



ClickHouse для инженеров и архитекторов БД

Мутация данных и манипуляции с партициями



Проверить, идет ли запись

Меня хорошо видно && слышно?

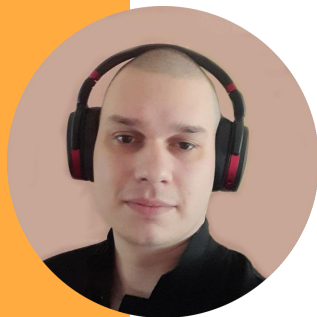


Ставим "+", если все хорошо
"-", если есть проблемы



Тема вебинара

Мутация данных и манипуляции с партициями



Константин Трофимов

Senior SRE / ClickHouse DBA в [VK](#)

Занимаюсь эксплуатацией ClickHouse с первых версий: 5 лет в VK, до этого в AdNow, до этого занимался Vertica. Сотни серверов, десятки кластеров, десятки петабайт данных.

Правила вебинара



Активно
участвуем



Off-topic обсуждаем
в учебной группе
#OTUS ClickHouse-2024-08



Задаем вопрос
в чат или голосом



Вопросы вижу в чате,
могу ответить не сразу

Условные обозначения



Индивидуально



Время, необходимое
на активность



Пишем в чат



Говорим голосом



Документ



Ответьте себе или
задайте вопрос

Карта курса



Темы модуля

«Масштабирование и манипуляции с данными»

1 (13). Проекция и материализованные представления

2 (14). Репликация и другие фоновые процессы

3 (15). Шардирование и распределенные запросы

4 (16). Мутация данных и манипуляции с партициями

5 (17). Сессия Q&A



Маршрут вебинара

Мутация данных и манипуляции с партициями



Мутация данных и манипуляции с партициями

Цели вебинара

После занятия вы сможете



1. Работать с данными ClickHouse на уровне партиций



Цели вебинара

Зачем это нужно



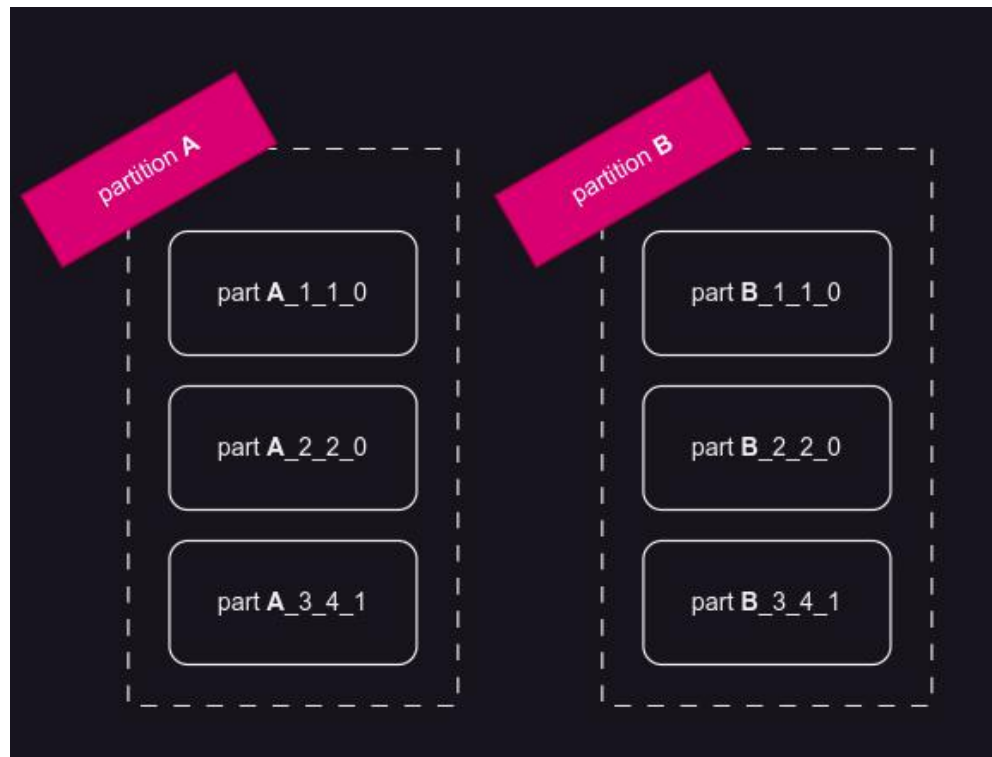
1. Резервное копирование, исправление данных

