

# COST-EFFECTIVE & CLOUD AGONISTIC LOGGING

(With – Clickhouse & Grafana)



Pushpender Kumar



# Pushpender Kumar


- Born and raised in Bijnor and Moved to Delhi to stay ahead in the race of life. 🏃💨
- Currently working at OLX India as DevOps Architect
- Cloud & DevOps Specialist | 10+ years in AWS, Kubernetes & automation | Cut log storage costs by 50% with Clickhouse.
- Unpopular Opinion – Docker containers are lightweight, but their issues can be heavy. 🐳




# AGENDA

- Overview of Logging Tools
  - Challenges in Traditional Logging Tools
  - Cloud Agnostic Approaches
  - ClickHouse Architecture
  - Comparing Clickhouse & Grafana with ELK Stack
  - Performance & Cost Insights
  - Conclusion and Q&A
- 


# THE IMPORTANCE OF LOGGING AND MONITORING

- Ensures application reliability and performance.
  - It's critical for identifying and resolving issues promptly.
  - Facilitates data-driven decision-making.
- 


# CHALLENGES IN TRADITIONAL LOGGING TOOLS

- Common tools like the ELK Stack (Elasticsearch, Logstash, Kibana) have been popular, but they come with challenges.
  - Key Challenges:
    - High operational costs (storage, processing, maintenance).
    - Scaling difficulties as data volume grows.
- 

# CLOUD AGNOSTICISM

- The ability to operate across different cloud platforms without vendor lock-in.
  - Clickhouse & Grafana Advantages:
    - Multi-cloud Deployment: Can be deployed on any cloud provider or on-premises.
    - Flexibility: Freedom to choose the best cloud provider based on cost and performance needs.
    - Interoperability: Integrates with various cloud services, supporting hybrid and multi-cloud environments.
- 

# CLICKHOUSE & GRAFANA

- What is Clickhouse?
    - An open-source columnar database, optimized for analytical queries and high-throughput logging.
    - Designed to handle large volumes of data efficiently.
  - What is Grafana?
    - A popular open-source platform for monitoring and observability, supporting various data sources.
    - Provides powerful visualization and alerting capabilities.
- 

# WHY CLICKHOUSE IS AN EXCELLENT FIT FOR LOGGING


Clickhouse, a column-oriented DBMS, is well suited for OLAP workloads. Logging workloads share similar attributes with OLAP workloads, such as being read-heavy, having large batches of inserts but rare mutations, and containing mostly immutable data.





