ChistaDATA Inc. Introduction

Shiv Iyer
Founder and CEO of ChistaDATA Inc.

*An enterprise-grade 24*7 Consultative Support and Managed Services Provider for Open Source/Free ClickHouse (both on-premises and Cloud)*

How is Real-Time Analytics Different from Traditional OLAP?

ChistaDATA Inc.

- **Data Latency:** Real-time analytics provides instant access to up-to-date data, while traditional OLAP may have a lag between data updates and analysis. In real-time analytics, data is processed and analyzed in near real-time, so insights can be quickly obtained and acted upon.
- **Data Volume:** Real-time analytics is typically designed to handle large volumes of streaming data, such as data generated by IoT devices, social media, or web traffic. Traditional OLAP is typically designed for batch processing of large volumes of historical data.
- **Query Performance:** Real-time analytics requires fast query response times to keep up with the speed of incoming data. In contrast, traditional OLAP may prioritize query accuracy and completeness over query speed.

- **Data Models:** Real-time analytics often uses flexible, dynamic data models that can adapt to changing data streams and evolving business needs. Traditional OLAP uses static, pre-defined data models that can limit the types of analysis that can be performed.
- **Use Cases:** Real-time analytics is often used for real-time decision-making, monitoring, and alerting, such as fraud detection or supply chain optimization. Traditional OLAP is often used for historical analysis and strategic decision-making, such as trend analysis or forecasting.

ClickHouse – ColumnStore for Real-Time Analytics

- ClickHouse is an open-source, column-oriented database management system that was initially developed by the Russian company Yandex in 2016 for the analytical processing of large-scale data sets in real-time. ClickHouse was designed to be highly scalable, efficient, and flexible, with the ability to process petabytes of data and support real-time data processing and analysis. The system was built using C++ and supports a wide range of data formats and protocols.
- The development of ClickHouse was motivated by Yandex's need for a highly performant database system that could process and analyze the massive amounts of data generated by its web search and advertising services.

ChistaDATA – Partner for building WebScale Real-Time Analytics on ClickHouse

- ChistaDATA provides full-stack ClickHouse Optimization. We deliver elite-class Consultative Support (24*7) and Managed Services for both on-premises ClickHouse infrastructure and Serverless/Cloud/ClickHouse DBaaS operations.
- ChistaDATA Server for ClickHouse (and all tools essential for Data Ops. @ Scale) will be Open Source (100% GPL forever) and free. We are committed to helping corporations build Open Source ColumnStore for high-performance Data Analytics.
- Global Team available 24*7 for ClickHouse Consultative Support and Managed Services.

Why do we recommend ClickHouse over many other columnar database systems?

- **Compact data storage** – Ten billion UInt8-type values should exactly consume 10GB uncompressed to efficiently use the available CPU. Optimal storage even when uncompressed benefits performance and resource management. ClickHouse is built is store data efficiently without any garbage.
- **CPU efficient** – Whenever possible, ClickHouse operations are dispatched on arrays, rather than on individual values. This is called “vectorized query execution,” and it helps lower the cost of actual data processing.
- **Massively Parallel Processing** – ClickHouse is capable of Massively Parallel Processing very large/complex SQL(s) optimally and cost-efficiently

Why do we recommend ClickHouse over many other columnar database systems?

- **Data compression** – ClickHouse supports two kinds of compression LZ4 and ZSTD. LZ4 is faster than ZSTD but the compression ratio is smaller.ZSTD is faster and compresses better than traditional Zlib but slower than LZ4. We recommend customers LZ4 when I/O is fast enough so decompression speed will become a bottleneck. When using super ultra-fast disk subsystems you have an option to specify “none” compression. ZSTD is recommended when I/O is the bottleneck in queries with large range scans.
- **Can store data in disk** – The columnar database systems like SAP HANA and Google PowerDrill can only work in the RAM

Why do we recommend ClickHouse over many other columnar database systems?

