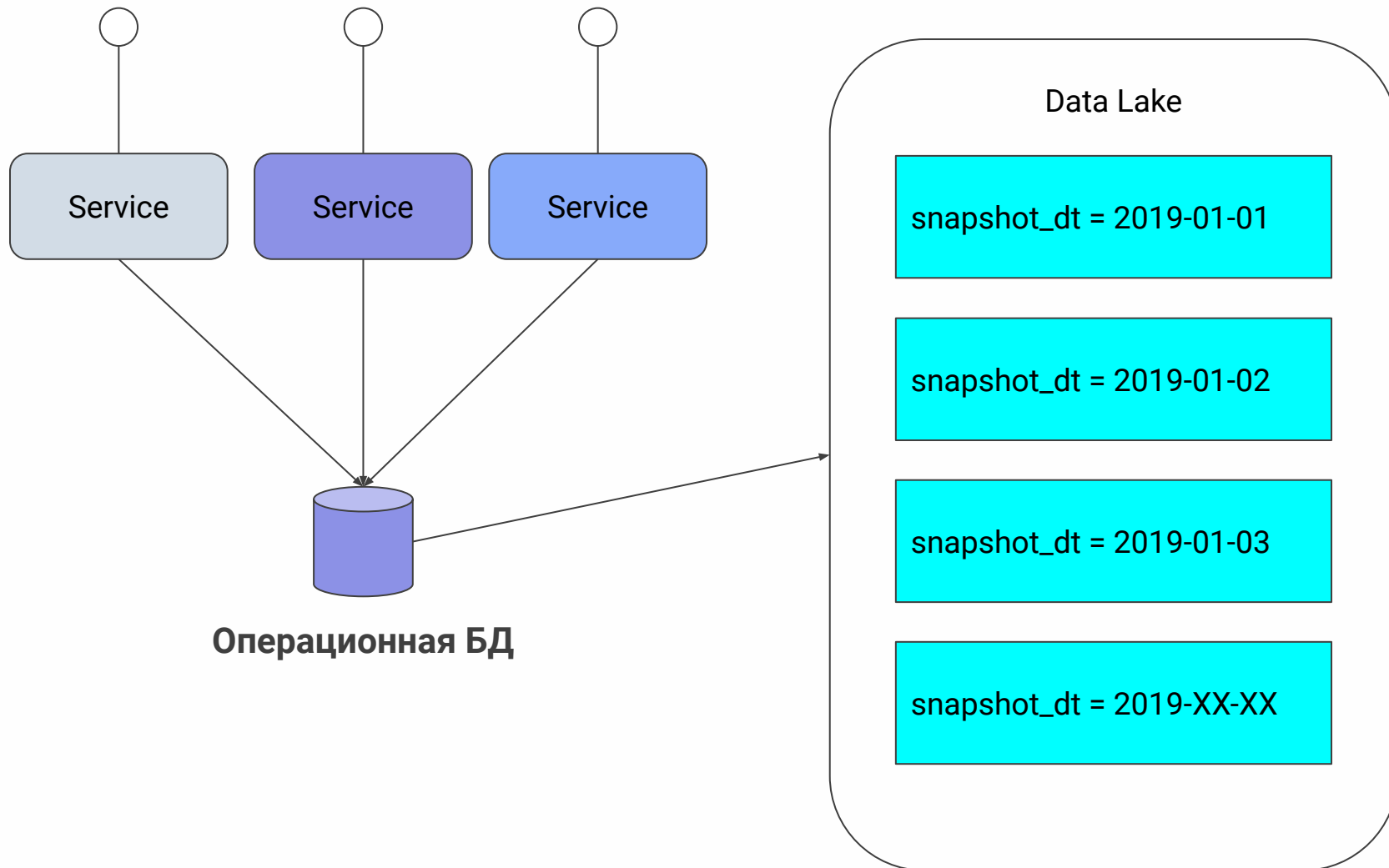


# 02

## Пассивная выгрузка

Snapshot

Раз в X - снимаем текущее состояние базы и сохраняем его (Snapshot)



## Pros:

- Архитектурно проще чем CDC
- Подходит для многих use-case
- Легкий мониторинг и шедулинг