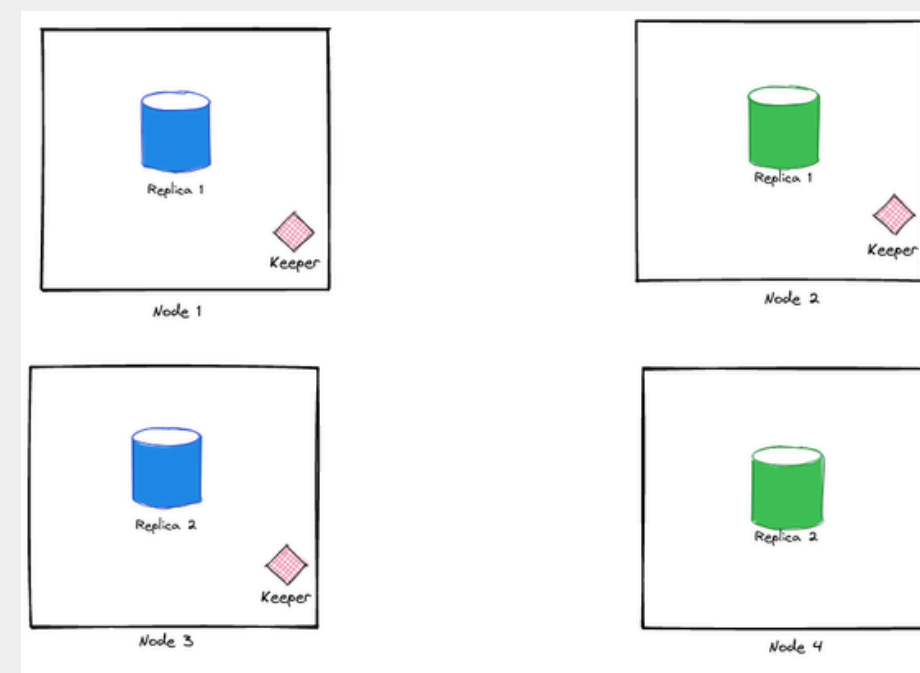
# CLICKHOUSE'S COMPRESSION CODECS

Clickhouse supports various compression codecs out of the box, including LZ4, Brotli, and ZSTD, at the block level. These codecs can reduce the amount of data that needs to be read from the disk, enhancing performance. Additionally, Clickhouse supports dictionary compression, which is useful for columns with low cardinalities.



# SCALABILITY WITH CLICKHOUSE

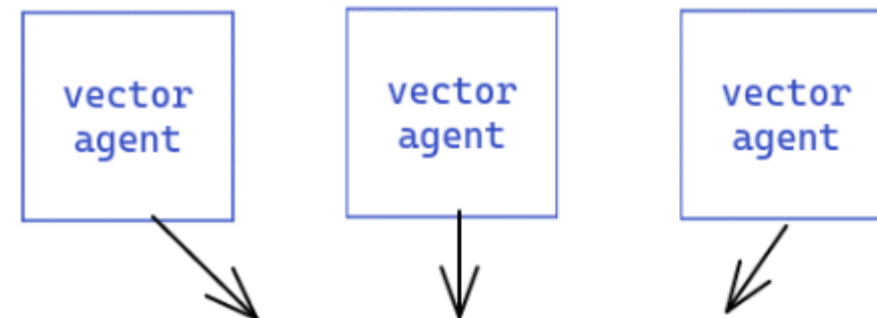
Clickhouse is relatively easy to scale and comes bundled with clickhouse-keeper, which makes it possible to set up a distributed Clickhouse cluster with multiple nodes replicating data with each other.



# CLICKHOUSE ARCHITECTURE

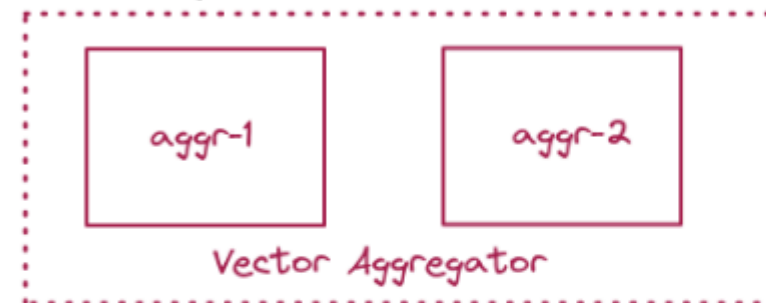
Stage 1:

Collect logs from  
file/stdout and parse them



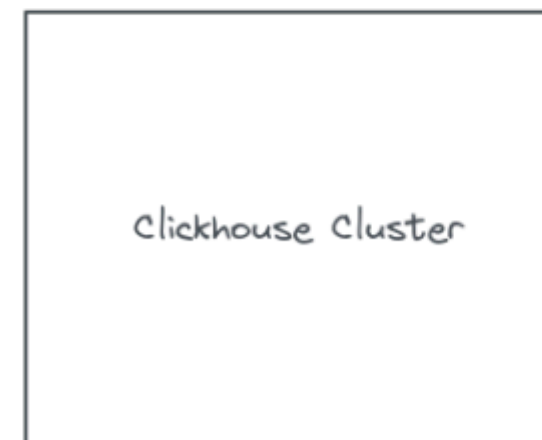
Stage 2:

Buffer the incoming logs in  
disk buffers and batch the  
writes to Clickhouse



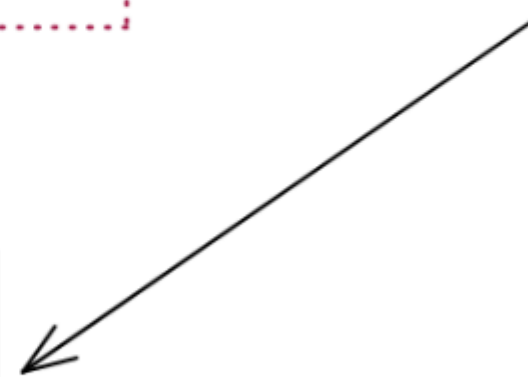
Stage 3:

Store the logs in different  
tables. Use distributed tables to  
replicate across nodes.