Парт и партиция

Парт (*part*) - минимальный атомарный наборный данных, в контексте репликации, фоновых процессов, записи, изменения данных.

Партиция (*partition*) - набор part-ов с одинаковым значением ключа партиционирования.

Подробнее: см. вебинар «4 (4). MergeTree и типы данных», слайды «Структура Part» и «Partition key»



Где находятся парты

У каждой *MergeTree-таблицы есть «data_paths» (можно найти в system.tables таблице), это каталог таблицы.

В каталоге таблицы находятся каталоги партов.

Каталог таблицы создается при создании таблицы, путь зависит от применяемого storage_policy или storage_policy по умолчанию, таковых каталогов может быть несколько если несколько дисков.

```
SELECT
  `table`,
  data_paths
FROM system.tables
WHERE data_paths != []
```

Query id: 8356ae12-f2af-4398-a6ec-11dc27907d0e

table	data_paths
1. asynchronous_metric_log	['/var/lib/clickhouse/store/591/59128e4d-287b-4165-95aa-1636746db9ff/']
2. asynchronous_metric_log_0	['/var/lib/clickhouse/store/c01/c01d121e-fb73-4cd7-bc8a-b19851a0f166/']
3. error_log	['/var/lib/clickhouse/store/b16/b161598b-9bde-442c-bba0-9cb81c5daf0d/']
4. metric_log	['/var/lib/clickhouse/store/c8b/c8b4cf3c-007b-435c-a234-f8a9686052b5/']
5. metric_log_0	['/var/lib/clickhouse/store/c8a/c8ae30da-ed41-4945-97f1-ca3088650db5/']
6. processors_profile_log	['/var/lib/clickhouse/store/785/7853acfe-4155-4271-8d6f-9acff8d97e2d/']
7. query_log	['/var/lib/clickhouse/store/715/715d71f9-d267-4a7d-acc8-487f4eefb629/']
8. query_log_0	['/var/lib/clickhouse/store/891/89121e3d-f4e4-420d-8267-7e410bfc7f44/']
9. text_log	['/var/lib/clickhouse/store/430/430b7818-2933-41e2-815f-36947d181070/']
10. trace_log	['/var/lib/clickhouse/store/aac/aac249aa-0068-416a-b531-4ae97fdc2552/']
11. trace_log_0	['/var/lib/clickhouse/store/cc4/cc416f97-6071-4d3f-8c71-5b56c02c1caf/']

11 rows in set. Elapsed: 0.006 sec.

```
ch-1S_1K :) Bye.
clickhouse@clickhouse:/$ ls /var/lib/clickhouse/store/591/59128e4d-287b-4165-95aa-1636746db9ff/
202411_10_10_0  202411_1_21_4  202411_1_41_8  202411_19_19_0  202411_24_24_0  202411_30_30_0  202411_35_35_0  202411_41_41_0  202411_7_7_0
202411_1_1_0   202411_1_26_5  202411_15_15_0  202411_20_20_0  202411_25_25_0  202411_31_31_0  202411_36_36_0  202411_42_42_0  202411_8_8_0
202411_11_11_0 202411_13_13_0 202411_1_6_1   202411_21_21_0  202411_26_26_0  202411_32_32_0  202411_37_37_0  202411_43_43_0  202411_9_9_0
202411_1_11_2  202411_1_31_6  202411_16_16_0  202411_2_2_0   202411_27_27_0  202411_3_3_0   202411_38_38_0  202411_4_4_0   detached
202411_1_16_3  202411_1_36_7  202411_17_17_0  202411_22_22_0  202411_28_28_0  202411_33_33_0  202411_39_39_0  202411_5_5_0   format_version.txt
202411_12_12_0 202411_14_14_0 202411_18_18_0 202411_23_23_0 202411_29_29_0 202411_34_34_0 202411_40_40_0 202411_6_6_0
```

