

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 , Server Programming Team Kwalee, India
Jun 2021 - Mar 2022	Engineer - Information Security , Cyber Fusion, Information Security FIS Global, India
Oct 2019 - Jun 2021	Project Engineer , Python Cloud Computing, Wipro Digital Wipro Limited, India
Jul 2018 - Oct 2019	Project Engineer , Big Data, Cyber Defense Wipro Limited, India
Mar 2017 - Apr 2017	Full Stack Developer Intern , Backend Team 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	Masters 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 , 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 , Ratneshwer. <i>Fault Modeling of an Object-Oriented System using CPN</i> , International Journal of Computer Sciences and Engineering

POSTERS

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

BACHELOR'S THESIS

Shubham Kaushik. *Fault Modelling of an Object-Oriented System using Colored Petri Nets*, 2018.

Advisor: [Dr. Ratneshwer](#), School of Computer and Systems Sciences, Jawaharlal Nehru University.

TECHNICAL SKILLS

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

PROJECTS

- **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](#)]
- **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](#)]
- **Heterogeneity-Aware Operator Placement for Stream Processing Systems at the Edge**: Streaming systems are widely used for real-time data processing. However, all operators within a cluster runs with a static configuration, which is suboptimal for dynamic workloads. In this project, we proposed an approach to dynamically place operators based on the selectivity and heterogeneity of the data. Toward this, I modified Apache Flink's scheduler to dynamically switch tasks at the edge devices (*Raspberry Pi*) and servers. This reduced the network traffic and improved system efficiency and resource utilization. [[readme](#)]
- **Finding Vulnerabilities in VS Code Extensions**: The use of third-party extensions can introduce potential security vulnerabilities, which can also render the base applications vulnerable. In this project, I identified security vulnerabilities in VS Code extensions and developed an automated tool for their detection. I devised a simulation framework using the Pyautogui library to install and execute extensions, and detect security vulnerabilities by analyzing the open ports associated with each extension. The analysis specifically targeted *Path Traversal* and *Zip Slip* attacks, and detected 5% of the extensions examined as vulnerable. [[report](#)] [[readme](#)]

TEACHING EXPERIENCE

Fall 2024	Teaching Assistant , <i>Introduction to Computer Networking (COSI 128A)</i> <i>Michtom School of Computer Science, Brandeis University, MA, United States</i>
Spring 2024	Teaching Assistant , <i>Database Management Systems (COSI 127B)</i>
Spring 2025	<i>Michtom School of Computer Science, Brandeis University, MA, United States</i>
Fall 2023	Teaching Assistant , <i>Data Mechanics (DS 310)</i>
Spring 2023	<i>Center for Computing & Data Sciences, Boston University, MA, United States</i>
Fall 2022	Teaching Assistant , <i>Computer Networks (CS 455)</i> <i>Department of Computer Science, Boston University, MA, United States</i>

EXTERNAL REVIEWER

2024 IEEE International Conference on Big Data (BigData 24)

CERTIFICATIONS

- 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

- Jan 2025 | Web Chair for *Northeast Database (NEDB) Day 2025*, Brandeis University.
- 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*.