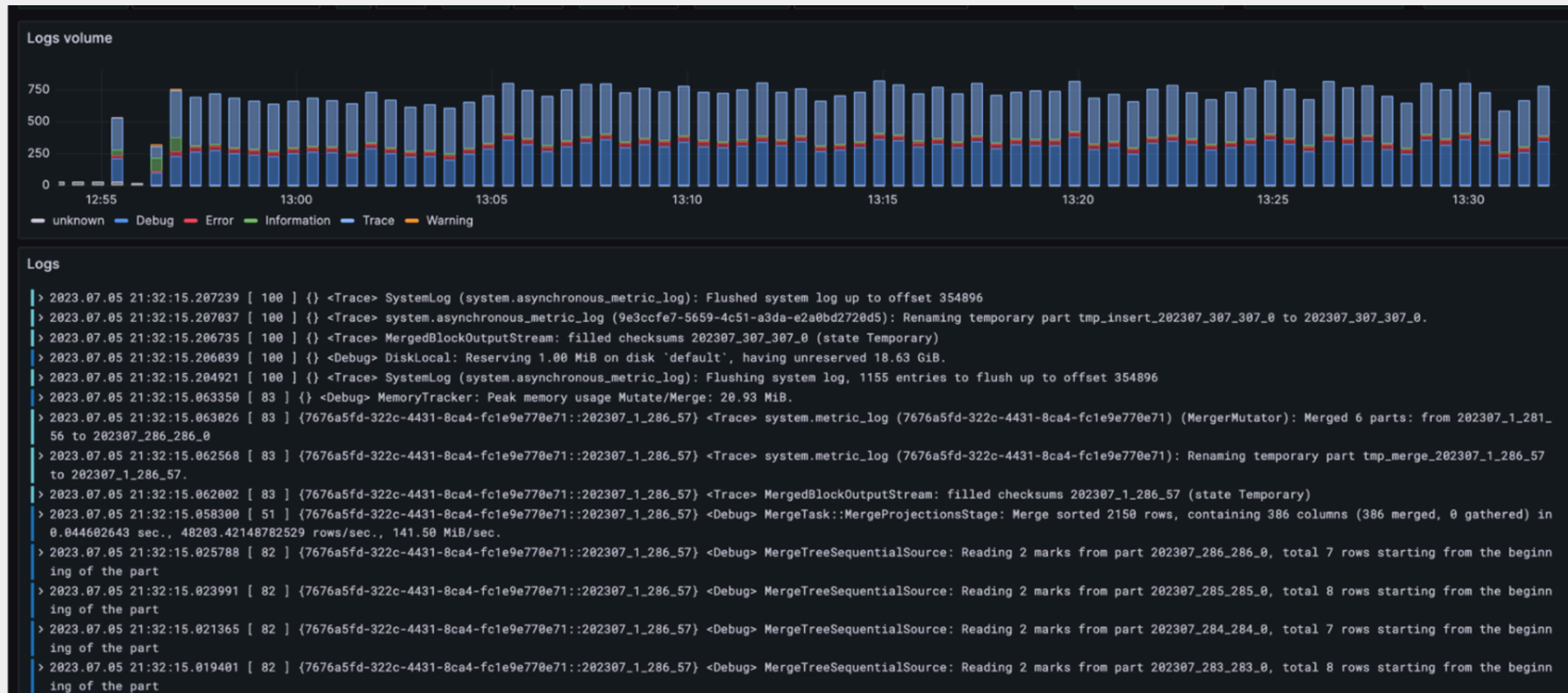
**Grafana**

Use SQL + query templates  
to create dashboards and  
visualize log data




# GRAFANA

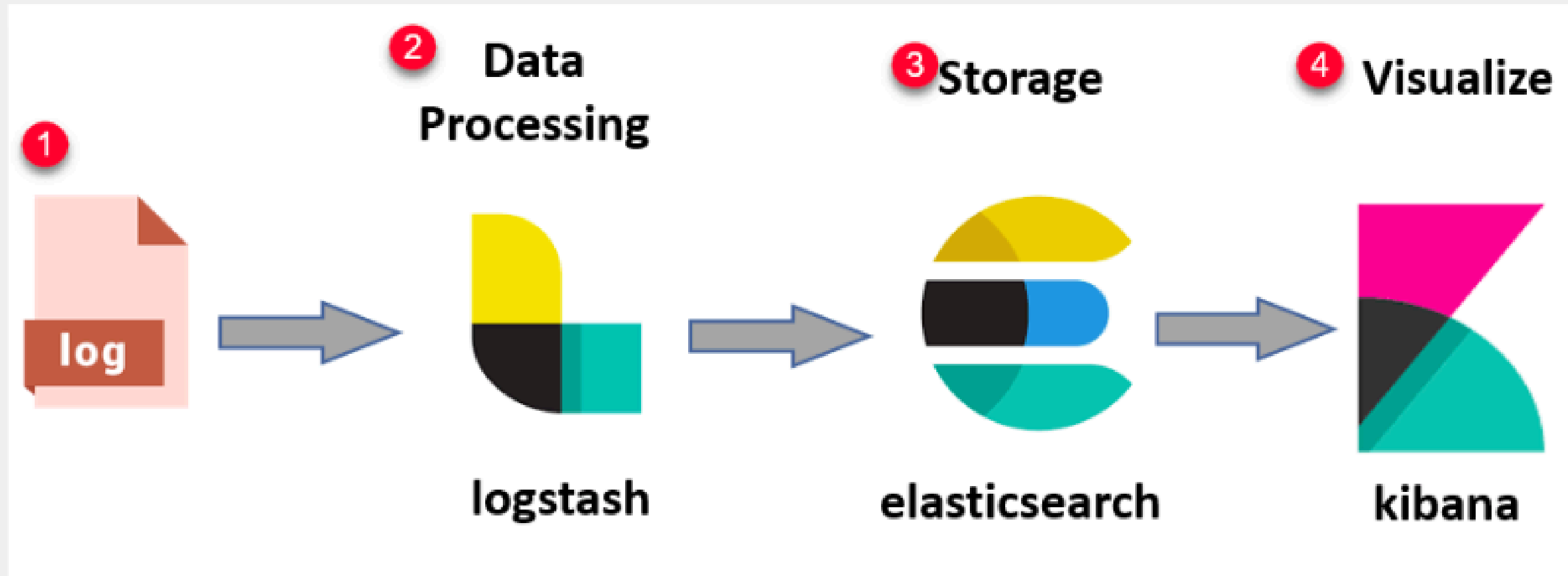
Grafana is a multi-platform open source analytics and interactive visualization web application. It can produce charts, graphs, and alerts for the web when connected to supported data sources



# COST EFFICIENCY WITH CLICKHOUSE & GRAFANA

- **Storage Optimization:** Clickhouse's columnar storage reduces disk usage, significantly lowering storage costs.
  - **Query Performance:** Efficient data retrieval with minimized computational overhead.
  - **Open Source:** No licensing fees, reducing overall costs.
  - **Horizontal Scalability:** Easily scales with your data, avoiding expensive vertical scaling.
- 

