

Возможности для интеграции PostgreSQL с ClickHouse



PostgreSQL



ClickHouse

ClickHouse

- высокопроизводительная
распределенная аналитическая
столбцовая СУБД

Q clickhouse...

- Векторизация выполнения запросов
- Сжатие данных
- Параллельное и распределенное выполнение запросов
- Обработка запросов в реальном времени
- Высокая скорость работы с большим объемом данных
- ...

Движки таблиц и баз данных

Движок таблицы (тип таблицы) определяет:

- Как и где хранятся данные, куда их писать и откуда читать.
- Какие запросы поддерживаются и каким образом.
- Использование индексов, если есть.
- Параметры репликации данных.
- ...

1 Семейство MergeTree

2 Семейство Log

3 Движки для интеграции

4 Специализированные
движки

Возможности для интеграции с PostgreSQL



Движки PostgreSQL

- Данные хранятся в PostgreSQL
- Для одних движков могут храниться метаданные в ClickHouse, для других узнаются во время запроса
- Database engine *PostgreSQL* - view в реальном времени на все таблицы базы в PostgreSQL

Движки MaterializedPostgreSQL

- Данные хранятся в ClickHouse
- Метаданные тоже в ClickHouse
- Database engine *MaterializedPostgreSQL* - данные уже на диске в ClickHouse

```
clickhouse :) CREATE TABLE postgres_table
