

ClickHouse Cloud Update

Sept 7, 2023

||||· ClickHouse

ClickHouse Cloud - from 0 to 1 in under a year



Private Preview

May - July 2022

Public Beta

Oct 2022

General Availability

Dec 2022



Serverless hosted ClickHouse for key design partners

- AWS with 3 regions initially
- Limited scalability within predefined limits
- Basic cloud console for ClickHouse service and user management
- Strong security and privacy with SOC 2 Type I compliance

Serverless hosted ClickHouse for early adopters

- Ecosystem of first-party connectors to onboard and work with data
- Automatic scaling up and down for compute
- Integrated billing using pay as you go pricing model
- Enhanced security features such as Private Link, IP Filtering, Auditing

Serverless hosted ClickHouse for the broader market

- Enhanced cloud console for analytics and operational controls
- Support for more AWS regions and AWS marketplace billing
- Uptime SLA and additional operations tools
- Advanced security features with SOC 2 Type II compliance



ClickHouse Cloud



Soon

Metrics Backups Connection Settings

Running

Location Ohio (us-east-2) Version ClickHouse 23.8 Last successful backup 6 hours ago Created Oct 18, 2022

Aggregation period 1 hour Time period Last week

Data stored **650.99 GB** Increase of 9 GB in last week

Data stored over time

Memory allocation **720 GiB**

Total read size **2.21 TB** Average read size per second: 3.65 MB

Read throughput (bytes per second)

Total inserts **5 M** Average inserts per second: 8.3

Successful queries • 100% Failed queries • 0%

SQL statements over time

Expensive properties

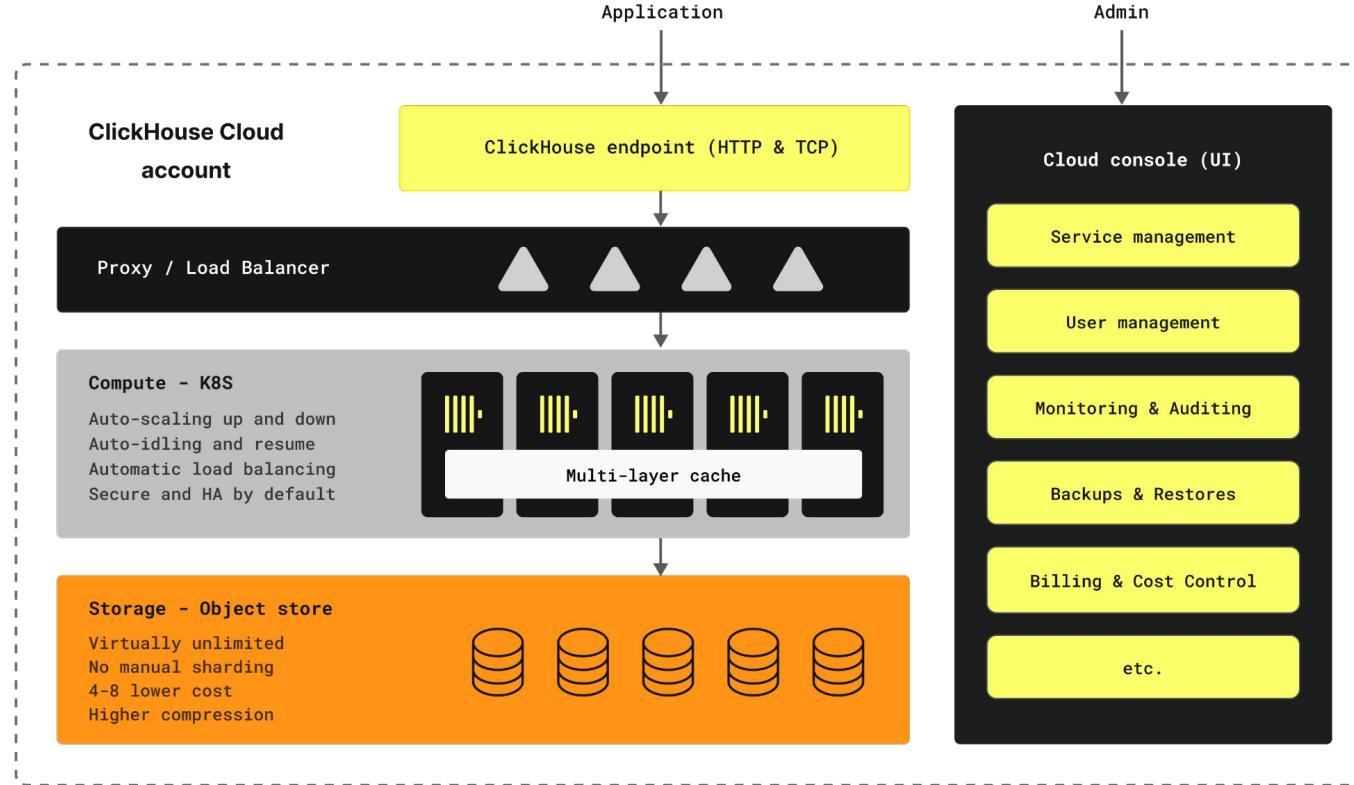
```
-- the price, street and district of the most expensive properties in London in 2020 where the post code starts with W15
SELECT price, time, street, addr1, addr2, district, postcode
FROM pp_complete
WHERE town ILIKE '%London%' AND postcode ILIKE 'W15%' AND time LIKE '2020%'
ORDER BY price DESC
LIMIT 10;
```

Search results... Elapsed: 0.785s Read: 27,450,499 rows (937.59 MB)

#	price	time	street	addr1	addr2
1	366180000	2020-10-08 00:00	NEW BOND STREET	159	
2	366180000	2020-10-08 00:00	NEW BOND STREET	158	
3	305211030	2020-10-08 00:00	NEW BOND STREET	158	
4	114712500	2020-07-02 00:00	NEW BOND STREET	141	
5	95086435	2020-10-07 00:00	SAVILLE ROW	25	
6	560000000	2020-03-20 00:00	MADDOX STREET	23	
7	560000000	2020-03-20 00:00	MADDOX STREET	29	
8	560000000	2020-03-20 00:00	MADDOX STREET	27	
9	421080000	2020-06-30 00:00	SACKVILLE STREET	1	