# ELK ARCHITECTURE



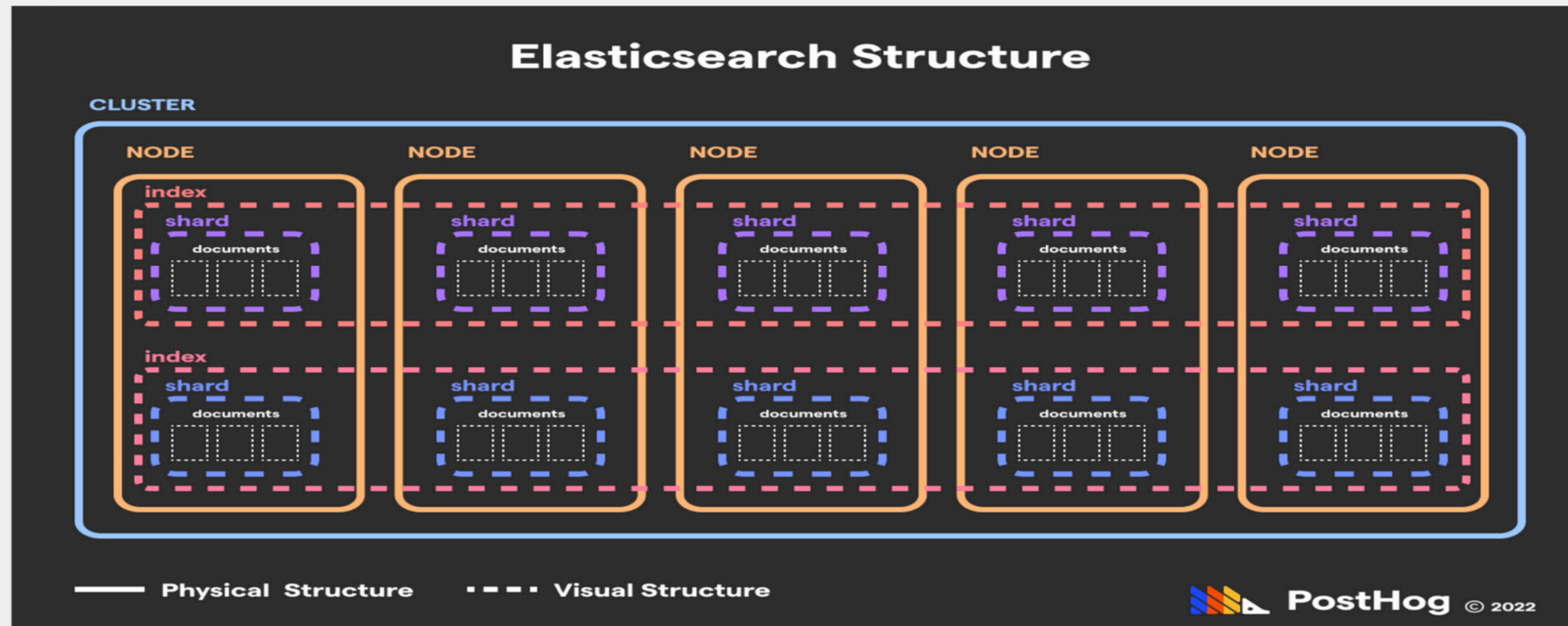
# DATA AND INFRASTRUCTURE

The biggest, defining difference between Elasticsearch and ClickHouse is their respective techniques for storing and organizing data.

- ClickHouse is a columnar database; it stores data in a table, just with an inverted structure (in disk) relative to a traditional MySQL or PostgreSQL table. ClickHouse's columnar data store simplifies aggregating data.
- Elasticsearch isn't columnar – it isn't even a table-based database. It stores data as documents, grouping sets of documents into shards, which are part of physical and virtual collections respectively known as nodes and indices



# ELASTICSEARCH STRUCTURE





# CLICKHOUSE STRUCTURE

## Row-based versus Columnar Database in Storage

### Row-Based Disk Storage

	Name	Industry	Age	Spines
	Percy	Construction	3	2,483
	Dylan	Construction	5	1,902
	Mary	Healthcare	4	2,190
	Dennis	Healthcare	8	1,828
	Megan	Space	5	1,231