- **Built for web-scale data analytics** – ClickHouse supports sharding and distributed processing. This makes ClickHouse the most preferred columnar database system for web-scale. Each shard in ClickHouse can be a group of replicas addressing maximum reliability and fault tolerance.
- **ClickHouse support Primary Key** – ClickHouse permits real-time data updates with a primary key (there will be no locking when adding data). Data is sorted incrementally using the merge tree to perform queries on the range of primary key values.
- **Supports data replication** – ClickHouse supports asynchronous multi-master and master-slave replication.

Why do we recommend ClickHouse over many other columnar database systems?

- **Built for statistical analysis and supporting partial aggregation** – ClickHouse is a statistical query analysis-ready columnar database store supporting aggregate functions for approximated calculation of the number of various values, medians, and quantiles. ClickHouse supports aggregation for a limited number of random keys, instead of for all the keys. You can query on a part (sample) of data and generate approximate results reducing disk I/O operations considerably.
- **Supports SQL** – ClickHouse supports SQL, Subqueries are supported in FROM, IN, and JOIN clauses, as well as scalar subqueries. Dependent subqueries are not supported.

Why is ClickHouse recommended for a time-series Database?

- **Column-oriented storage:** ClickHouse uses a column-oriented storage model, which means that data is stored by columns rather than by rows. This allows for efficient compression and faster data retrieval, especially for time-series data, where the data is often read in time-based chunks.
- **Advanced analytical functions:** ClickHouse supports advanced analytical functions such as window functions, aggregate functions, and SQL-based data filtering, which are useful for time-series data analysis. This allows users to perform complex queries on large data sets quickly and efficiently.
- ClickHouse Supports High Availability through Replication.

Why is ClickHouse recommended for a time-series Database?

- **Real-time query performance:** ClickHouse is designed to handle high write and read performance, making it suitable for real-time data analysis. It can handle millions of writes per second and return results in milliseconds, even on large datasets.
- **Scalability:** ClickHouse is a distributed system, which means that it can scale horizontally by adding more servers. This allows it to handle very large data sets and handle high write and read loads.
- **Compression:** ClickHouse supports advanced compression techniques, which can significantly reduce the size of the data stored on disk, making it more cost-efficient for storing large data sets.

Why migrate from Hadoop to ClickHouse for Real-Time Analytics?

- **Performance:** Hadoop is designed for batch processing and data warehousing, which can result in longer query times. It's not optimized for high-performance analytical queries, which are required for real-time analytics.
- **Latency:** Hadoop's batch processing approach means that data is processed in large chunks, which can result in significant latency. This makes it difficult to provide near real-time analytics.
- **Scalability:** Hadoop can scale horizontally, but it requires more resources and management than other technologies.

Why migrate from Hadoop to ClickHouse for Real-Time Analytics?

- **Complexity:** Hadoop requires a significant amount of configuration and management, which can be complex and time-consuming. It also requires a knowledge of programming languages such as Java or Python to work with the data.
- **Real-time streaming:** Hadoop is not well-suited for real-time streaming data, which is becoming increasingly important for real-time analytics use cases.
- **Cost:** Hadoop can be expensive, as it requires expensive commercial licenses for some of its components, such as for HDFS and YARN.

How can ChistaDATA help you build web-scale real-time streaming data analytics using ClickHouse?

- **Consulting** – We are experts in building optimal, scalable (horizontally and vertically), highly available and fault-tolerant ClickHouse powered streaming data analytics platforms for planet-scale internet / mobile properties and the Internet of Things (IoT). Our elite-class consultants work very closely with your business and technology teams to build custom columnar database analytics solutions using ClickHouse.

How can ChistaDATA help you build web-scale real-time streaming data analytics using ClickHouse?

- **Database Architect services** – We architect, engineer and deploy data analytics platforms for you. We will take care of your data analytics ecosystem so that you can focus on business.
- **ClickHouse Enterprise Support** – We have 24*7 enterprise-class support available for ClickHouse. Our support team will review and deliver guidance for your data analytics platforms architecture, SQL engineering, performance optimization, scalability, high availability and reliability.