Каталог *detached*

Помимо каталогов партов, в каталоге таблицы так же лежит:

- файл `format_version.txt` - используется для обратной совместимости при обновлении КХ
- каталог `detached`, с неиспользуемыми партами

Парты из `detached` каталога не используются таблицей, каталог `detached` нужен для импорта данных в формате партов, или как хранилище партов убранных из таблицы вручную или при восстановлении данных распознанных как поврежденные.

Содержимое каталогов `detached` можно смотреть прямо из ClickHouse, таблица ***system.detached_parts***.

Как получить список партов и партий

Список part-ов можно получить из **system.parts**, там же можно получить и список партий, а так же дополнительную информацию об их размере, участии в фоновых процессах, размеру до сжатия, и так далее.

```
SELECT
  `table`,
  partition,
  formatReadableSize(sum(bytes_on_disk)) AS size,
  groupArray(name) AS parts
FROM system.parts
WHERE `table` = 'asynchronous_metric_log'
GROUP BY ALL
FORMAT Vertical
```

Query id: 8d35505e-607b-4a0c-a176-8af09bb79225

Row 1:

table:	asynchronous_metric_log
partition:	202411
size:	4.21 MiB
parts:	['202411_1_361_72', '202411_1_366_73', '202411_1_371_74', '202411_1_376_75', '202411_1_381_76', '202411_1_386_77', '202411_1_391_78', '202411_1_396_79', '202411_1_401_80', '202411_1_406_81', '202411_1_411_82', '202411_1_416_83', '202411_1_421_84', '202411_1_426_85', '202411_1_431_86', '202411_362_362_0', '202411_363_363_0', '202411_364_364_0', '202411_365_365_0', '202411_366_366_0', '202411_367_367_0', '202411_368_368_0', '202411_369_369_0', '202411_370_370_0', '202411_371_371_0', '202411_372_372_0', '202411_373_373_0', '202411_374_374_0', '202411_375_375_0', '202411_376_376_0', '202411_377_377_0', '202411_378_378_0', '202411_379_379_0', '202411_380_380_0', '202411_381_381_0', '202411_382_382_0', '202411_383_383_0', '202411_384_384_0', '202411_385_385_0', '202411_386_386_0', '202411_387_387_0', '202411_388_388_0', '202411_389_389_0', '202411_390_390_0', '202411_391_391_0', '202411_392_392_0', '202411_393_393_0', '202411_394_394_0', '202411_395_395_0', '202411_396_396_0', '202411_397_397_0', '202411_398_398_0', '202411_399_399_0', '202411_400_400_0', '202411_401_401_0', '202411_402_402_0', '202411_403_403_0', '202411_404_404_0', '202411_405_405_0', '202411_406_406_0', '202411_407_407_0', '202411_408_408_0', '202411_409_409_0', '202411_410_410_0', '202411_411_411_0', '202411_412_412_0', '202411_413_413_0', '202411_414_414_0', '202411_415_415_0', '202411_416_416_0', '202411_417_417_0', '202411_418_418_0', '202411_419_419_0', '202411_420_420_0', '202411_421_421_0', '202411_422_422_0', '202411_423_423_0', '202411_424_424_0', '202411_425_425_0', '202411_426_426_0', '202411_427_427_0', '202411_428_428_0', '202411_429_429_0', '202411_430_430_0', '202411_431_431_0', '202411_432_432_0']

Вопросы?