Direction of Storage Scan

### Column-Based Disk Storage

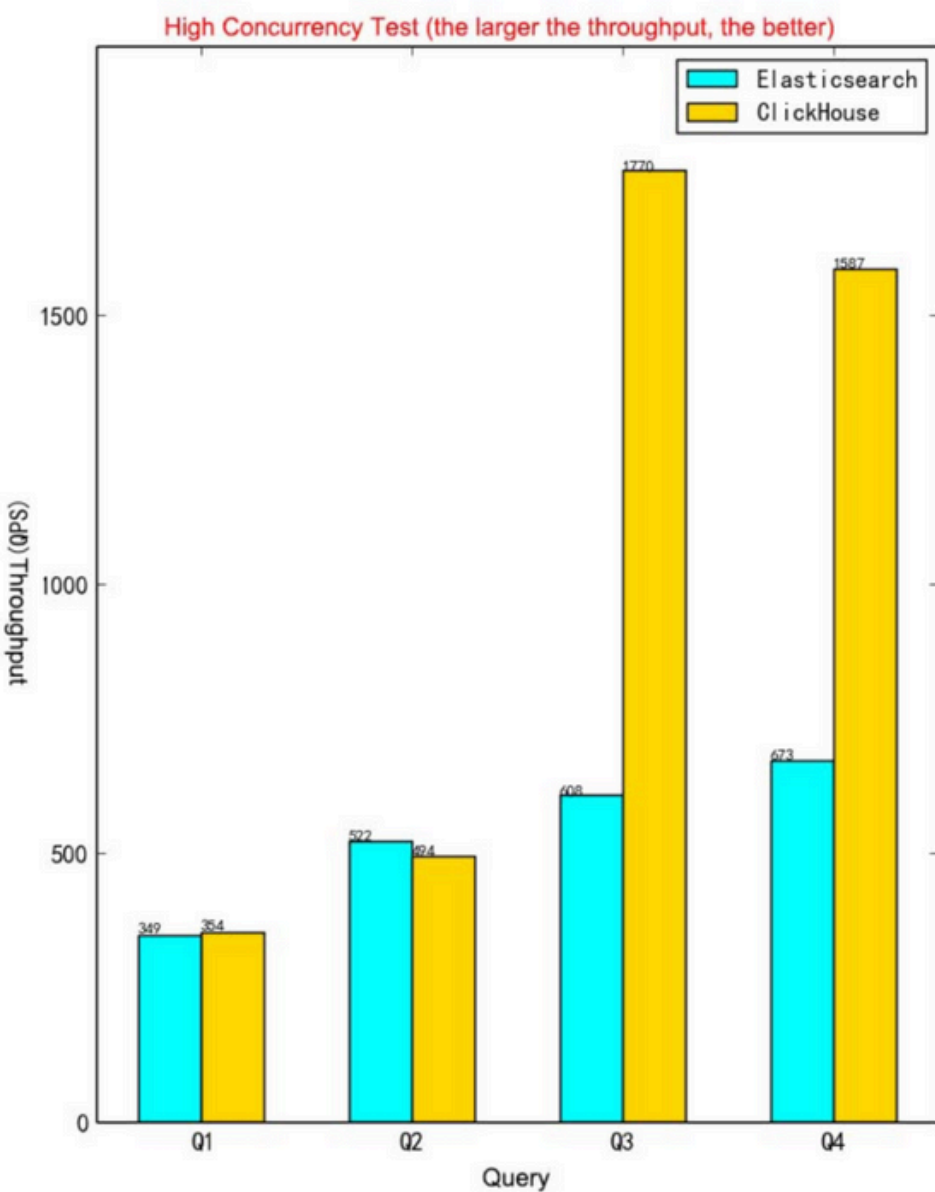