              ENGINE = PostgreSQL('postgres1:5432', 'postgres_database', 'postgres_table',
                                   'user', 'pwd')

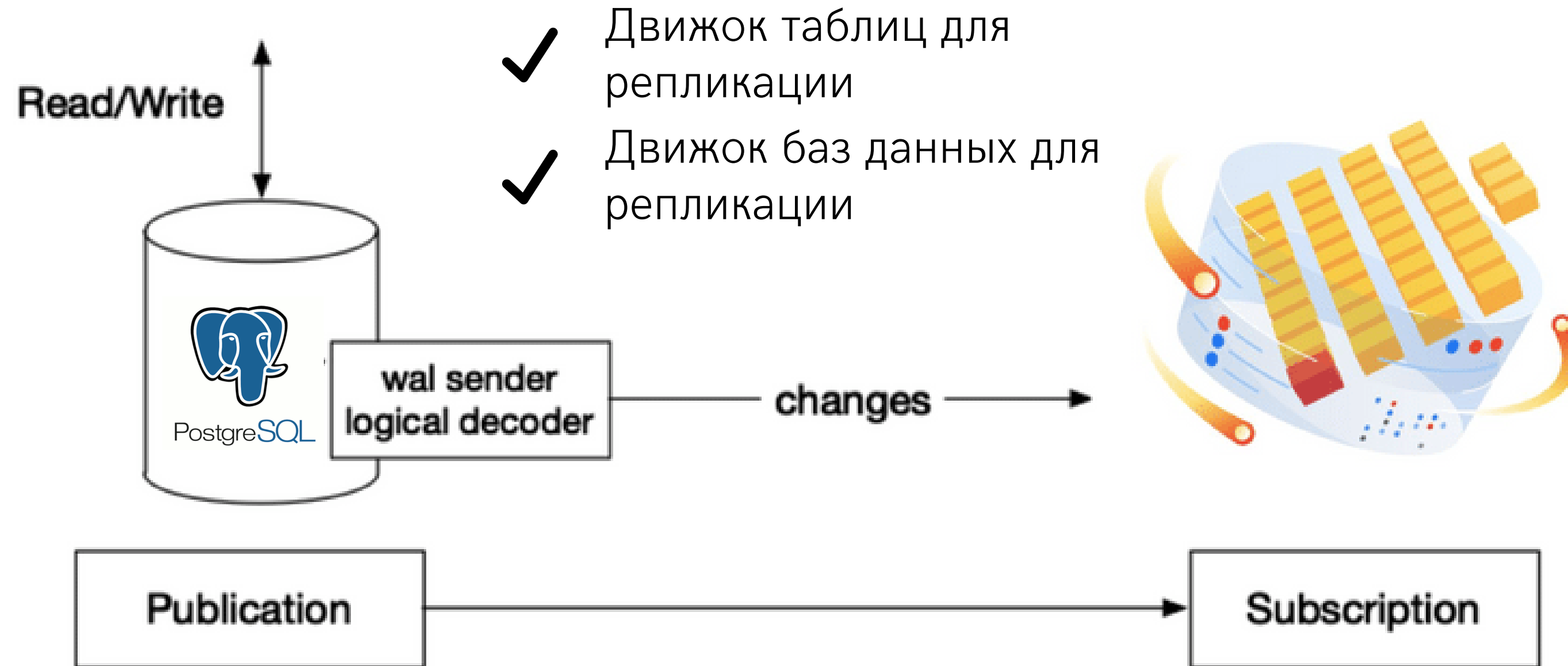
clickhouse :) CREATE DICTIONARY postgres_dict (id UInt32, val UInt32)
              PRIMARY KEY id
              SOURCE(PostgreSQL(
                port 5432
                host 'localhost'
                user  'postgres_user'
                password 'pwd'
                db 'postgres_database'
                table 'postgresql_table'))
              LIFETIME(MIN 1 MAX 2)
              LAYOUT(HASHED());
```

```
clickhouse :) CREATE DATABASE postgres_database_database
              ENGINE = PostgreSQL('postgres1:5432', 'postgres_database', 'postgres_table',
                                   'user', 'pwd')

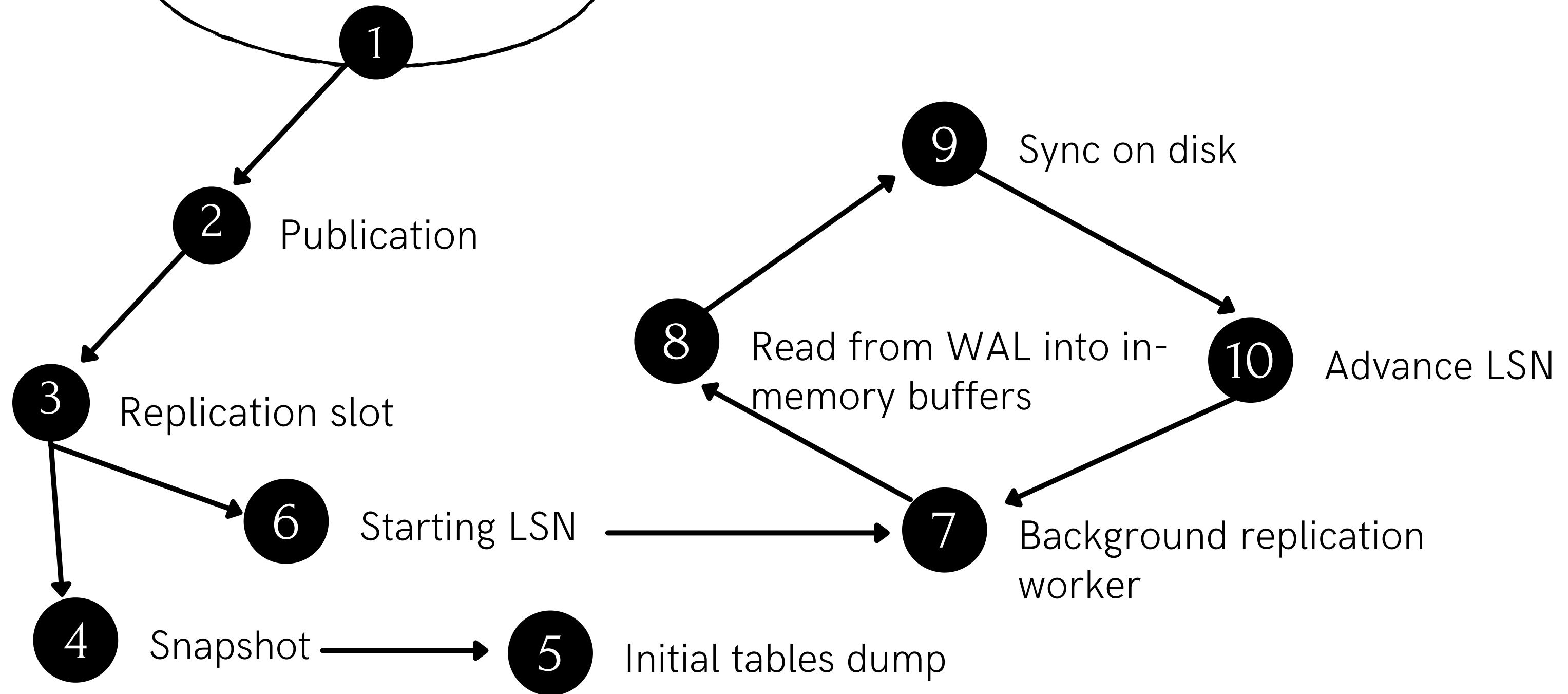
clickhouse :) SELECT * FROM postgres_database.table1;
clickhouse :) INSERT INTO postgres_database.table1;
clickhouse :) SELECT joined FROM postgres_database.table1 ANY LEFT JOIN (SELECT number * 2
AS number, number * 10 + 1 AS joined FROM system.numbers LIMIT 10) js2 USING number LIMIT 10

clickhouse :) INSERT INTO FUNCTION postgresql('localhost:5432', 'postgres_database',
        'postgres_table', 'user', 'pwd') SELECT number FROM numbers(10)
```