All Rows

ClickHouse Cloud Architecture



Cloud Pricing

Development

Great for smaller workloads and starter projects

\$50 - \$193 / Month

- ✓ Up to 1 TB storage
- ✓ 16 GiB total memory
- ✓ Burstable CPU
- ✓ Backups every 24h, retained 1 day
- ✓ Replicated across 2 AZs
- ✓ Expert support with 24h response time

Storage

\$35.33 per TB/mo

Compute

\$0.2160 per unit/hr

Production

Designed to handle production workloads

Usage Based

- ✓ Unlimited storage
- ✓ 24GiB+ total memory
- ✓ Dedicated CPU
- ✓ Backups every 24h, retained 1 day
- ✓ Replicated across 3 AZs
- ✓ Expert support with 1h SLA
- ✓ AWS private link support
- ✓ Automatic scaling

Storage

\$47.10 per TB/mo

Compute

\$0.6888 per unit/hr

Dedicated

Designed for the most demanding latency-sensitive workloads

Capacity Based

- ✓ Unlimited storage
- ✓ Custom compute options
- ✓ Dedicated environment
- ✓ Advanced isolation and security
- ✓ Customized backup controls
- ✓ Scheduled upgrades
- ✓ Uptime SLAs
- ✓ Consultative migration guidance
- ✓ Dedicated support engineers

Contact us



Topics

01

Cloud Platform

Krithika & Zach / 15 mins

02

Integrations

Ryadh / 15 mins

03

ClickHouse Core

Alexey / 15 mins

04

Roadmap

Tanya / 10 mins

05

Q&A



01

Cloud platform

Krithika & Zach

Cloud Platform

- API & Terraform provider
- Secure S3 access via assumed role
- CMEK/BYOK for on-disk encryption
- ISO27001 compliance
- Improved data loading (DEMO)





Automation and Programmability

- APIs to support programmatic access to manage service lifecycle
- Terraform provider to automate provisioning and configuration management

A screenshot of a ClickHouse Terraform provider page. At the top left is the ClickHouse logo. Next to it is the text "clickhouse" and a blue square icon. Below this is a large black square containing the ClickHouse logo. To the right of the logo is the word "clickhouse". Underneath is a blue button labeled "Partner" and the text "by:ClickHouse". Below that is a dark button labeled "Database". Further down are sections for "VERSION 0.0.3", "PUBLISHED 20 days ago", and "SOURCE CODE ClickHouse/terraform-provider-clickhouse".

A screenshot of a JSON API response titled "result": [. . .]. The response contains a list of ClickHouse services, each represented by a JSON object. One service entry is shown in full:

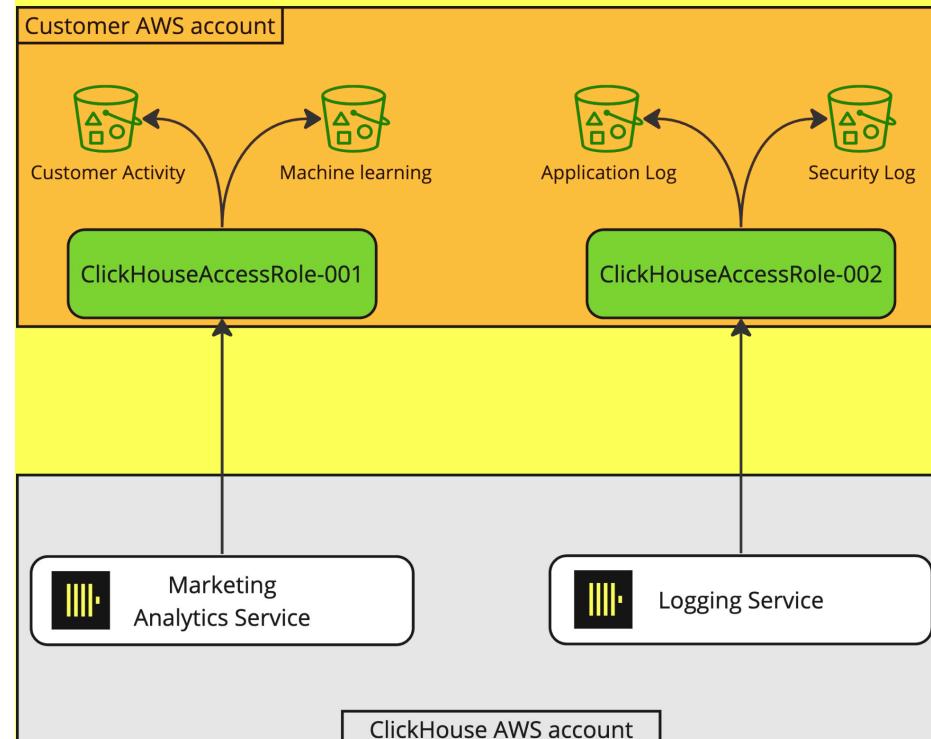
```
    {
        "id": "62ac2533-eb9b-486c-b3ce-17983fdc635b",
        "name": "AnalyticsService",
        "provider": "aws",
        "region": "us-east-2",
        "state": "running",
        "endpoints": [
            {
                "protocol": "nativesecure",
                "host": "dz2kb5ctwz.us-east-2.aws.clickhouse.cloud",
                "port": 9440
            },
            {
                "protocol": "https",
                "host": "dz2kb5ctwz.us-east-2.aws.clickhouse.cloud",
                "port": 8443
            }
        ],
        "tier": "production",
        "idleScaling": true,
        "idleTimeoutMinutes": 5,
        "minTotalMemoryGb": 192,
        "maxTotalMemoryGb": 720,
        "ipAccessList": [
            {
                "source": "111.52.22.111",
                "description": "RemoteOfficeIP(Asia)"
            },
            {
                "source": "104.51.40.101",
                "description": "HQOfficeIP(Portland)"
            }
        ],
        "createdAt": "2023-06-21T23:26:45Z"
    }
```

The JSON object includes fields like id, name, provider, region, state, endpoints, tier, idleScaling, idleTimeoutMinutes, minTotalMemoryGb, maxTotalMemoryGb, ipAccessList, and createdAt.



