|  |  |
| --- | --- |
| **Shubham Kaushik** | Ph.D. Researcher @ Brandeis University |
| Phone: +1 (774) 519-0913 | Waltham, MA |
| Email: [kaushiks@brandeis.edu](mailto:kaushiks@brandeis.edu) |  |

Dear Hiring Manager,

I am a Ph.D. researcher at the Department of Computer Science at Brandeis University, advised by Prof. Subhadeep Sarkar. Prior to joining Brandeis, I earned my Master’s degree in Computer Science with a specialization in data-centric computing from Boston University. Before this, I spent over four years working with both established IT giants and agile startups, where I built a solid foundation in developing data-driven products from inception through to production deployment.

My research interests include databases, storage engines, and distributed systems with a particular focus on designing and optimizing NoSQL key-value stores. Specifically, over the two years, I have devoted my time to investigating two key challenges of log-structured merge (LSM) tree-based key-values stores: (i) **optimizing ingestion performance** for mixed (interleaved inserts and queries) workloads and (ii) **improving range query performance**.

In LSM-based data stores, memory buffers are a small but crucial component of the overall architecture, serving as the entry point for every operation. The performance of an LSM-based data store is largely governed by the implementation and tuning of the memory buffer. The goal of this project is to find out the **optimal data structure and tuning for the LSM buffer based on the workload and dynamically switch to the optimal configuration as the workload shifts**. As a first step toward this, we performed an analytical study of six different buffer implementations under a wide variety of workloads to quantify their performance implications. The results of this work were **published in DBTest 2024**, a workshop associated with ACM SIGMOD. Presently, I am exploring **Cassandra’s trie-based buffer implementation** to further enrich the solution space with the next milestone being designing a workload-aware dynamic buffer switching policy.

In parallel, I am also working on improving range query performance in LSM-based data stores. Range queries in LSM-engines are realized by scanning through the qualifying data spread across multiple sorted runs and merging them in memory to create the result set. In the presence of updates and deletes in a workload, this can lead to significant read amplification, especially for tiered LSM designs. The problem is further exacerbated when there are many range queries with overlapping ranges. To this end, I introduced RangeReduce, a **range query-driven compaction algorithm that takes advantage of the data read during range queries by conditionally compacting them** and writing them back as a single sorted run on disk after the query is terminated. Piggybacking compactions on range queries allows us to (1) **avoid superfluous I/Os** during future range queries, (2) **reduce write amplification** by purging the logically invalidated entries (by updates and deletes), as well as (3) **reduce space amplification**. By removing duplicates within each sorted run, RangeReduce significantly reduces data movement based on query selectivity, enhancing performance in workloads with frequent updates, deletes, and range queries. We plan to submit this work to the next SIGMOD cycle.

I am very excited about the prospect of joining Dolby as an intern, as the company’s mission to advance high-performance, distributed database technologies resonates strongly with my professional experience and research focus. My background in data-centric computing and hands-on experience in creating scalable, production-ready solutions align well with Dolby's mission to support companies in deriving real-time insights and operational intelligence from their data. I am excited to contribute to the company’s mission through a successful internship.

Thank you for considering my application. I look forward to the possibility of learning from and contributing to the team and making a meaningful impact.

Sincerely,  
Shubham Kaushik  
*PhD researcher,  
Department of Computer Science  
Brandeis University*