

16. Соколова В.П., Габура А.В. К вопросу обогащения окисленных железных руд Ингулецкого месторождения /Сб.тр. КТУ. Разработка рудных месторождений, №84, 2003г. – С.153-158.

Рукопись поступила в редакцию 15.04.2020

УДК 681.03

І.Н. ВДОВИЧЕНКО, канд. техн. наук, доцент  
Криворізький національний університет

## СИСТЕМНИЙ АНАЛІЗ CLICKHOUSE - КОЛОНОЧНОЇ СУБД ДЛЯ ONLINE ОБРОБКИ АНАЛІТИЧНИХ ЗАПИТІВ

**Мета.** Проведення системного аналізу СУБД ClickHouse. Оцінка доцільності використання її у аналітичних системах.

**Методи дослідження.** У роботі використані: метод експертних оцінок, статистичні та аналітичні методи, комп'ютерні інформаційні та програмні технології. Залучення методів системного аналізу для вирішення проблем необхідно, перш за все, тому, що в процесі прийняття рішень доводиться здійснювати вибір в умовах невизначеності, яка обумовлена наявністю факторів, що не підлягають адекватній кількісній оцінці.

**Наукова новизна.** Проаналізовано сучасні потреби аналітичних обробок інформації, відмічено необхідність зберігати, аналізувати, перетворювати трильйони записів і петабайти даних. У роботі досліджено бази даних стовпчикового типу в цілому, інструменти для керування ними та їх переваги і недоліки. Предметом дослідження є система управління базами даних стовпчикового типу ClickHouse.

**Практична значимість.** При переході від рядкових СУБД до колоночних СУБД вираш у часі обробки запитів майже в 100 тисяч разів.

**Результати.** Було проведено аналіз будови СУБД ClickHouse, визначено можливості та особливості об'єкту, з'ясовано структурні особливості, які потенційно допомогли б впровадити її у роботу. Розглянуто показники, за якими оцінюють сучасні програмні продукти класу баз даних, виявлено основні можливості СУБД ClickHouse, порівняння її з іншими подібними інструментами, означено переваги і основні недоліки перед конкурентними продуктами галузі. У роботі конкретизовано характеристики колоночних СУБД. Представлено результати тестування колоночних СУБД. Запропоновано приклад для перевірки використання SQL для колоночних БД. Обґрунтовано доцільність переходу від рядкових БД до колоночних. Розглянуто спосіб зберігання інформації в ClickHouse. Виконуючи системний аналіз СУБД ClickHouse було отримано характеристики, на базі яких можна зазначити, що дана СУБД – це перспективний, швидкісний напрям, що розвивається, має великі можливості застосування, як самостійне ПЗ, так і в таких технологіях, як Big Date, Data Mining та інших.

**Ключові слова:** бази даних, колоночні СУБД, аналітичні запити, системний аналіз, швидкість.

doi: 10.31721/2306-5451-2020-1-50-165-169

**Проблема та її зв'язок з науковими і практичними завданнями.** Актуальність теми дослідження полягає у тому, що інформатизація майже усіх галузей призводить до стрімкого росту об'єму даних, які потрібно зберігати, аналізувати, перетворювати у людино-зрозумілий вигляд і надавати для подальших дій. Раніше із цими задачами могли справлятися реляційні «горизонтальні» бази даних - тобто бази, у яких інформація на фізичному рівні розташовувалася по рядках. З об'ємами даних, які наразі визначають основні вимоги до сховища даних, таким базам даних впоратися не під силу, тому постають питання:

використання іншого типу баз даних - «вертикальних», які дозволяють зберігати і швидко отримувати дуже великі об'єми даних;

вибір найбільш зручної системи управління такою базою даних.

Одним із інструментів, що вирішують ці питання, є ClickHouse - стовпчикова СУБД від компанії Яндекс, що була розроблена для задоволення потреб аналітичної системи компанії.

**Аналіз досліджень і публікацій.** Найчастіше тема СУБД ClickHouse піднімається на конференціях і в спец. лекціях для розробників ПЗ. Розглядаються принципи роботи систем аналітики, особливості збору і обробки даних та структуру СУБД ClickHouse. Література пропонує всілякі тестові аналізи і порівняння сучасних колоночних СУБД.