Secure S3 Access

- Connect securely to private S3 bucket using assumed roles
- Enables secure access to private file storage in AWS and ingest to ClickHouse Cloud



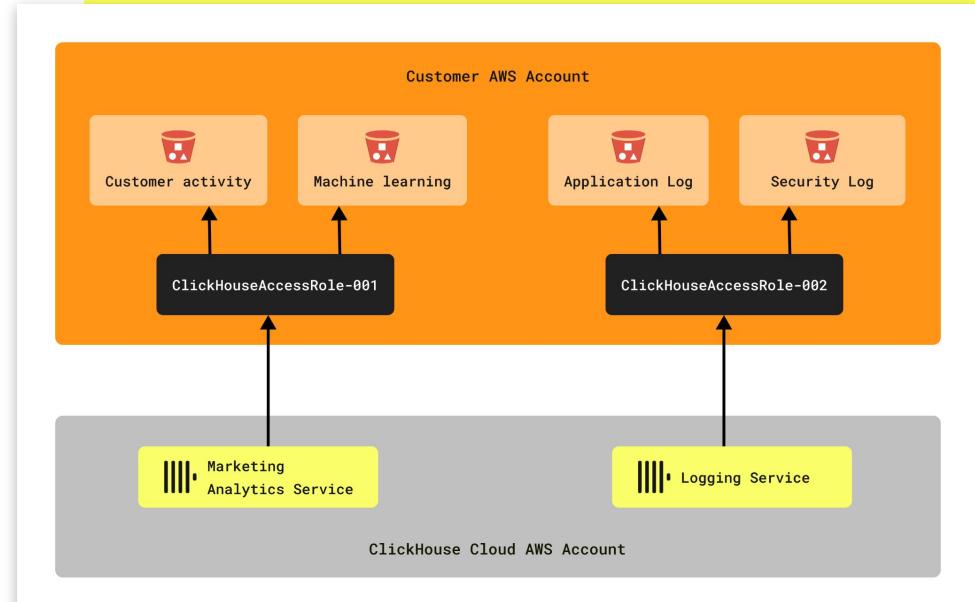
Contact support to request access





Secure S3 Access

- Connect securely to private S3 bucket using assumed roles
- Enables secure access to private file storage in AWS and ingest to ClickHouse Cloud

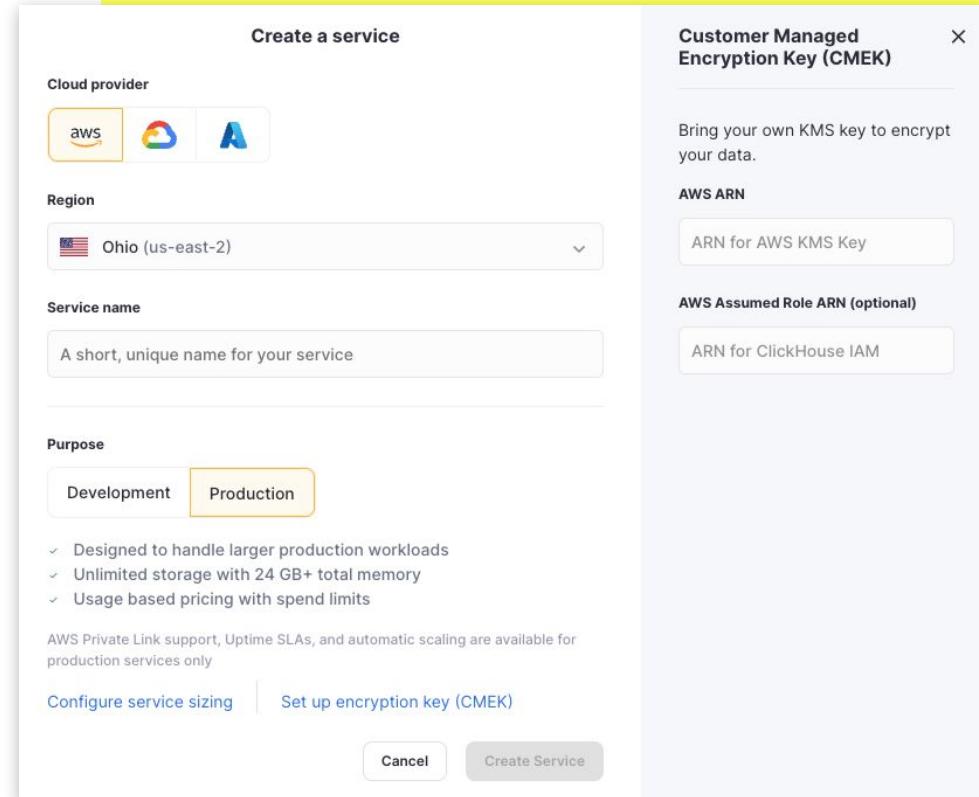


Contact support to request access



Customer Managed Encryption Key (CMEK) aka BYOK (Bring Your Own Key)

- Advanced protection of data at rest by allowing users to manage keys that control encryption/decryption of data
- Supports AWS KMS (Key Management Service), Cloud Key Management (GCP) - coming soon



Contact support to request access, supported for Production Services only



✓ ISO 27001

Request access to Compliance reports at

<https://trust.clickhouse.com/> -

SOC 2 Type II, ISO 27001,
GDPR, Pentesting and
Vulnerability reports.

