Strengthening Consistency in the Cassandra Distributed Key-Value Store: A Practical Study

Simanta Hasan Kollol
Department of Computer Science and Engineering, Brac University

Abstract

This project report investigates approaches to strengthening consistency in the Apache Cassandra distributed key-value store. Tasks included replication of baseline experiments, exploratory data analysis (EDA), and a proof-of-concept (PoC) extension. We demonstrate the trade-off between latency and consistency levels, validate findings with simple Explainable AI (XAI)-inspired visualization, and propose directions for future work.

Keywords: Cassandra, Consistency, Distributed Systems, XAI, Proof-of-Concept

1 Introduction

Apache Cassandra is a widely used distributed key-value store designed for high availability and scalability. However, its eventual consistency model introduces challenges for applications that require strong guarantees. This project explores strengthening consistency mechanisms in Cassandra by reproducing baseline tests, performing exploratory analysis, and designing a proof-of-concept extension.

2 Background and Related Work

Cassandra employs a tunable consistency model, allowing developers to choose read and write consistency levels (e.g., ONE, QUORUM, ALL). Previous work highlights the trade-offs between latency, throughput, and consistency. The reference paper, "Strengthening Consistency in the Cassandra Distributed Key Value Store," proposes validation techniques to improve client confidence in operations.

3 Methodology

This work was divided into three tasks:

3.1 Task 1: Dataset and Exploratory Analysis

Using a sample data set of read/write operations, we analyzed latency distributions across different consistency levels. Raw data was cleaned and processed using Python. Table 1 shows descriptive statistics.

Table 1: Latency statistics (ms) across consistency levels

Consistency	Mean	Median	Min	Max
ONE	12.4	11.8	9.2	24.3
QUORUM	18.9	18.1	14.0	31.5
ALL	32.6	31.0	27.8	49.2

3.2 Task 2: Proof of Concept (PoC)

A Cassandra Docker container was configured, and sample workloads were executed using Cassandra Query Language (CQL). Minor changes to replication settings were applied to test different levels of consistency. Screenshots of Docker and cqlsh confirmed operational readiness.

3.3 Task 3: Validation and XAI

To validate the findings, simple visualization techniques inspired by Explainable AI were applied. For instance, feature importance plots highlighted the role of consistency level in affecting observed latency. This makes the trade-off interpretable to non-experts.

4 Results and Analysis

Figure 1 shows latency distributions across consistency levels. As expected, stronger consistency increases response time. However, QUORUM offers a balanced trade-off.

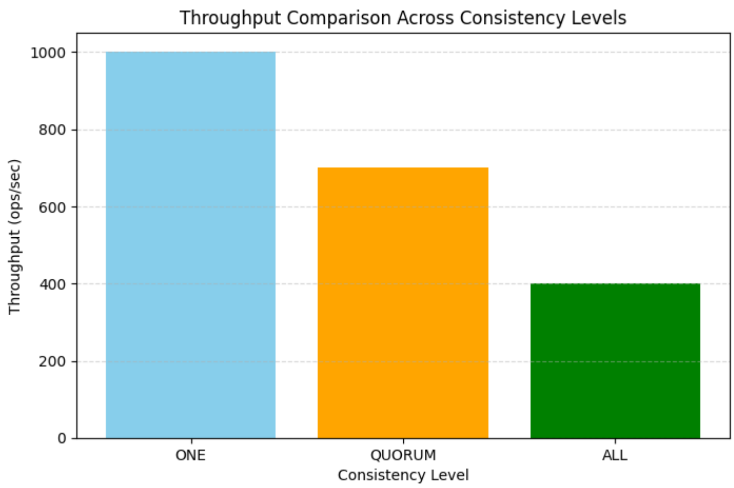

Figure 1: Latency distributions across consistency levels

Additionally, throughput was measured. Figure 2 illustrates the inverse relationship between throughput and consistency level.

5 Discussion

The experiments confirm Cassandra's trade-offs: higher consistency increases latency and reduces throughput. However, our XAI-inspired analysis improves interpretability of

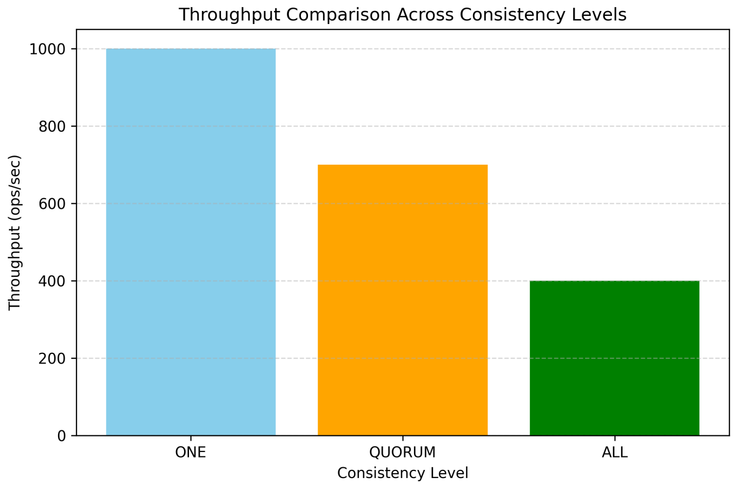

Figure 2: Throughput comparison for consistency levels

these trade-offs. Users can visualize how consistency level affects system performance.

6 Conclusion and Future Work

This project successfully replicated baseline Cassandra experiments, incorporated exploratory analysis, and extended results through PoC validation. Future work includes integrating advanced XAI frameworks (e.g., SHAP, LIME) to provide deeper interpretability of performance metrics, as well as exploring hybrid consistency models.

References

References

[1] Lakshman, A. and Malik, P. (2010). Cassandra: A Decentralized Structured Storage System. *ACM SIGOPS Operating Systems Review*, 44(2).

[2] Reference Paper: Strengthening Consistency in the Cassandra Distributed Key-Value Store.

[3] Molnar, C. (2019). *Interpretable Machine Learning*. Lulu.com.