**Постановка задачі.** Об'єктом дослідження є бази даних стовпчикового типу в цілому, інструменти для керування ними та їхні переваги і недоліки.

Предметом дослідження є система управління базами даних стовпчикового типу ClickHouse.

Завданням дослідження є виявлення основних можливостей СУБД ClickHouse; порівняння її з іншими подібними інструментами; визначення переваг і основних недоліків перед конкурентними продуктами галузі.

Метою дослідження є проведення системного аналізу СУБД ClickHouse. Оцінка доцільності використання її у аналітичних системах.

Залучення методів системного аналізу для вирішення проблем необхідно, перш за все, тому, що в процесі прийняття рішень доводиться здійснювати вибір в умовах невизначеності, яка обумовлена наявністю факторів, що не підлягають адекватній кількісній оцінці.

Для системного аналізу властиво формування таких загальних уявлень про систему, як:

виявлення головних функцій (властивостей, цілей, призначення) системи;

виявлення основних частин (модулів) та їх функцій;

виявлення основних процесів, їх ролі, умов здійснення, виявлення змін станів у функціонуванні системи;

врахування змін і невизначеностей в системі;

дослідження функцій і процесів в системі з метою управління ними та інших.

**Виклад матеріалу і результати.** Популярні реляційні СУБД - Oracle, SQL Server, MySQL, PostgreSQL базуються на архітектурі, що була розроблена у минулому столітті, коли головною задачею баз даних було підтримати перехід від паперового обліку діяльності до комп'ютерного. Величезна (на той час) кількість інформації з паперових документів переносилася в БД облікових систем, які повинні були надійно зберігати всі вхідні відомості і, при необхідності, швидко знаходити їх. Такі вимоги зумовили архітектурні особливості реляційних СУБД, що залишилися до теперішнього часу практично незмінними: зберігання даних по рядках, індексування записів і журнал операцій [5].

Однак відносно нещодавно, з поширенням аналітичних інформаційних систем і сховищ даних, які застосовувались для управлінського аналізу, стало зрозуміло, що характер навантаження в цих двох видах систем радикально відрізняється. Якщо звичайним транзакційним додаткам властиво часті, але дрібні операції додавання/зміни записів, то для аналітичних систем навпаки, важливі вибірки великих обсягів даних при відносно рідкісному їх додаванні/оновленні (запис даних найчастіше також іде великими блоками, тому операція відбувається набагато рідше за вибірку). Варто зазначити, що у таких вибірках є важлива особливість - зазвичай, вони містять невелику кількість полів (дуже рідко – більше 7-8 полів, хоча у аналітичній таблиці кількість полів зазвичай переважає 25-30).

Для таких вимог, які поставили аналітичні системи, виникла потреба у зміні способу зберігання і доступу до даних задля прискорення виконання операцій з великими вибірками. Таким рішенням стало використання «вертикальних» (стовпчикових) баз даних, які «повернули» дані на 90°. Основна ідея стовпчикових СУБД - це зберігання даних не по рядках, як це роблять традиційні СУБД, а по колонках. Це означає, що з точки зору SQL-клієнта дані представлені як за звичайно, у вигляді таблиць, але фізично ці таблиці є сукупністю колонок, кожна з яких по суті являє собою таблицю з одного поля. При цьому фізично на диску значення одного поля зберігаються послідовно один за одним (рис. 1).

На рис.1 та рис. 2 показано відмінність у зберіганні та доступі до даних у стовпчикових та рядкових БД. При зберіганні за стовпцями відбувається читання 2-х стовпців, а при зберіганні по рядках відбувається читання всіх стовпців.

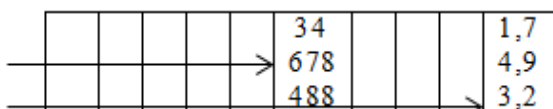


Рис. 1. Зберігання за стовпцями

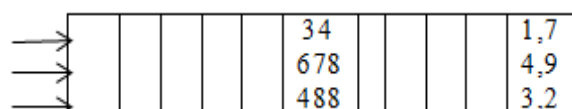


Рис. 2. Зберігання по рядках

Однією із перших стовпчикових баз даних, що стала доступною для використання є Sybase TQ. Якийсь час ця СУБД була монополістом у своїй галузі, тому встигла завоювати прихильників серед користувачів. Найбільшою перевагою цієї СУБД є швидкість завантаження та індексації даних: згідно офіційного сайту компанії, станом на 2013 рік за цим показником було

поставлено світовий рекорд. Також вона надає функціональність збережених процедур, що є важливим аспектом гнучкості системи в отриманні масивів даних. Серед недоліків – направленість на використання у корпоративних проектах enterprise-рівня, що робить його доволі громіздким для впровадження у малі та середні проекти [4].