ClickHouse Trust Center

Subscribe

Get access to this Trust Center

- ✓ Review sensitive security details
- ✓ Unlock documents
- ✓ Submit security questionnaires
- ✓ Ask for more information

Request access

Had access before? [Reclaim access](#)

Find Item...

Overview

Welcome to ClickHouse's Trust Center. Our commitment to data privacy and security is embedded in every part of our business. Use this Trust Center to learn about our security posture and request access to our security documentation.

Compliance

CCPA (CCPA), GDPR (GDPR), ISO (ISO 27001), PCI DSS (PCI DSS)

Documents

All Public Private

Bulk Download

Request Access to Private Documents

PCI DSS, Pentest Report, SOC 2 Report, Vulnerability Assessment Report, ISO 27001, Transfer Impact Assessment, Data Classification Policy, Information Security Policy



Analyst experience demo

- Demo Data Loading
 - File upload
 - s3 import
- Demo ‘running queries’ UI
 - view status of s3 import from data loading
 - view status of long-running query originated from CLI, cancel query
- Parameterized Queries

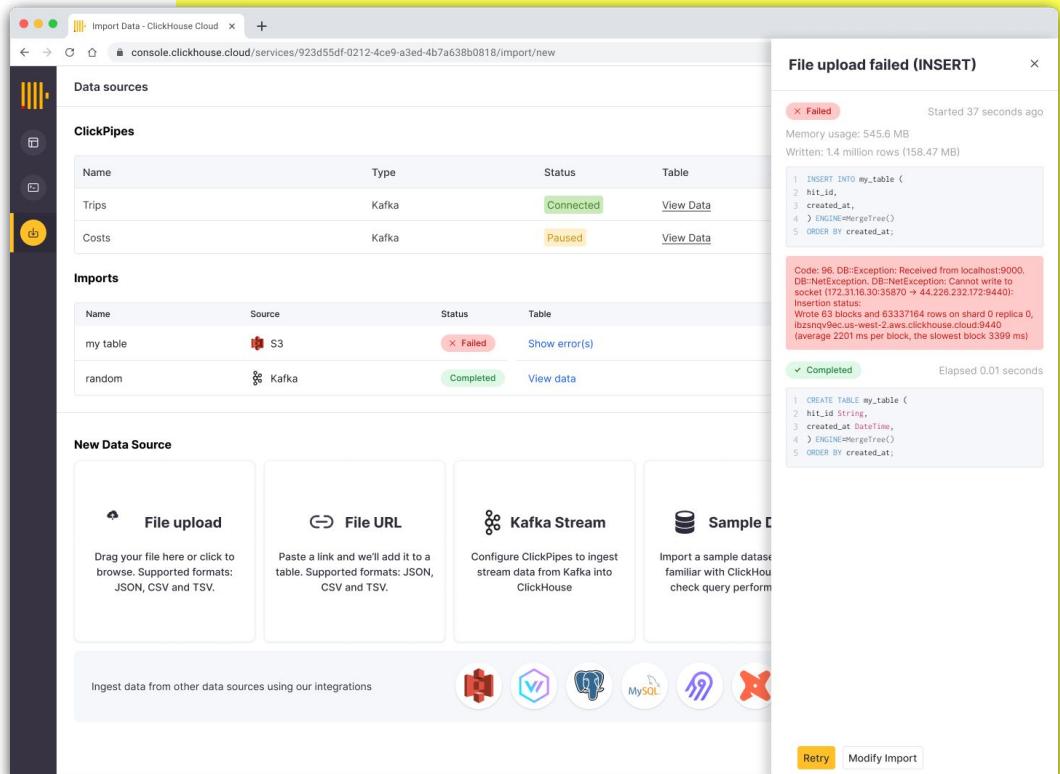


Feature Sneak Peek(s)



Monitoring for Data Imports

- View detailed progress and status heuristics for data import jobs
- Inspect, modify, and retry failed imports



The screenshot shows the ClickHouse Cloud Import Data interface. On the left, there's a sidebar with icons for Data sources, ClickPipes, and a search bar. The main area has three tabs: Data sources, ClickPipes, and Imports.

Data sources: Shows two Kafka-based data sources: "Trips" (Connected) and "Costs" (Paused).

ClickPipes: Shows a single ClickPipe named "my table" connected to S3, with a status of Failed. A "Show error(s)" button is available.

Imports: Shows two imports: "my table" from S3 (Failed) and "random" from Kafka (Completed). A "View data" button is available for the completed import.

New Data Source: A section with four options: File upload, File URL, Kafka Stream, and Sample DB. "File upload" is selected, showing instructions to drag files or browse for CSV, JSON, and TSV formats.

