

Shubham Kaushik

Ph.D. Researcher @ Brandeis University

+1 (774) 519-0913 | [kaushiks\[at\]brandeis\[dot\]edu](mailto: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

TPCTC 2025	Alexander H. Ott, Shubham Kaushik , James Chen, and Subhadeep Sarkar. <i>Tectonic: Bridging Synthetic and Real-World Workloads for Key-Value Benchmarking</i> , TPC Technology Conference on Performance Evaluation & Benchmarking
DBTest 2024	Shubham Kaushik and Subhadeep Sarkar. <i>Anatomy of the LSM Memory Buffer: Insights & Implications</i> , In Proceedings of the International Workshop on Testing Database Systems
JCSE 2019	Shubham Kaushik and Ratneshwer. <i>Fault Modeling of an Object-Oriented System using CPN</i> , International Journal of Computer Sciences and Engineering

POSTERS

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

TALKS

10/2025	“Range Queries and LSM-Trees: From Foes to Friends”, Salesforce , USA
09/2025	“Tectonic: Bridging Synthetic and Real-World Workloads for Key-Value Benchmarking”, TPCTC , UK
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

- **Programming Languages:** C, C++, Python, SQL, Rust
- **Markup Languages:** HTML, CSS, JSON, YAML, \LaTeX , Markdown
- **Databases:** RocksDB, Postgres, MongoDB, Redis, SQLite, ORM
- **Tools & Systems:** Kafka, Hadoop, gRPC, Microservices, Asyncio, Git, ETL, Flink, AWS

KEY PROJECTS

- **Tectonic** (*Ongoing*): A Rust-based synthetic workload generator designed for benchmarking and performance evaluation. The existing workload generators struggles to capture the diverse and dynamic patterns of real-world scenarios, such as shifting access or sortedness-aware operations. Tectonic addresses these limitations by modeling these properties directly, providing a more realistic and flexible evaluation environment. It achieves up to $2\times$ higher throughput and reduces memory consumption by 84% compared to state-of-the-art benchmarks. [[Paper PDF](#)] [[GitHub](#)]
- **RangeReduce** (*Ongoing*): Log-Structured Merge (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. RangeReduce achieved up to 18% lower range-query latency, 90% reduction in compaction debt, 12% less overall data movement, and 20% lower space utilization, enhancing overall LSM engine efficiency. [[Blog](#)] [[GitHub](#)]
- **LearnedLSMBuffer** (*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) which provides a framework for optimal buffer selection for various workloads, enabling data-driven design choices. [[Paper PDF](#)] [[Blog](#)] [[GitHub](#)]
- **MorphStream**: 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. [[Blog](#)] [[GitHub](#)]
- **Exposé**: 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. [[Blog](#)] [[Report](#)] [[GitHub](#)]

TEACHING ASSISTANT

Spring 2025/2024 | *Database Management Systems (COSI 127B)*, Brandeis University [[SP '25](#)] [[SP '24](#)]
Fall 2024/2025 | *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

2026 | Member, VLDB 2026 Shadow PC.
2025 | Web Chair for *Northeast Database (NEDB) Day 2025*.
2024/2025 | 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](#)]

EXTRA CURRICULAR ACTIVITIES

Sep 2023/2025 | 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*.