Vertica – потужний інструмент, однією із великих переваг якого є відсутність обмежень у виборі апаратної платформи, є набір конфігурацій для необхідного користувачу обладнання. Vertica має безкоштовну версію, що сприяє поширенню серед невеликих проектів, що потребують обробки аналітичних даних. Одним із вагомих недоліків СУБД є відсутність індексів, замість них використовується поняття «проекцій» даних, але це не є оптимальним із точки зору швидкодії системи.

Exasol – поширений представник стовпчикових СУБД. Серед переваг цієї СУБД є зберігання даних у таблицях у стислому вигляді, що значно прискорює аналітичні запити, а також орієнтованість системи на зберігання даних у оперативній пам'яті, що також прискорює обробку даних. Exasol підтримує SQL-синтаксис, що дозволяє користувачу звичайних баз даних легше перейти на використання стовпчикової БД. Але Exasol орієнтований на enterprise-розробку, як і Sybase не має безкоштовної версії та не дуже поширений в Україні.

ClickHouse - це стовпчикова СУБД для OLAP систем (online обробки аналітичних запитів). Вона була розроблена компанією Яндекс для задоволення потреб Яндекс.Метрики - другої за величиною аналітичною системою у світі. Серед усіх перерахованих вище СУБД ClickHouse обрано за об'єкт дослідження через велику кількість переваг, серед яких:

- можливість лінійного масштабування додаванням серверів;
- відмовостійкість;
- паралельна обробка процесів на багато процесорних ядрах;
- СУБД оптимізована для роботи на жорстких дисках (можна обробляти дані, на які не вистачає оперативної пам'яті);
- векторні обчислення - за рахунок чого досягається висока ефективність і швидкодія;
- open-source проект, безкоштовний, велика спільнота користувачів на території СНД;
- та багато інших особливостей [1].

Стовпчикові бази такі називаються тому, що зберігають дані не по рядках, а по стовпчиках. Кожен окремий стовпчик - це ніби окрема таблиця з однієї колонки, яка зберігає тільки свої значення. Це дозволяє вибирати лише необхідні дані із певних стовпчиків, не перебираючи усі стовпчики у рядках, як це роблять звичайні СУБД.

Всі дані в стовпчиковій базі даних зберігаються в відсортованому вигляді (кожен стовпчик окремо). Для запитів на вибірку даних, що складають переважну більшість запитів від аналітичної системи до СУБД, це критично важливо: такі запити містять декілька умовних операторів, групові операції, сортування за певними критеріями. Ці запити, завдяки описаному принципу зберігання даних виконуються майже миттєво і можуть обробляти величезні об'єми даних (10 мільйонів записів - не проблема).

На більш високому рівні БД складається зі звичних реляційних СУБД таблиць. ClickHouse надає величезну кількість вбудованих функцій для отримання аналітичних даних - квантілі візних рівнів (у тому числі медіана вибірки), дисперсію вибірки, коефіцієнт кореляції тощо.

Для роботи із СУБД надається широкий спектр можливостей, а саме:

- клієнт командної строки;
- HTTP-інтерфейс;
- JDBC-драйвер;
- бібліотеки сторонніх розробників для таких мов програмування, як PHP, Python, NodeJs, Ruby, Java, C++ та багато інших;
- веб-інтерфейс Tabix та UI/IDE для OSX, Linux и Windows - HouseOps.

СУБД ClickHouse було впроваджено для збирання статистичних даних щодо конверсії та повноти читання матеріалів на сайтах періодичних видань республіканського інформаційного центру. Щоденний обсяг відвідування порталу, на якому базуються сайти перевищує 50 тисяч унікальних користувачів, отже для зберігання і обробки усіх дій цих користувачів протягом тривалого періоду часу можливостей і швидкості звичайної СУБД не вистачило б. СУБД ClickHouse було налаштовано на окремому сервері, та налагоджено процес передачі даних до

нього, використовуючи модифікований PHP-компонент, СУБД MySQL для тимчасового зберігання та веб-інтерфейс [2].