Задаем
вопросы в чат



Ставим “-”,
если вопросов нет

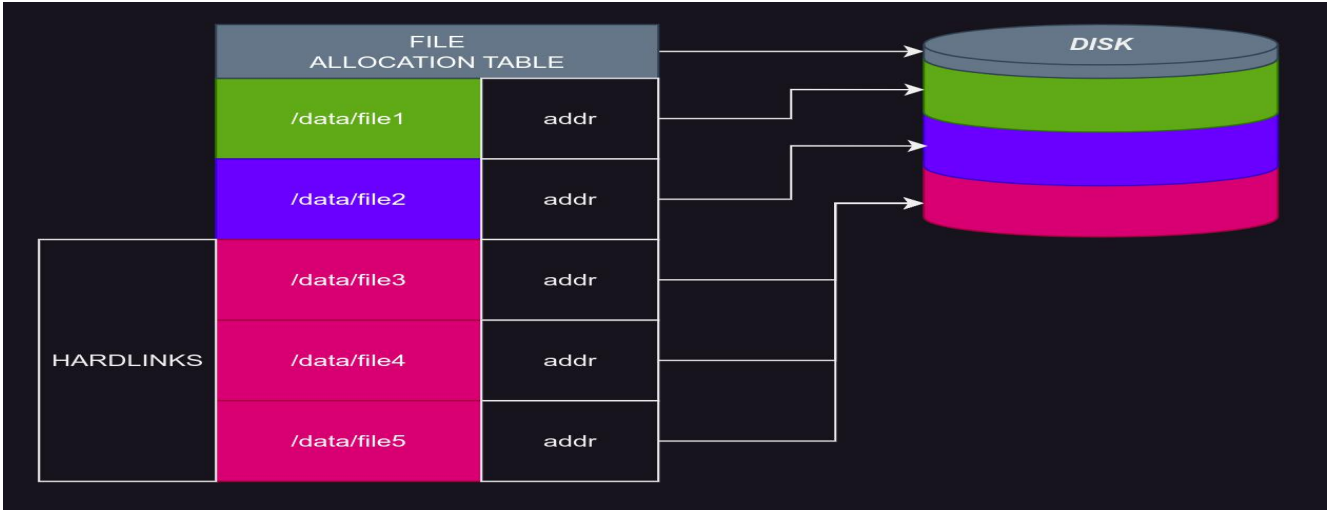


Что такое хардлинк

зачем нам знать что такое хардлинк

ClickHouse использует хардлинки при копировании партов, и создает с их (хардлинков) помощью копии данных таблиц для систем резервного копирования, поскольку это несравнимо дешевле и быстрее обычного копирования.

Хардлинк - ещё одно отображение файла в файловой системе, по дополнительному пути. Файл будет окончательно удален тогда, когда будет удалено последнее отображение. Отображения файла равнозначны.



Вопросы?



Задаем
вопросы в чат



Ставим “-”,
если вопросов нет



Операции над партами и партициями

что можно делать с партициями и партами

ClickHouse позволяет:

- удалять partition/part
- перемещать partition/part между таблицей и detached каталогом
- скачать партицию с другого шарда
- перемещать, копировать, заменять partition между таблицами одинаковой структуры
 - - между Replicated и не-Replicated можно!
 - - с меньшим набором тех же колонок можно!
 - - с другими ключами нельзя
- создавать на хардлинках копии данных таблиц, используется как артефакт утилит резервного копирования

Важно! Операции с партициями являются реплицируемыми.

Все операции над партициями имеют синтаксис «***ALTER TABLE database.table ОПЕРАЦИЯ***».