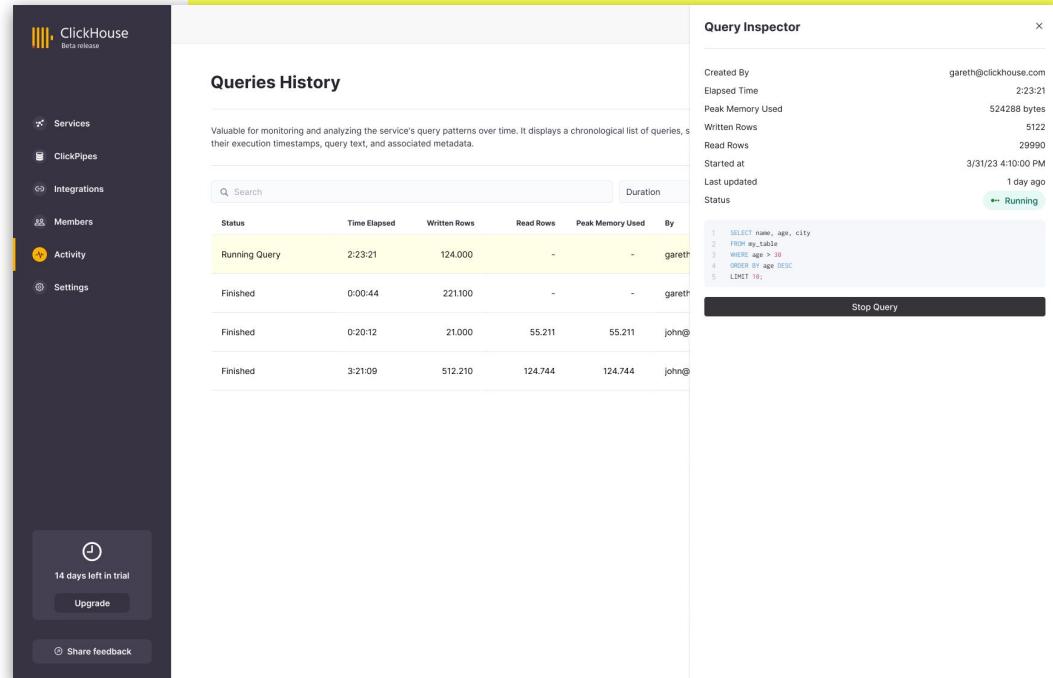
Logs and Metrics: On the right, there are two log sections. The top one for "File upload failed (INSERT)" shows a failed attempt with a memory usage of 545.6 MB, 1.4 million rows written, and the SQL query used. The bottom one for "CREATE TABLE my_table" shows a successful completion with an elapsed time of 0.01 seconds and the corresponding SQL query.

Integrations: At the bottom, it says "Ingest data from other data sources using our integrations" followed by icons for various databases and services like MySQL, PostgreSQL, and MongoDB.

Buttons: At the bottom right are "Retry" and "Modify Import" buttons.

Query Log Visualization

- View, filter, and inspect past and current queries



The screenshot shows the ClickHouse Query Inspector interface. On the left is a dark sidebar with navigation links: Services, ClickPipes, Integrations, Members, Activity (which is highlighted in yellow), and Settings. A trial upgrade notice is at the bottom of the sidebar.

The main area has two tabs: "Queries History" (selected) and "Query Inspector".

Queries History: This tab displays a table of query logs. The columns are: Status, Time Elapsed, Written Rows, Read Rows, Peak Memory Used, and By. The data is as follows:

Status	Time Elapsed	Written Rows	Read Rows	Peak Memory Used	By
Running Query	2:23:21	124.000	-	-	gareth@
Finished	0:00:44	221.100	-	-	gareth@
Finished	0:20:12	21.000	55.211	55.211	john@
Finished	3:21:09	512.210	124.744	124.744	john@

Query Inspector: This tab shows detailed information for the currently running query. It includes fields for Created By, Elapsed Time, Peak Memory Used, Written Rows, Read Rows, Started at, Last updated, and Status (Running). Below this is a code editor containing the SQL query:

```
1 SELECT name, age, city
2 FROM my_table
3 WHERE age > 30
4 ORDER BY age DESC
5 LIMIT 10;
```

A "Stop Query" button is located at the bottom right of the code editor.

SQL Console Access Control

- Configure SQL console access at user- and role-level
- Enables ‘passwordless’ access to ClickHouse Cloud instances for all users

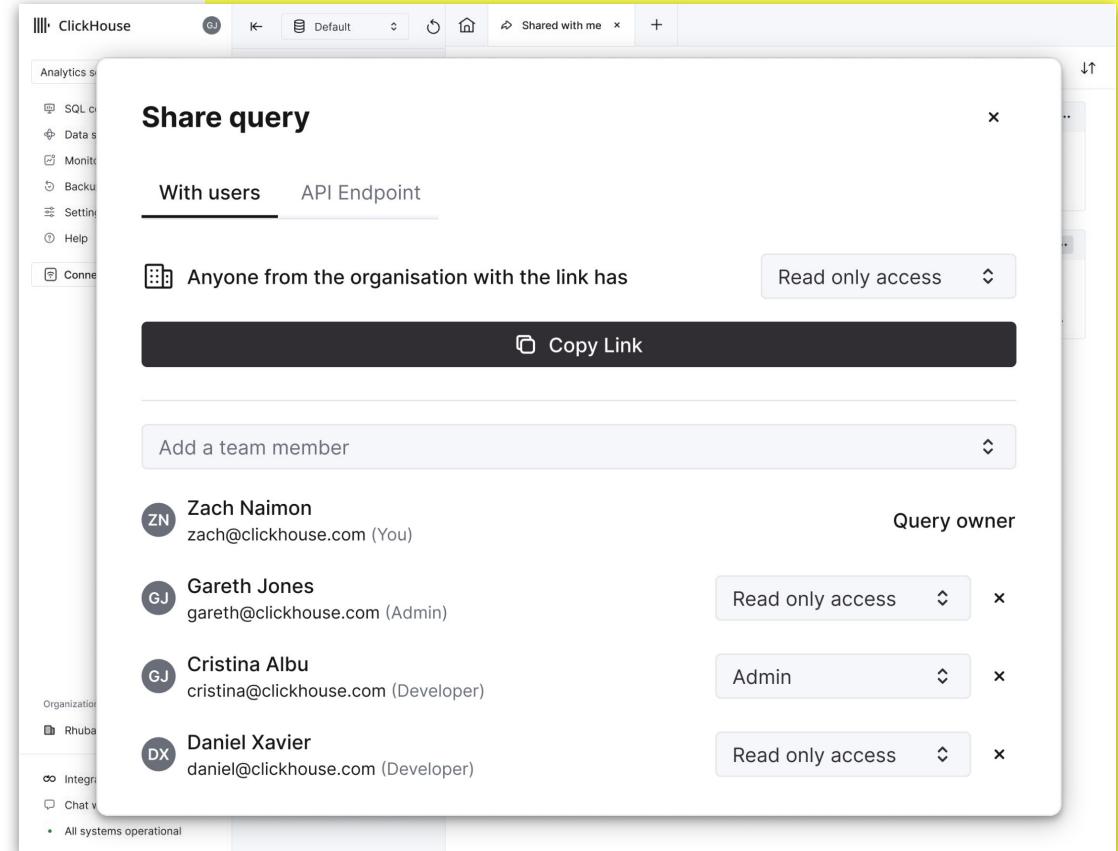
SQL console access

Provision SQL console access to members of your organization based on role. Roles without access will be prompted to manually enter credentials when opening the SQL console