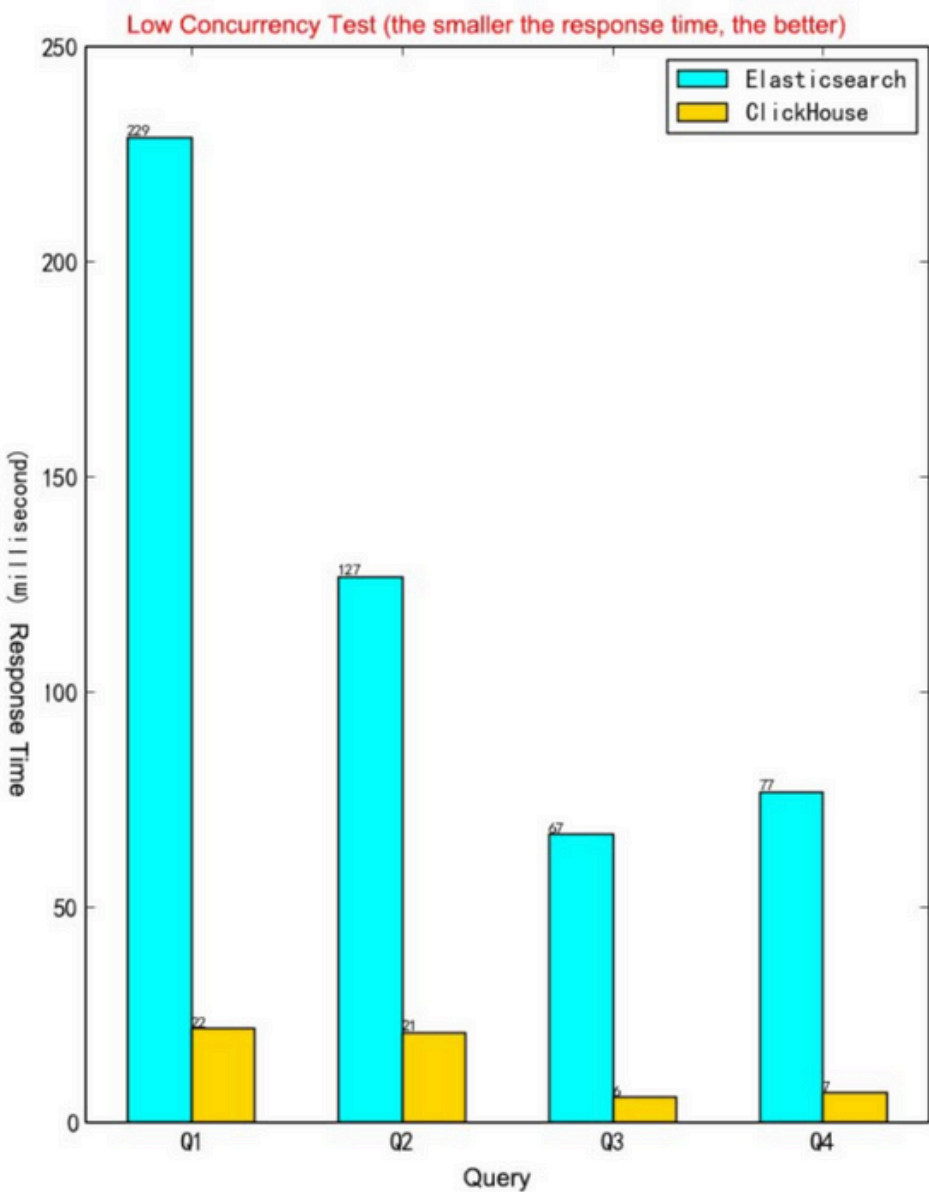
	#1	#2	#3	#4	#5
					
