Shubham Kaushik

Ph.D. Researcher @ Brandeis University

+1 (774) 519-0913 | kaushiks[at]brandeis[dot]edu | LinkedIn | Github | Waltham, MA

RESEARCH INTERESTS

Databases, Data systems, Storage systems, Distributed systems, Data streaming

PROFESSIONAL EXPERIENCE

Jan 2024 - Present	**Ph.D. Researcher** Brandeis University, MA, United States
Mar 2022 - Aug 2022	**Software Engineer**, *Scalability & Infrastructure* Kwalee, India
Jun 2021 - Mar 2022	**Engineer - Information Security**, *Information Security Automation* FIS Global, India
Oct 2019 - Jun 2021	**Project Engineer**, *Cloud Computing* Wipro Limited, India
Jul 2018 - Oct 2019	**Project Engineer**, *Big Data Security* Wipro Limited, India
Mar 2017 - Apr 2017	**Full Stack Developer Intern**, *Full Stack Development* SoPo Internet Private Limited, India

EDUCATION

Jan 2024 - Present	**Doctor of Philosophy (Ph.D.)** Brandeis University, MA, United States Major: **Computer Science**
Sep 2022 - Dec 2023	**Master of Science (M.S.)** Boston University, MA, United States Major: **Computer Science** with specialization in *"Data-Centric Computing"* GPA: *3.88/4.0*
Jul 2014 - Jun 2018	**Bachelor of Technology (B.Tech.)** Maharshi Dayanand University, Haryana, India Major: **Computer Science** & **Engineering** Thesis: *"Fault Modelling of an Object-Oriented System using Colored Petri Nets"*

PUBLICATIONS

DBTest 2024	**Shubham Kaushik** and Subhadeep Sarkar. *Anatomy of the LSM Memory Buffer: Insights & Implications*, In Proceedings of the International Workshop on Testing Database Systems
JCSE 2019	**Shubham Kaushik** and Ratneshwer. *Fault Modeling of an Object-Oriented System using CPN*, International Journal of Computer Sciences and Engineering

POSTERS

NEDB Day 2025/2024	**Shubham Kaushik**, Manos Athanassoulis, and Subhadeep Sarkar *RangeReduce: A Range Query Driven Compaction for LSM-Trees*, North East Database Day

TALKS

09/2024	"Advancements in LSM Buffering: Performance and Implementations", **Brandeis University**, USA
06/2024	"Anatomy of LSM Memory Buffers: Insights and Implications", **DBTest**, Santiago, Chile
05/2024	"LSM Trees & Buffering: Fundamentals and Research Trends", **Brandeis University**, USA

TECHNICAL SKILLS

o **Programming Languages**: C, C++, Python, SQL, Rust (*learning*)

o **Markup Languages**: HTML, CSS, JSON, YAML, LaTeX, Markdown

o **Databases**: RocksDB, Postgres, MongoDB, Redis, SQLite, ORM

o **Tools & Systems**: Kafka, Hadoop, gRPC, Microservices, Asyncio, Git, ETL, Flink, AWS

PROJECTS

o **Designing Range Query-Aware Log-Structured Merge (LSM) Trees** (*Ongoing*): LSM trees are at the heart of several NoSQL data stores due to their ingestion-optimized design. However, this superior ingestion performance comes at the cost of poor range query performance and increased write amplification. In this project, we introduce a new family of data reorganization strategies and data layouts, driven by range queries. These strategies allow us to (i) reduce the overall data movement during workload execution and (ii) reduce the I/O cost for future range queries. I am currently in the process of integrating our solutions on RocksDB, a widely used commercial LSM-based data store. [readme]

o **Anatomy of LSM Memory Buffer** (*Ongoing*): In LSM-based data stores, memory buffers are often overlooked due to their small footprint, yet they serve as a critical gateway for data ingestion. This project focuses on optimizing buffer implementations that dynamically adapt to shifting workloads to maximize performance. As a first step, we analyzed existing buffer designs, with our findings accepted at DBTest 2024 (SIGMOD). Currently, we are exploring Cassandra's trie-based buffer and other novel implementations to expand the solution space, with the next milestone focused on workload-aware dynamic buffer switching policy. [PDF]

o **Enabling Efficient Range Deletes in LSM-Trees** (*Ongoing*): LSM-based data stores perform data deletion logically, without physically deleting the target data objects. This leads to significant performance bottlenecks when deleting ranges of data, as the logically deleted data continues to 'live' in the database, increasing the overall cost of operations. In this project, we introduce a light weight and updatable range delete filter to avoid superfluous accesses to slow storage in exchange for a small amount of metadata in memory. The proposed solution substantially reduces the execution cost for workloads with range deletes. [readme]

o **Heterogeneity-Aware Operator Placement for Stream Processing Systems at the Edge**: Streaming systems process real-time data but rely on static operator configurations, which are suboptimal for dynamic workloads. This project proposed dynamic operator placement based on data selectivity and heterogeneity. I modified Apache Flink's scheduler to dynamically switch tasks at the edge devices (*Raspberry Pi*) and servers, reducing network traffic and improving system efficiency and resource utilization. [readme]

o **Finding Vulnerabilities in VS Code Extensions**: The third-party extensions can introduce security risks, compromising base applications. I developed an automated tool to detect vulnerabilities in VS Code extensions. I built a simulation framework to install, execute, and analyze extensions for open-port risks, specifically targeting *Path Traversal* and *Zip Slip* attacks, identified vulnerabilities in 5% of examined extensions. [readme]

TEACHING ASSISTANT

Spring 2025/2024	*Database Management Systems* (*COSI 127B*), Brandeis University [SP '25] [SP '24]
Fall 2024	*Introduction to Computer Networking* (*COSI 128A*), Brandeis University
Fall/Spring 2023	*Data Mechanics* (*DS 310*), Boston University
Fall 2022	*Computer Networks* (*CS 455*), Boston University

PROFESSIONAL SERVICES

2025	Web Chair for *Northeast Database (NEDB) Day 2025*.
2024	External reviewer for IEEE International Conference on Big Data.
2024 - Present	Member, Association for Computing Machinery (ACM).
2024 - Present	Student Member, Institute of Electrical and Electronics Engineers (IEEE).

CERTIFICATION

Jul 2023	*"The Ultimate Hands-On **Hadoop**: Tame your **Big Data!**"* - Udemy [link]
Jul 2023	*"Beginning **C++** programming from Beginner to Beyond"* - Udemy [link]
Oct 2018	Statement of accomplishment for *"**Python** Track"* - DataCamp [link]

CURRICULAR ACTIVITIES

Sep 2023	Judged and mentored at *HackMIT 2023*, aiding teams with technical challenges.
Nov 2022	Mentored 4 teams, with an average of 20 participants at *BostonHacks*.
Jan 2017	Volunteered in the Program Event Management team at the *National Youth Festival*.