 Admin 5 users	<div>Full access</div> <div>▼</div>
 Developer 13 users	<div>Read only</div> <div>▼</div>

Improved Query Sharing

- Save queries for private use or collaborate with team
- Introduces granular saved query access control



Query API Endpoints

- Expose REST API endpoints for saved queries, begin consuming query results in your application in just a few clicks
- Configure access and monitor request volume by API key(s)

Share query

x

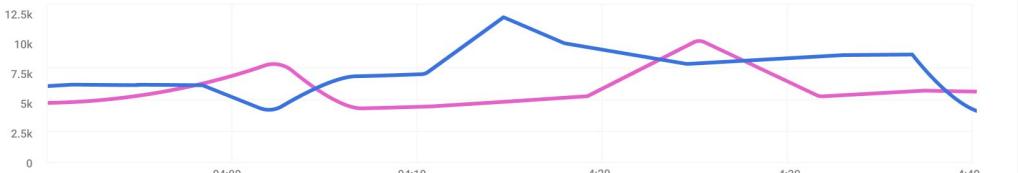
API endpoint

```
curl -sS --user '<key_id>:<key_secret>' \
https://api.clickhouse.cloud/query/4rnokq12?
param_foo=<value>&param_bar=<value>&format=CSVWithNames
```

...

Requests over time

Last hour



12.5k
10k
7.5k
5k
2.5k
0

04:00 04:10 04:20 04:30 04:40

● Gareth's API Key 104,579 requests ● HouseClick Application 89,162 requests

Total requests **1,412,743**

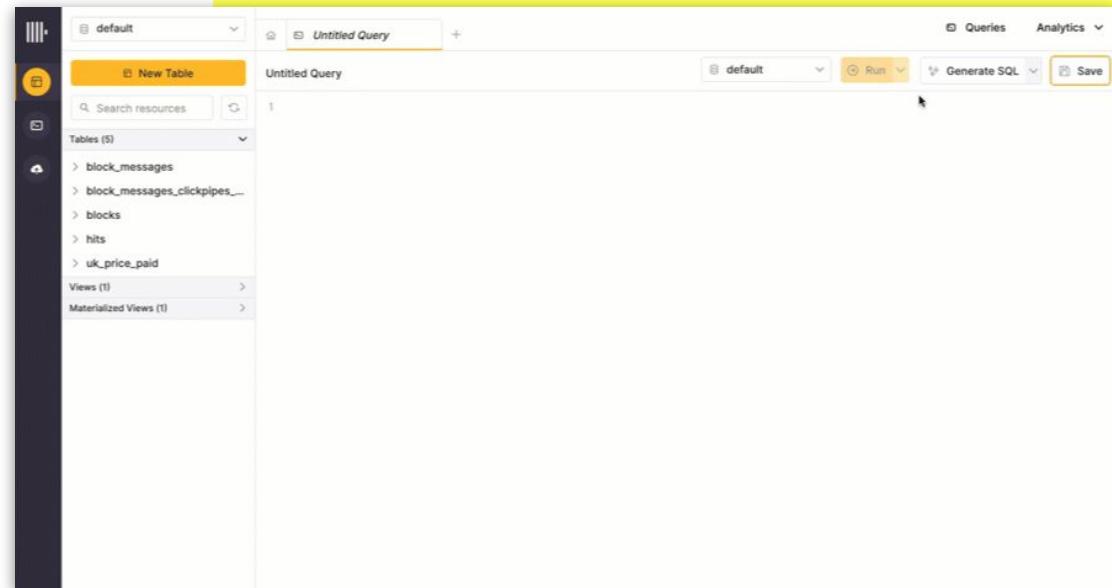
Last request **4 minutes ago**

You can now monitor how your API endpoint is used [here](#)

Disable endpoint

LLM-Enabled Query Generation

- Generate ClickHouse SQL queries from natural language prompts
- Fix errors in queries



02 Integrations

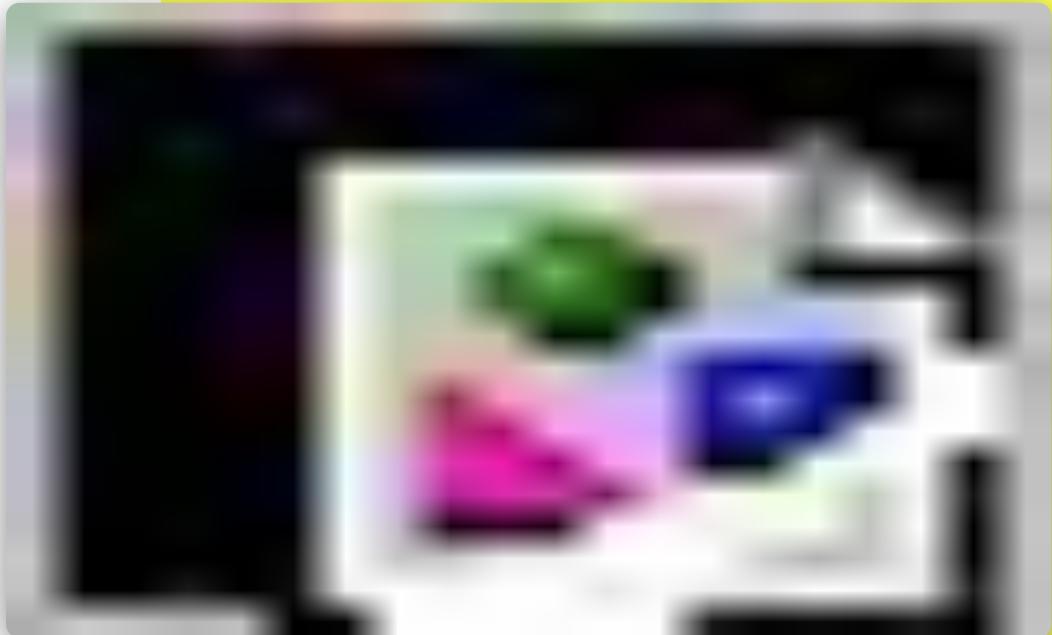
Ryadh



ClickPipes *noun*

/klɪk paɪps/

- Cloud-Native experience to ingest data from remote data sources to ClickHouse Cloud