Наступним етапом було побудова користувацького інтерфейсу додатку згідно із вимогами щодо аналітичної інформації.

ClickHouse використовує власний діалект SQL близький до стандартного, але містить різні розширення: масиви та вкладені структури даних, функції вищого порядку, імовірнісні структури.

Система оптимізована для зберігання даних на жорстких дисках (використовуються переваги лінійного читання, стиснення даних). У багатьох тестах ClickHouse показує дуже високу продуктивність, вигравши за цим показником у таких конкурентів як Greenplum, Vertica, Amazon Redshift, Druid, InfiniDB / MariaDB ColumnStore, Apache Spark, Presto, Elasticsearch.

Спочатку в Яндекс.Метрика для побудови звітів використовувалися попередньо агреговані дані. Цей підхід дозволяв зменшити розмір збережених даних, але мав ряд обмежень і недоліків:

- необхідність заздалегідь зафіксувати список доступних для користувача звітів (відсутність можливості побудувати довільний звіт);

- преагрегації по великій кількості ключів або по ключам високою кардинальністю (таким як URL) може призводити до зворотного ефекту (збільшення обсягу даних);

- підтримку логічної цілісності при зберіганні великої кількості різних агрегацій важко.

Альтернативний підхід полягає в зберіганні «сирих» не агрегованих даних, виробляючи всі необхідні обчислення в момент запиту користувача. Для цього була необхідна СУБД, яка змогла б обробляти не агреговані дані Яндекс.Метрики (петабайти даних) з дуже високою ефективністю і в реальному часі. До кінця 2014 року була запущена Метрика 2.0, що працює на базі ClickHouse, яка дозволила користувачам будувати довільні звіти. У червні 2016-го року вихідний код системи був викладений в open-source під ліцензією Apache 2.0 [3].

Основні характеристики ClickHouse:

- колоночна СУБД, поруч з даними не зберігаються інші величини, наприклад, в ClickHouse підтримуються значення постійної довжини, щоб не зберігати поруч розмір типу;

- лінійна масштабованість. Здатність працювати з великими даними (трильйони записів і петабайт даних).

ClickHouse розроблялась для OLAP сценарію роботи:

- в системі кілька слабко пов'язаних таблиць з великою кількістю параметрів / стовпців (порядку сотень);

- запити можуть використовувати велику кількість рядків, але тільки невелика підмножина стовпців;

- в основному, запити на читання є досить рідкісні (зазвичай не більше 100 грс на сервер);

- потрібна висока пропускна здатність при обробці одного запиту (до мільярдів рядків в секунду на один сервер);

- результат виконання запиту істотно менше вихідних даних - тобто, дані фільтруються або агрегуються;

- порівняно простий сценарій поновлення даних, зазвичай тільки додавання даних пачками; немає складних транзакцій [9].

ClickHouse можна використовувати як внутрішнє сховище даних для аналітиків.

Вже відмічалось, що ClickHouse це колончаста система, отже, стиснення даних виконується добре, тому що дані з однієї колонки зберігаються, в одному файлі на жорсткому диску і, якщо там приблизно однакова інформація, то стискається дуже ефективно - а якщо зберігати дані по рядках, то інформація різна в кожному полі (з них складається рядок) і стискається погано.

Швидкісні характеристики ClickHouse неймовірні, в сотні тисяч разів вище, в порівнянні з Hadoop, MySQL, PostgreSQL [10].

У ClickHouse використовується векторна обробка даних, отже дані не обробляються по рядках, обробляються тільки колонками. Це дозволяє дуже прискорити обробку [6-8].

Для перевірки можливостей SQL в ClickHouse створимо таблицю expert зі стовпцями fio, gr та stag типу String, Int, Int відповідно, для цього вводимо наступні команди:

```
CREATE TABLE expert(fio String, gr Int, stag Int) ENGINE=Memory
```

Для введення записів використовуємо команду INSERT.

