

Shubham Kaushik

Graduate Student @ Boston University

Contact Information

Contact.: +1 (774) 519-0913
Email: kaushiks@bu.edu ; shubhamk00020@gmail.com
Website: shubhamkaushik.com ; [Linkedin](#) ; [Github](#)
Address: 569 Cambridge Street, Apt. #2, Allston, MA 02134

Research Interests

Databases, Data systems, Storage systems, Distributed systems, Data streaming, Cyber security

Education

Sep 2022 - Present	Masters of Science (M.S.) Boston University , MA, United States Major: Computer Science with specialization in “Data-Centric Computing” GPA: 3.83/4.0
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”

Work Experience

Mar 2022 - Aug 2022	Kwalee , India Software Engineer , Server Programming Team
Jun 2021 - Mar 2022	FIS Global , India Engineer - Information Security , Cyber Fusion, Information Security
Oct 2019 - Jun 2021	Wipro Limited , India Project Engineer , Python Cloud Computing, Wipro Digital
Jul 2018 - Oct 2019	Wipro Limited , India Project Engineer , Big Data, Cyber Defense
Mar 2017 - Apr 2017	SoPo Internet Private Limited, India Full Stack Developer Intern , Backend Team

Publication

- o **Shubham Kaushik**, Ratneshwer. [Fault Modelling of an Object-Oriented System using CPN](#), International Journal of Computer Sciences and Engineering (JCSE), 2019.

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.

My thesis analyzes the formal representation of object-oriented properties using colored petri nets to enhance the understanding of software fault behavior, offering insights for the software development, testing, and maintenance phases. I modeled the various faults in object-oriented systems, emphasizing on faults related to object interactions, such as inheritance and polymorphism states.

Teaching Experience

Fall 2023	Teaching Assistant , Data Mechanics (DS 310)
Spring 2023	Center for Computing & Data Sciences, Boston University, MA, United States
Fall 2022	Teaching Assistant , Computer Networks (CS 455) Department of Computer Science, Boston University, MA, United States

Technical Skills

- o **Programming Languages:** Python, C, C++, Java, JavaScript, SQL, Go (*learning*)
- o **Markup Languages:** HTML, CSS, JSON, YAML, \LaTeX , Markdown
- o **Databases:** RocksDB, MySQL, MongoDB, Redis, Elasticsearch, SQLite, ORM
- o **Tools & Frameworks:** Django, Flask, Microservices, Asyncio, Kafka, Git, ETL, Hadoop, AWS

Projects

Academic Research

- **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. [[ppt](#)] [[report](#)] [[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. [[report](#)] [[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. [[ppt](#)] [[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](#)]

Industry Research

- **Asynchronous IP-Scanning**: In this project, I designed an *asynchronous* IP scanning module to pipeline the scanning of a large number (more than 100K) of IPs from internet. This module was implemented in *Python* while leveraging *multithreading*, *socket programming*, and *asyncio* techniques to maximize the performance. The automation enhanced the security posture, emphasizing the importance of continuously monitoring IP addresses, and improved the productivity of the security team by up to 50%.
- **Data Masking**: I developed a RESTful microservice to securely mask personally identifiable information in various file formats, including CSV, Excel, mainframe, and HTML. I also designed an efficient masking algorithm that pseudonymizes the users' personal data stored in flat files. This service streamlined the process of masking thousands of files, enhancing the system performance and reducing manual intervention by 70%.
- **Data Inquiry**: I optimized the Data Inquiry service using the *asyncio framework* in *Python* which enabled *asynchronous execution* and effective coordination of HTTP requests. It allowed the inquiry service to process multiple requests concurrently and reduce the turnaround time. These enhancements improved the system efficiency by up to 50%, allowing faster responses and better utilization of computational resources.
- **Anomaly Detection**: In this cyber-defense project, I developed a novel anomaly detection agent that is capable of monitoring network traffic from multiple systems using *Scapy*. The agent continuously analyzed the data in real time, identifying and flagging any suspicious or unusual activities. The agent also identified malicious IPs using *K-means clustering* and a repository of malicious IPs. It triggers alerts for harmful events, providing valuable information to security analysts for further investigation and threat mitigation.

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](#)]
- May 2016 | *"Core Java"* - Oracle's Workforce Development Program [[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 Program Event Management team at the *National Youth Festival*.