<https://clickhouse.com/cloud/clickpipes>





MySQL Interface in Cloud

- Opt-in UI to enable the MySQL interface in ClickHouse Cloud
- Expands the Cloud BI tools and clients compatibility
- Tested for Looker Studio
- Currently in private preview

Connect to Ryadh's Demo

We created credentials to help you securely log into ClickHouse. If you do not know the default user's credentials, you can [reset your password](#).

You can use your connection string to connect to your new service. See our docs for more details around connecting [natively](#) or using [HTTPS](#).

Native HTTPS MySQL

By enabling [ClickHouse MySQL Interface](#), your service will expose the network port 3306 for the MySQL wire protocol.

Enable MySQL Interface

```
mysql -h usipq6vb8i.eu-west-1.aws.clickhouse.cloud -u mysql4usipq6vb8i -P 3306 --password
```

[Close](#) [Open SQL console](#)

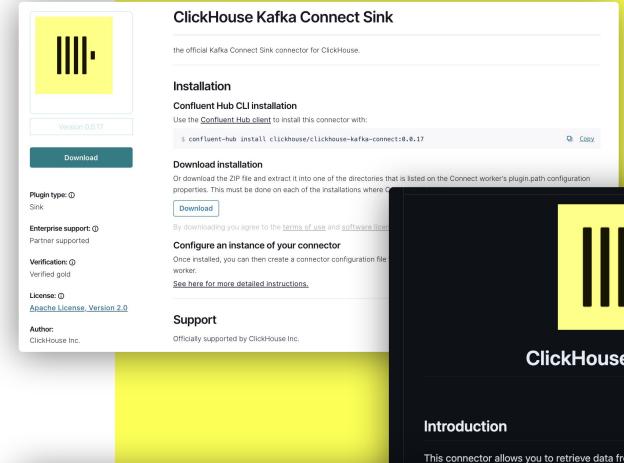
SUMMARY	
NAME	Ryadh's Demo
STATUS	Running
PLAN	Serverless
PROVIDER	aws





Also happening in integrations ...

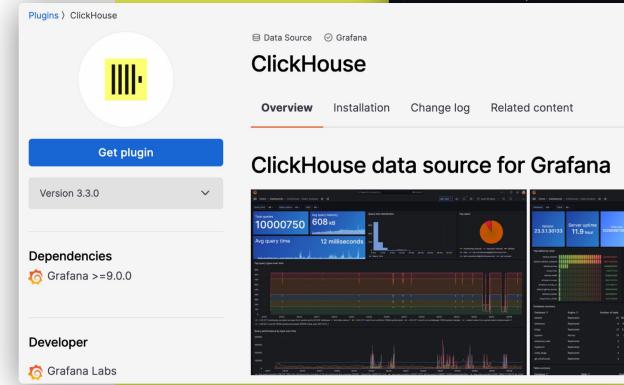
- The clickhouse-kafka-connect sink is now GA 🎉
- Our very own PowerBI Connector (beta) has landed 🚀
- In-progress: Grafana clickhouse-datasource v4 🎨
- JS client: Released a stable version for web-based environments 🌐
- Golang/Python/JS/Java: reliability and performance improvements 💪



The screenshot shows the ClickHouse Kafka Connect Sink page on the Confluent Hub. It includes a plugin card with the ClickHouse logo, a download button, and details about the plugin type (Sink), enterprise support (Partner supported), verification (Verified gold), and license (Apache License, Version 2.0). The author is ClickHouse Inc.



The screenshot shows the ClickHouse Connector for Power BI page. It features a large ClickHouse logo and bar charts, with a download button and a brief introduction: "This connector allows you to retrieve data from ClickHouse directly into Power BI for analysis and visualization." It also mentions "Query mode, allowing you to work with large datasets in real-time" and "DBC driver settings to meet your specific requirements."



The screenshot shows the ClickHouse data source for Grafana page. It includes a "Get plugin" button, dependency information ("Grafana >= 9.0.0"), developer information ("Grafana Labs"), and two preview screenshots of Grafana dashboards showing various metrics and data visualizations.



03

ClickHouse Core

Alexey

ClickHouse Core

ClickHouse Cloud uses separated storage & compute with **shared storage** - S3 and GCS and write-through cache on **local SSD** and RAM.

Is it scalable?

Yes

Is it fast?

Yes

But how? 🤔



Numbers That Were Easy To Get

How quickly can you insert a trillion records to ClickHouse Cloud?

- **1 hour 10 min: 250 million records/second**
(on 10 replicas - parallel data loading from s3 for one of the customers)

How quickly can you scan the data in ClickHouse Cloud?

- **191 GiB/sec**
(aggregation query on PyPi usage statistics)



Numbers That Were Hard To Get

Short queries latency

— optimized by adding cache pre-warming.

Example: loading 440 bn rows of Wikistat data with 4 connections and running concurrent SELECT queries for charts

Cloud, before/after, p50: **422/157** ms, p90: **750/294** ms, p99: **1250/756** ms.

ClickBench results

— optimized task distribution for reading.

14% improvement on average comparing to Beta.

First queries after idling

— optimized by adding compressed indices.