INSERT INTO expert VALUES ('Ivanov T.K.', 1984, 8), ('Petrov C.D.', 1977, 17), ('Sidorov A.I.', 1987, 5)

Для перегляду використовуємо команду SELECT:  
SELECT \* FROM expert

Результат представлено в табл.1.

Визначимо експерта, який народився після 1985 року.

SELECT fio FROM expert WHERE gr >1985

Видаємо таблицю командою DROP TABLE.

DROP TABLE expert

Видно, що при елементарних операціях - це звичайний SQL.

Таблиця 1

Таблиця expert

fio	gr	stag
Ivanov T.K.	1984	8
Petrov C.D.	1977	17
Sidorov A.I.	1987	5

**Висновки і напрямки подальших досліджень.** В ході дослідження було обстежено ринок продуктів в предметній області БД, виявлено продукти, що найбільш часто використовують, проаналізовано їх переваги та недоліки.

Після виокремлення предмету дослідження було проведено аналіз будови СУБД ClickHouse, було визначено можливості та особливості об'єкту, визначено структурні особливості, які потенційно допомогли б впровадити її у роботу.

Виконуючи системний аналіз СУБД ClickHouse було отримано вище вказані характеристики, на базі яких можна зазначити, що дана СУБД – це перспективний, швидкісний напрям, що розвивається, має великі можливості застосування, як самостійне ПЗ, так і в таких технологіях, як Big Date, Data Mining та інших.

#### Список літератури

1. ClickHouse. <https://ru.wikipedia.org/wiki/ClickHouse>.
2. ClickHouse. Руководство. [http://clickhouse.readthedocs.io/en/latest/reference\\_ru.html](http://clickhouse.readthedocs.io/en/latest/reference_ru.html).
3. Александр Календарев ClickHouse в системах сбора статистики. // Системный администратор, 2017 г. - №1-2 С. 56-59.
4. Башкеев А.А., Миловидов А.Н. Коротко о СУБД ClickHouse. Тезисы докладов конференции «Интеллектуальные системы хранения и обработки информации». www.osp.ru/iz/tbd\_dbms Москва, 29-30 ноября 2016 г.
5. Томас Кайт, Кун Дарл Oracle для профессионалов. - М. Вильямс, 2016. - 960с.
6. Урсатьев А.А. Большие Данные. Аналитические базы данных и хранилища: Vertica, Kdb. //Управляющие системы и машины. 2018, - № 1. - С. 58-71.
7. Vertica. <http://www.vertica.com/>
8. Майкл Стоунбрейкер. Один размер пригоден для всех»: идея, время которой пришло и ушло. [http://citforum.ru/database/articles/one\\_size\\_fits\\_all/](http://citforum.ru/database/articles/one_size_fits_all/)
9. Колоночные СУБД — принцип действия, преимущества и область применения. <https://habr.com/ru/post/95181/>
10. Алексей Князев. Колоночные индексы, особенности использования. [http://www.t-ql.ru/post/columnstore\\_index\\_faq.aspx](http://www.t-ql.ru/post/columnstore_index_faq.aspx)

Рукопис подано до редакції 09.04.2020

УДК 621.867.157 (088.8)

Ю.Г. ГОРБАЧОВ, канд. техн. наук, проф., А.О. ХРУЦЬКИЙ, канд. техн. наук, А.С. ГРОМАДСЬКИЙ, д-р техн. наук, проф., О.С. ЛІФЕНЦОВ, ст. викладач Криворізький національний університет

### ОБГРУНТУВАННЯ ПАРАМЕТРІВ ЕКСПЕРИМЕНТАЛЬНОГО СТЕНДУ ДЛЯ ДОСЛІДЖЕННЯ ВІБРАЦІЙНИХ МАШИН ТА ПРОЦЕСУ ВІБРОТРАНСПОРТУВАННЯ

**Мета.** Необхідність подальшого розвитку гірничо-металургійного комплексу потребує оснащення гірничих підприємств новим технологічним обладнанням. Метою роботи є забезпечення можливості експериментального дослідження особливостей технологічного процесу вібраційного транспортування сипких вантажів та вібраційних машин для його реалізації. Важливість проблеми обумовлюється необхідністю подальшого розширення застосування вібраційних технологій і техніки у гірничому виробництві для підвищення його ефективності. Для вибору та обґрунтування раціо-