Репликация данных из PostgreSQL в ClickHouse



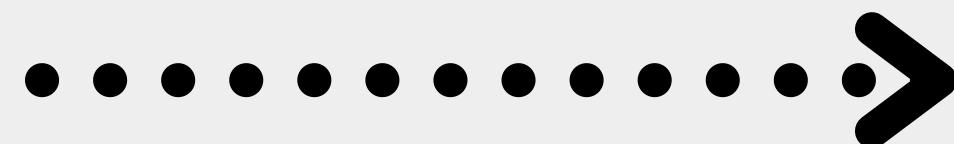

```
CREATE DATABASE postgresql_replica
ENGINE = MaterializedPostgreSQL(...params...)
```



Особенности

UPDATE -- DELETE

В ClickHouse доступна только
вставка данных, но не их изменение
или удаление



ReplacingMergeTree
table engine

Фоновая дедупликация данных на
основе на основе ключа сортировки
+
виртуальные колонки sign & version

+ FINAL

Дополнительная фильтрация при запросе

```
clickhouse :) CREATE TABLE postgresql_replica  
              ENGINE = MaterializedPostgreSQL('postgres1:5432', 'postgres_database',  
                                               'postgres_table', 'user', 'pwd')
```

```
clickhouse :) SELECT * FROM postgresql_replica WHERE ...
```



```
clickhouse :) SELECT * FROM postgresql_replica WHERE ... FINAL
```

key	_version	_sign
a	1	1
b	2	1
a	3	1

UPDATE



key	_version	_sign
a	3	1
b	2	1

key	_version	_sign
a	3	1
b	2	1
b	4	-1

DELETE



key	_version	_sign
a	3	1

key	_version	_sign
a	1	1

key	_version	_sign
b	2	1

key	_version	_sign
a	3	1

UPDATE

key	_version	_sign
a	3	1

key	_version	_sign
b	2	1

key	_version	_sign
a	3	1

DELETE

key	_version	_sign
a	3	1

key	_version	_sign
b	2	1

key	_version	_sign
b	4	-1

Репликационный идентификатор

LSN - Зачем?
Транзакционная целостность

Что делать с DDL?

- Не поддерживаются при логической репликации

Int16

Number of columns.

