

ClickHouse

2023 Product Roadmap

Tanya Bragin

Jan 2023

 ClickHouse

Table of contents

01

ClickHouse Cloud

02

ClickHouse Integrations

03

ClickHouse Core

04

Questions



ClickHouse Core

ClickHouse Core 2022 Recap

Completed

- ✓ Make clickhouse-keeper production ready
- ✓ Support for backup and restore
- ✓ AArch64 hardware support
- ✓ Replicated database engine
- ✓ Type inference for data import
- ✓ Hudi, DeltaLake format support
- ✓ Implement Analyzer in ClickHouse

In progress

- PR Lightweight delete (experimental)
- PR Semi-structural Data (experimental)
- PR Support for transactions (experimental)
- PR SQL Compatibility improvements
- PR JOIN improvements



ClickHouse Core Roadmap 2023

<https://github.com/ClickHouse/ClickHouse/issues/44767>

- **Query analysis and optimization**

- Enable Analyzer by default
- Correlated subqueries (with decorrelation)
- Automatically select optimal JOIN algorithms

- **Formats and integrations**

- Support for Apache Iceberg
- Optimization of reading for Parquet

- **Security and access control**

- Bcrypt or other PBKDF
- Data masking in row-level security
- JWT support

- **Separation of storage and compute**

- Parallel replicas (production readiness)
- Parallel replicas with custom sharding key
- Optimization of high-cardinality operations

- **Data Storage**

- Asynchronous inserts by default
- Lightweight deletes (production readiness)
- Lightweight updates
- Transactions for ReplicatedMergeTree

- **Testing and hardening**

- Green CI: <https://aretestsgreenyet.com/>
- More complete fuzzing coverage
- Fully-static ClickHouse builds
- Embedded documentation



Faster Data Onboarding

Core recent features and roadmap

Asynchronous inserts

Ensure deduplication when using `async_insert`

- ✓ `async_insert` setting
- ✓ dedup when using `async_insert`

Datalakes integrations

Integrate natively with popular datalakes formats

- ✓ Hudi, Deltalake, Iceberg format support
- ✓ Optimized Parquet reading
(watch 24.4 release call)

Semi-structured data

Add support for natively storing semi-structured data

- PR JSON Object type
(experimental)
- PR SQL/JSON support



More flexible analytics

Core recent features and roadmap

Analyzer by default

Enabling Analyzer by default opens many possibilities in ClickHouse

- ✓ Implement Analyzer in ClickHouse
- PR Enable Analyzer by default
- PR Simplify existing code
 - Implement cost based optimizer

Extended JOIN support

Faster JOINS for more use cases; automatic JOIN algorithm selection

- ✓ New JOIN algorithms
 - Dynamic JOIN algorithm selection and JOIN rewrite (dep. Analyzer)
 - TPC-H and TPC-DS benchmarks

Indexing / Search

More functionality and seamless migrations / ecosystem integrations

- ✓ Inverted indices (experimental)
- PR More vector search functions
- PR More Standard SQL support



Extended data management capabilities

Core recent features and roadmap

Lightweight operations

Delete data frequently without impacting performance

- ✓ Lightweight DELETES
(production-ready in 23.3)
- PR Lightweight UPDATES

Transactions

Bring ACID properties to use cases important to ClickHouse users

- PR BEGIN TRANSACTION,
COMMIT and ROLLBACK
statements (experimental)

Query cache

Speed up repeated queries with built-in query cache

- ✓ Query cache (experimental)
- PR Query cache improvements



ClickHouse Integrations

Ecosystem Integrations status and roadmap

	Available	Recent/Next
Data Ingestion / Orchestration	<ul style="list-style-type: none">✓ S3✓ Kafka✓ DBT✓ Airbyte✓ Vector (community)✓ PostgreSQL✓ MySQL✓ Decodable (partner)✓ Metaplane (partner)	<ul style="list-style-type: none">🔥 Kafka Connect (beta)🟡 Confluent Cloud (partner)🟠 RedPanda Cloud (partner)✓ OpenTelemetry (community)✓ FluentBit (community)🔥 Apache Beam (community)🟠 Google Dataflow (partner)🟠 Azure Event Hub (partner)✓ EMQX (partner)
Data Visualization	<ul style="list-style-type: none">✓ Superset / Preset✓ Deepnote (partner)✓ HEX (partner)✓ Mode (partner)🔥 Grafana (partner)🔥 Metabase	<ul style="list-style-type: none">✓ Tableau (community)🟡 GCP Looker (partner)🟠 Azure Power BI (partner)🟠 AWS QuickSight
Language Clients	<ul style="list-style-type: none">🔥 Go🔥 Python🔥 Node.js✓ Java✓ C# (community)	<ul style="list-style-type: none">✓ C# (community)🟠 Ruby (community)🟠 Rust (community)
SQL Clients	<ul style="list-style-type: none">✓ DBeaver✓ Datagrip	

Legend

- ✓ Completed
- 🟡 In development
- 🟠 Coming Next / Evaluating
- 🔥 Recently improved

75+ integrations listed on

<https://clickhouse.com/docs/en/integrations>



ClickHouse Cloud

ClickHouse Cloud - from 0 to 1 in under a year



Private preview

May -July 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

Public Beta

Oct 2022

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

General Availability

Dec 2022

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



Administration

Cloud recent features and roadmap

Security / Privacy

Continue to address advanced security and privacy needs

- ✓ AWS IAM roles support
- PR BYOK for data encryption
- PR ISO 27001 compliance

Programmability

Continue to address advanced security and privacy needs

- ✓ API for control plane orch
 - API for database operations
 - Terraform provider

Operations

Additional capabilities to make it easier to operate Cloud

- PR Incremental backups
- PR Centralised user management
 - Metrics export



Analyst productivity

Cloud recent features and roadmap

Kafka ecosystem

Data onboarding platform built into ClickHouse Cloud

- ✓ Kafka Table Engine (CH)
- ✓ KafkaConnect support
- ✓ Confluent Cloud support
- ✓ Amazon MSK support
- ✓ RedPanda support
 - Kafka data onboarding UI

Adhoc data loading

Continue to address advanced security and privacy needs

- PR UI for create/edit schema
- PR UI for data loading from S3
 - Instant sample datasets

Collaboration

Support analyst and developer productivity

- Query performance UI
- Publish query as an API
- Query sharing/private queries
- GPT enabled SQL queries



Questions?