Name	Percy	Dylan	Mary	Dennis	Megan
Industry	Cons...	Cons...	Heal...	Heal...	Space
Age	3	5	4	8	5
Spines	2,483	1,902	2,190	1,828	1,231

Direction of Storage Scan

# PERFORMANCE COMPARISION

ClickHouse	Elasticsearch	Node Number
CPU: 8 cores	CPU: 8 cores	4
Memory: 32 GB	Memory: 32 GB	
Storage: ESSD PL1 1,500 GB	Storage: ESSD PL1 1,500 GB	

Performance comparison access\_log (Data Volume: 197,921,836 Records)

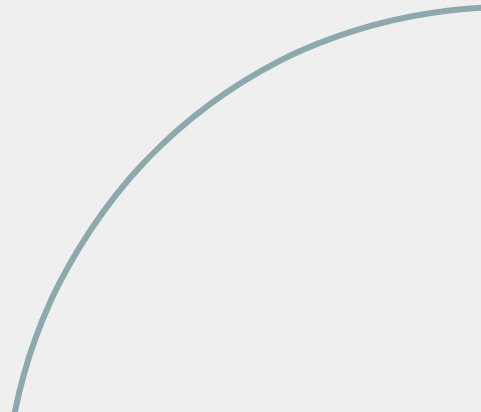


# DISK SPACE AND COSTS COMPARISONS

Stack	Disk Space	Rows	Avg Size (Bytes/Row)
Elasticsearch	22.1GB	65 million	340
Clickhouse	1.13GB	65 million	17.38