ChistaDATA Cloud for ClickHouse DBaaS

- ChistaDATA Cloud Infrastructure for ClickHouse is a Columnar Database Service built for performance, scalability and reliability, which is operationally simple and cost-efficient. Cloud Infrastructure for ClickHouse is fully compatible with Standard ClickHouse Server, so you can migrate existing applications and tools to run without requiring any modifications.
- The ChistaDATA Cloud Services team will assist you in Capacity Planning and Sizing so you can scale compute, memory and storage (SSD only) resources enabling your ClickHouse deployment up or down

How is ChistaDATA Cloud different?

- **Autonomous and Driverless Database Infrastructure for Columnar Analytics:**
  - Automatic Provisioning of Optimal and Reliable ClickHouse Operations
  - Automatic Configuration of ClickHouse for Performance and Scalability
- Fully Autonomous System Failure Detection and Repair
- **Secured ClickHouse Database Infrastructure:**
  - Fully Autonomous ClickHouse Bug Fixing and Patching
  - Customized ClickHouse Infrastructure for Data Privacy and Query (both successful and unsuccessful) Audit

How is ChistaDATA Cloud different?

- Fully Autonomous ClickHouse DR – Maximum Reliability Architecture for Zero Data Loss
- Fully Autonomous ClickHouse Replication Services
- Fully Autonomous ClickHouse Partitioning (both vertical and horizontal)
- AI-Based ClickHouse Observability and Monitoring Platform
- Advanced SQL IDE for Data Analytics Developer Success
- Fully Autonomous Data Archiving Toolkit for MySQL, MariaDB and PostgreSQL

ChistaDATA Cloud – Why building Data Analytics on ChistaDATA Cloud is equally recommended for both startups and large corporations?

- **Flexible subscription plans:**
  - Flexible Serverless ClickHouse Infrastructure available for Development, Build/Release Engineering and Production environment
  - ChistaDATA Cloud provides ClickHouse infrastructure for both compute and storage ecosystems
  - Elastic: You can always upgrade/downgrade your subscription plan without any operational challenges

ChistaDATA Cloud – Why building Data Analytics on ChistaDATA Cloud is equally recommended for both startups and large corporations?

- **Just bring your Data to ChistaDATA Cloud – Driverless and fully autonomous Database Infrastructure for ClickHouse:**
  - Advanced ClickHouse SQL development IDE provided for developer success and Time-To-Market
  - **Are you already on any other DBaaS like Amazon Aurora (PostgreSQL and MySQL), Amazon RedShift and Google CloudSQL?**
    - Larger OLTP Database Systems are extremely expensive operationally, with frequent performance outages
    - ChistaDATA Cloud Archiving Toolkit for MySQL, MariaDB, PostgreSQL and other DBaaS will help you in building lean OLTP databases for both on-premises and serverless infrastructure
    - ChistaDATA Proxy helps in transparent load-balancing between OLTP Database Infrastructure and ClickHouse

ChistaDATA Cloud – Why building Data Analytics on ChistaDATA Cloud is equally recommended for both startups and large corporations?

- AI-Based Troubleshooting for ClickHouse Infrastructure Maintenance Operations
- ChistaDATA Audit Logs for detailed forensics of ClickHouse Server Operations
- Point-in-Time-Recovery for ClickHouse Server
- ChistaDATA ClickHouse Proxy as a Service:
  - READ-WRITE Query Distribution / Horizontal Scalability
  - Query Caching
  - ClickHouse Firewall Implementation with ChistaDATA Proxy

A partial list of customers from the ChistaDATA Portfolio

- Blue Dart
- PayPal
- Morgan Stanley
- Sony
- Nintendo
- Netflix
- Carlsberg
- Burberry
- PRADA
- Cambridge Investment Research
- VISA
- Unilever
- Garmin
- National Geographic

Contact ChistaDATA for building Real-Time Analytics at WebScale

ChistaDATA(California)

ChistaDATA Inc.,
340 S LEMON AVE #9718
WALNUT 91789 CA, US

ChistaDATA(Houston)

ChistaDATA Inc., 1321 Upland Dr. PMB 19322, Houston,
TX, 77043, US

ChistaDATA Sales Contact

- (844)395-5717
- +1(209)314-2364 - FAX
- Info(at)chistadata.com