Example: a point query for Wikistat data: `SELECT * FROM wikistat WHERE path= 'ClickHouse'`

Cloud, Sep 2022: **9.9 s.** Jun 2023: **2.3 s.** Self-managed (baseline): **5.04 s.** (without compressed indices)

Inserts latency

Example: do 1000 small inserts sequentially into “hits” table from ClickBench dataset.

Cloud, Sep 2022: **100.3 s.** Jun 2023: **68.5 s.**



SharedMergeTree Table Engine

A new table engine to be used by default in ClickHouse Cloud, optimized specifically for dynamic scaling.

It is only available in ClickHouse Cloud and partner cloud providers.

<https://clickhouse.com/blog/clickhouse-cloud-boasts-performance-with-sharedmergetree-and-lightweight-updates>



Better scalability



Better performance of INSERTs



Cost savings



Faster server start-up



Less operational complexity



Lightweight Updates

```
SET apply_mutations_on_fly = 1;
```

Multi-stage update process:

1. The expression recorded and applied on the fly.
2. Patch-parts are calculated, and merged on the fly.
3. Updates are fully materialized.

Will be implemented for both **SharedMergeTree**
and **ReplicatedMergeTree!**

✓ Stage 1 is already available for SharedMergeTree.

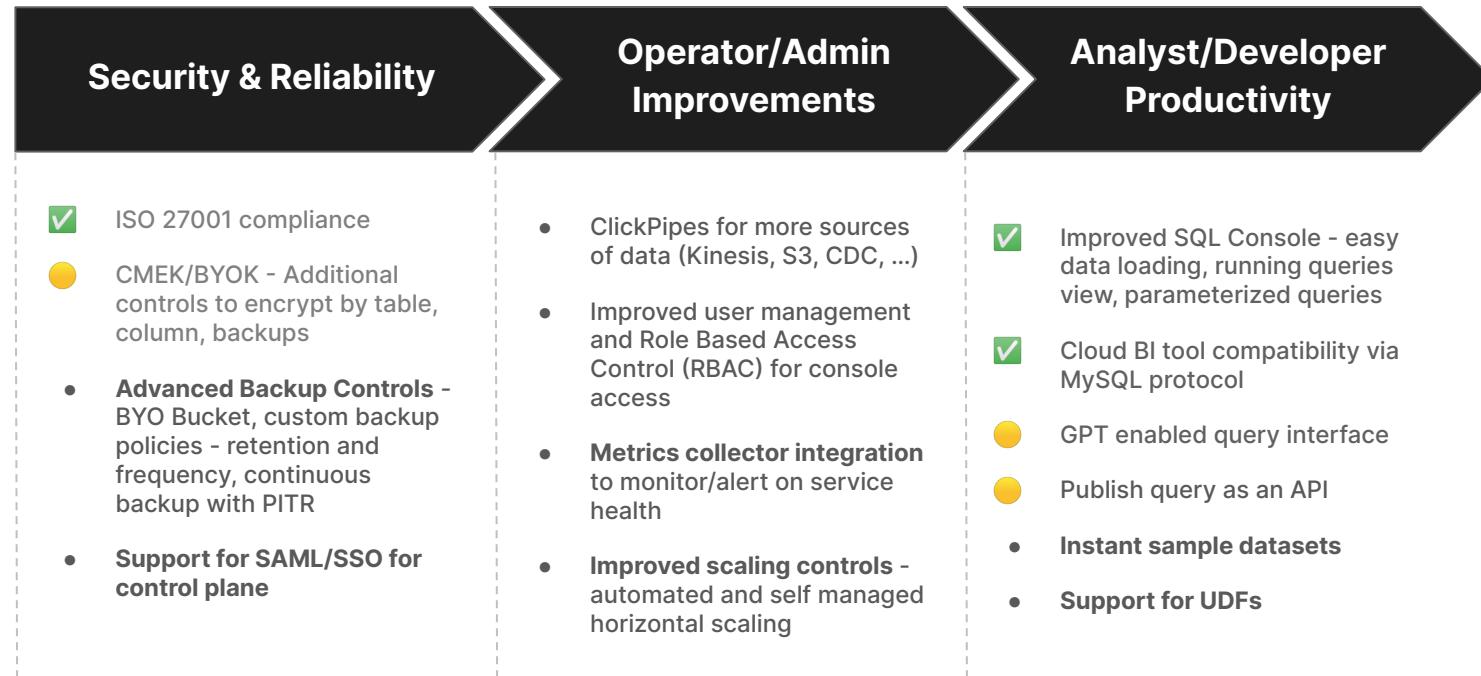


04

ClickHouse Cloud Roadmap

Tanya

ClickHouse Cloud - Roadmap



Cloud Provider Expansion

Microsoft Azure

Work in progress

Tentative ETA Q1 2024

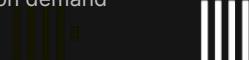


Registering interest in private preview
clickhouse.com/cloud/azure-waitlist

Additional regions



- Currently available
- In consideration based on demand



Q&A

Appendix

SharedMergeTree Table Engine

A new table engine to be used by default in ClickHouse Cloud, optimized specifically for dynamic scaling.

Advantages:

- Better scalability;
- Better performance of INSERTs;
- Cost savings;
- Faster server startup;
- Less operational complexity;

It is only available in ClickHouse Cloud and partner cloud providers.

<https://clickhouse.com/blog/clickhouse-cloud-boasts-performance-with-sharedmergetree-and-lightweight-updates>



The ClickHouse Difference

1

Very fast OLAP queries

Analytical queries - aggregations over large sets of data

2

Highly resource efficient

Disruptive data compression - 10-100x storage efficiency over alternatives

3

Easy to use

Self-service data onboarding from many sources (S3, Delta Lake, Iceberg, Hudi), analyst-friendly standard SQL syntax.

Deutsche Bank 

ClickHouse helps serve the Client Analytics platform for reporting, deep data analysis

- ✓ Platform for reporting and deep data analysis
- ✓ Advanced data science
- ✓ Provide clear view of client's activity and profitability

[Read the case study >](#)