## Cons:

- Требуется много места для хранения
- Потеря изменений внутри одного snapshot period
- Большая нагрузка на чтение / трансфер / запись
- Изменение схемы данных

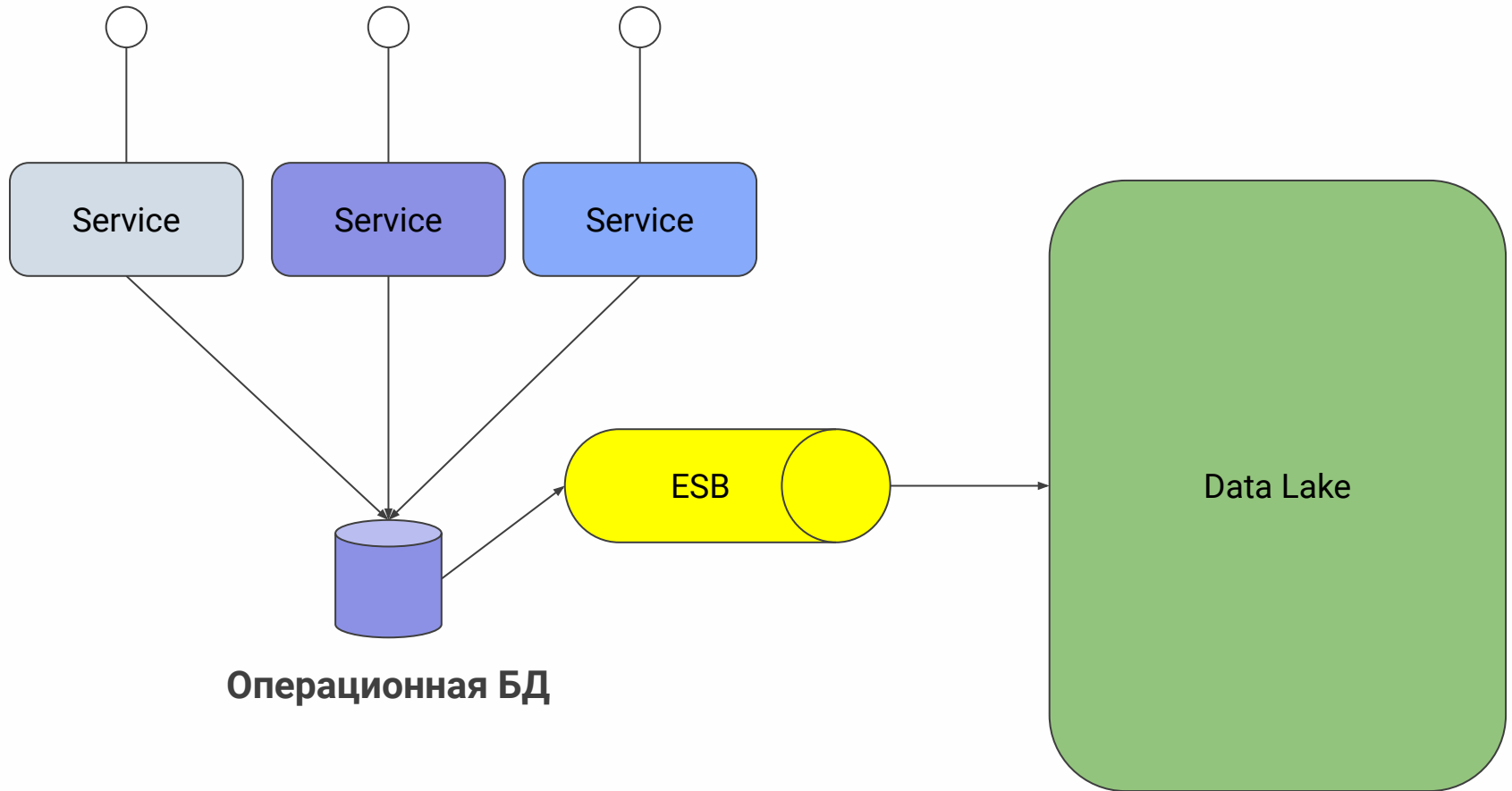


# 03

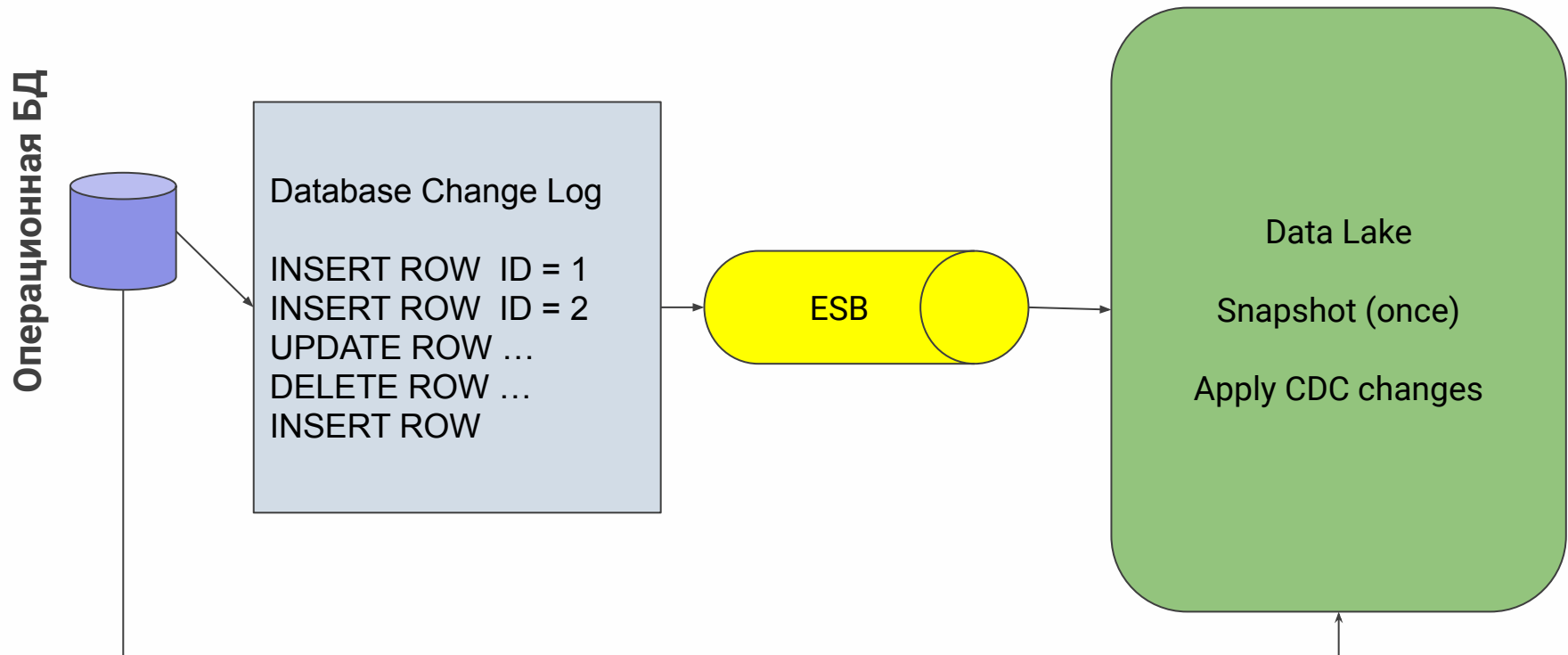
## Активная выгрузка

CDC

База самостоятельно отдает Change Log



## Change Data Capture - захват изменений данных



## Pros:

- 100% захват всех изменений
- Можно делиться изменениями по ESB
- Near-real time changes
- Меньше траффика
- Меньше нагрузка на БД (в некоторых случаях)
- Изменяемость схемы

## Cons:

- Требуется ESB
- Требуется настройки всех систем
- Требуется контроля и логов
- Не всегда exactly-once



**04**

**Практический пример**



# Трусов Иван

[ivan.trusov.contact@gmail.com](mailto:ivan.trusov.contact@gmail.com)

<https://renardeinside.github.io/>

**Спасибо  
за внимание!**