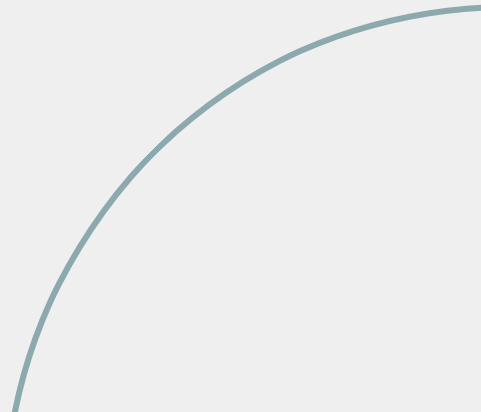
Stack	logd of line	Disk space	Instance type	Hourly cost (excluded EBS)
Elasticsearch	28 billion	13TB	3 m5a.8xlarge instances, each with 32 cores and 128GB RAM	\$2.67 per hour.
Clickhouse	30 billion	2.5 TB	4x r5a.2xlarge instances, which are overpowered for the given load with 8 cores and 64GB RAM	\$1.14 per hour

# PROS OF CLICKHOUSE


- **Exceptional Performance and Scalability:** ClickHouse stands out for its impressive query performance and ability to handle large datasets and high query loads. With its columnar storage and distributed architecture, it efficiently processes data, delivering fast and scalable performance.
  - **Comprehensive SQL Support and Advanced Analytics:** ClickHouse boasts full SQL support, empowering users to interact with the database using a familiar and powerful query language. It particularly excels in analytical workloads, enabling advanced analytics, complex queries, and robust aggregation capabilities.
  - **Optimal Data Storage and Compression:** ClickHouse optimizes data storage through its columnar storage format and effective compression techniques. By efficiently organizing and compressing data, it achieves significant storage reduction while ensuring quick query execution.
  - **Open Source Flexibility and Expandability:** As an open-source project, ClickHouse fosters transparency, flexibility, and extensibility. Its vibrant community allows for customization and extension, empowering users to tailor the database to their specific needs. Moreover, ClickHouse's ecosystem offers a wide range of integrations and connectors, facilitating seamless integration with various tools and platforms.
- 



# CONS OF CLICKHOUSE

- **Limited data modification:** ClickHouse is primarily designed for analytical workloads and may have restrictions when it comes to frequent data updates or real-time data ingestion.
  - **Steep learning curve:** ClickHouse's advanced features and configuration options may require a significant learning curve for users unfamiliar with analytical databases.
  - **Lack of native search capabilities:** While ClickHouse excels in analytics, it may not offer native search capabilities comparable to full-text search engines like Elasticsearch.
  - **Limited ecosystem and integrations:** ClickHouse's ecosystem and available integrations may be more limited compared to other popular databases, potentially requiring more custom development for certain use cases.
- 

# CLICKHOUSE & GRAFANA VS ELK STACK

- **Cost:**
    - Clickhouse & Grafana: Lower infrastructure and maintenance costs due to optimized storage and open-source nature.
    - ELK Stack: Higher costs due to heavier resource requirements and possible licensing fees (Elastic) for managed services.
  - **Performance:**
    - Clickhouse: Fast query performance, especially for time-series data.
    - ELK Stack: Powerful search capabilities but may slow down with large datasets.
  - **Scalability:**
    - Clickhouse: Scales efficiently horizontally.
    - ELK Stack: Requires more resources for scaling, leading to increased costs.
  - **Cloud Flexibility:**
    - Clickhouse & Grafana: Easily deployable across different cloud providers.
    - ELK Stack: For Managed Service - Tied more closely to specific cloud services, particularly Elasticsearch Service on AWS or Elastic Cloud.
- 

# CONCLUSION

- Clickhouse & Grafana offer a performant, cost-effective, cloud-agnostic alternative to traditional logging tools like the ELK Stack.
  - They provide superior performance and flexibility, especially in cloud-native and large-scale environments.
- 

Step 1 - Install ClickHouse and Grafana using Helm or any other package manager of your choice.

Step 2 - Install Vector to Ship Server Logs to ClickHouse Database.

Step 3 - Configure Vector

- Define Sources and Sinks.
- Configure custom Transformations of logging data if required.

Step 4 - Create Log Table Schema in ClickHouse Database.

Step 5 - Define required indexing and data partitioning keys if required.

Step 6 - Create Logging Dashboard on Grafana with ClickHouse as Datasource.



# THANK YOU

Scan connect to on LinkedIn



**Pushpender Kumar**

DevOps Architect specializing in  
Cloud Computing at OLX India