Next, the following message part appears for each column:

Int8

Flags for the column. Currently can be either 0 for no flags or

String

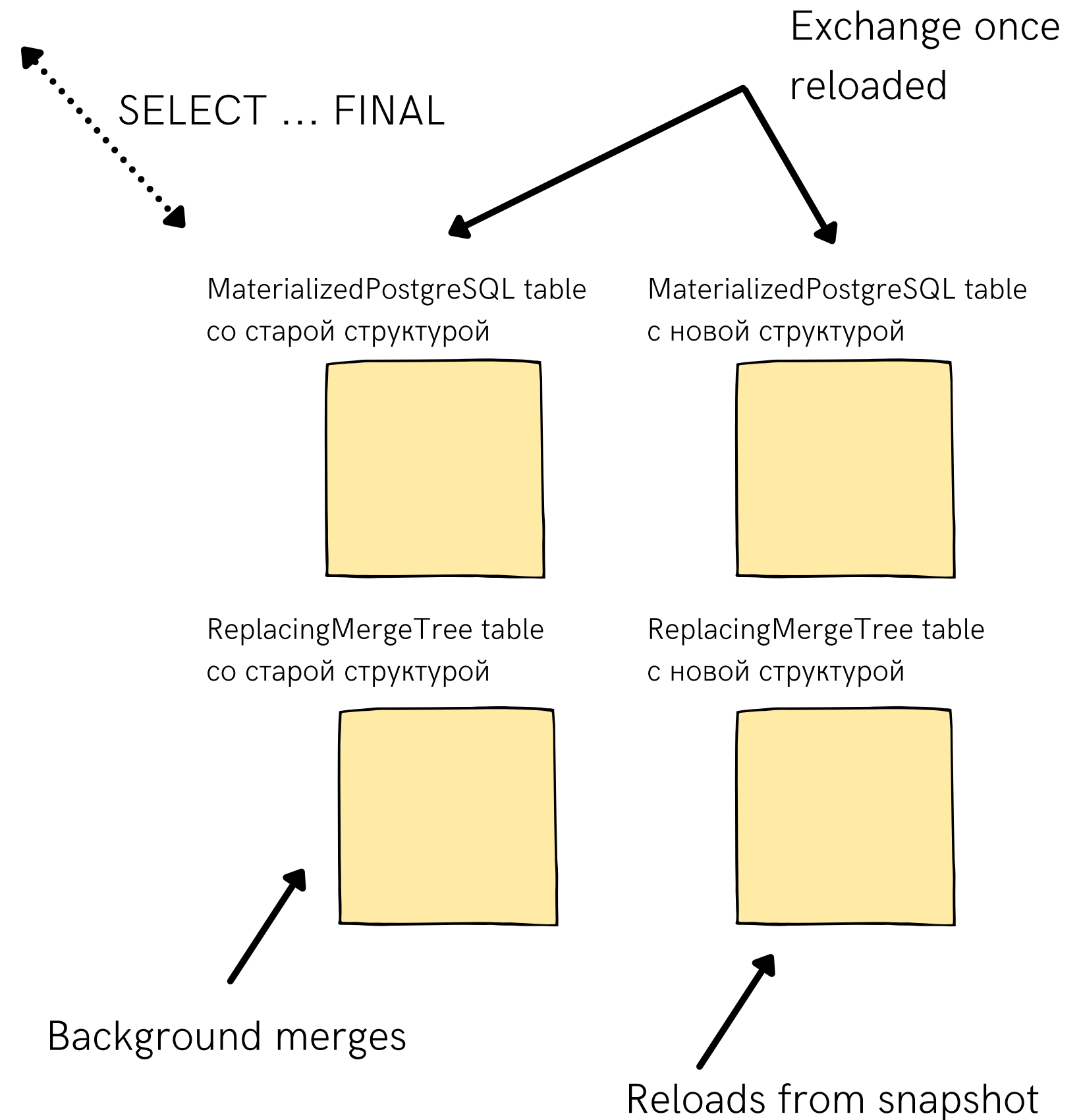
Name of the column.

Int32

ID of the column's data type.

Int32

Type modifier of the column (at ttypmod).



ИТОГ

- Быстрые запросы в ClickHouse над таблицами из PostgreSQL (в объединении с таблицами в ClickHouse) в реальном времени через:



- ClickHouse как реплика PostgreSQL



(Изменения на стороне публикации передаются в реальном времени и встака не блокирует чтение данных!)