Удаление

```
ALTER TABLE table_name DROP PARTITION|PART partition_expr
```

удаляет парт/партицию из таблицы

Переместить из/в detached

```
ALTER TABLE table_name DETACH PARTITION|PART partition_expr;
```

```
ALTER TABLE table_name ATTACH PARTITION|PART partition_expr;
```

```
ALTER TABLE table_name DROP DETACHED PARTITION|PART ALL|partition_expr;
```

подчищает парт из detached

Перемещение между таблицами

переместить:

```
ALTER TABLE table_source MOVE PARTITION partition_expr TO TABLE table_dest;
```

заменить:

```
ALTER TABLE table_source REPLACE PARTITION partition_expr FROM TABLE table_dest;
```

скопировать:

```
ALTER TABLE table_source ATTACH PARTITION partition_expr FROM TABLE table_dest;
```

Работа с хардлинками

```
ALTER TABLE table_name FREEZE [PARTITION partition_expr] [WITH NAME 'backup_name']
```

```
ALTER TABLE table_name UNFREEZE [PARTITION 'part_expr'] WITH NAME 'backup_name'
```

Скачать с другого сервера

```
ALTER TABLE users FETCH PARTITION 201902 FROM '/clickhouse/tables/01-01/visits';
```

скачивает партицию в каталог detached, после этого её надо сделать ATTACH

Вопросы?



Задаем
вопросы в чат



Ставим “-”,
если вопросов нет

Мутация данных

Мутация данных

Мутация данных - изменение данных в ClickHouse.

Работает как фоновая задача, создающая новую копию всех данных партиции (или всех партиций если не указана конкретная), изменяющая в ходе копирования данные уже в копии. После завершения старые данные заменяются новыми.

Удаление данных:

```
ALTER TABLE [db.]table DELETE [IN PARTITION partition_id] WHERE filter_expr
```

Обновление данных:

```
ALTER TABLE [db.]table UPDATE column1 = expr1 [, ...] [IN PARTITION partition_id] WHERE filter_expr
```

Статус мутации и прогресс

После выполнения команды на запуск мутации, ClickHouse добавляет необходимые операции в `system.replication_queue`, а общий прогресс мутации можно наблюдать в `system.mutations`.

В момент выполнения задачи из `system.replication_queue`, относящейся к мутации, можно так же увидеть эту операцию в `system.merges`.

LIGHTWEIGHT DELETE

Физически не удаляет данные, только помечает их как удаленные (`_row_exists` виртуальный столбец).

Данные затем удаляются при следующих мержах.

Не работает с проекциями.

Синтаксис:

```
DELETE FROM [db.]table [IN PARTITION partition_expr] WHERE expr;
```



Вопросы?



Задаем
вопросы в чат



Ставим “-”,
если вопросов нет

Рефлексия

Цели вебинара

Проверка достижения целей

1. Работать с данными ClickHouse на уровне партиций

Вопросы для проверки

Вопросы для проверки

1. Как быстро скопировать данные в соседнюю таблицу, удалить данные по условию `where user='jhon'`, и заменить оригинальные данные?

Рефлексия



С какими впечатлениями уходите с вебинара?



Как будете применять на практике то, что узнали на вебинаре?

Следующие вебинары



27 ноября 2024 (среда)

Сессия Q&A



2 декабря 2024 (понедельник)

RBAC контроль доступа, квоты и ограничения



Ссылка на вебинар
будет в ЛК за 15 минут



Материалы
к занятию в ЛК —
можно изучать



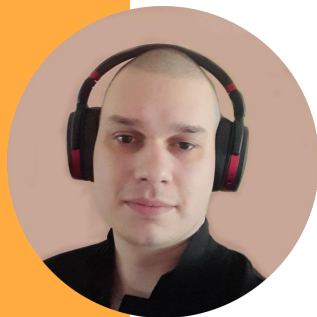
Обязательный
материал обозначен
красной лентой



**Заполните, пожалуйста,
опрос о занятии
по ссылке в чате**

Спасибо за внимание!

Приходите на следующие вебинары



Senior SRE / ClickHouse DBA в [VK](#)

Занимаюсь эксплуатацией ClickHouse с первых версий: 5 лет в VK, до этого в AdNow, до этого занимался Vertica. Сотни серверов, десятки кластеров, десятки петабайт данных.