SUPUN ABEYSINGHE

supun.online supunabeysinghe@gmail.com github.com/supunab in/supunabeysinghe
West Lafayette, IN, USA +1 (765)-775-8679

EDUCATION

Purdue University West Lafayette, IN
Ph.D. in Computer Science - **GPA: 4.0/4.0** *Expected Dec' 23*
Advisor: Prof. Tiark Rompf
Relevant Courses: Operating Systems, Algorithms, Compilers, Database Systems, Distributed Systems

University of Moratuwa Sri Lanka
B.Sc Engineering (Hons) (Computer Science & Engineering) *Dec' 18*
First Class Honours - **GPA: 4.06/4.20**

EXPERIENCE

Research Intern, Microsoft *Sep 2022 - Dec 2022*
- Developed a compiler prototype to generate specialized ONNXRuntime (ORT) custom ops from ONNX models, utilizing efficient specialized kernels (CUTLASS). Demonstrated significant performance improvements for critical models such as GPT2 and BERT compared to the standard ORT implementation.

Software Engineer Intern, SambaNova Systems *May 2022 - July 2022*
- Worked on optimization efforts on a critical analysis pass in the compiler stack (in MLIR) to reduce resource consumption. Achieved a significant increase in resource utilization and successfully integrated the optimizations into production.

Graduate Research Assistant, Purdue University *May 2020 - Present*
- Led three research projects focused on applying runtime code generation techniques to develop high-performance data analytics (SQL and Datalog) and ML systems. Submitted three papers for publication, with one already accepted at SIGMOD '22 and others currently under review. Co-authored one paper accepted at OOPSLA '23.

Research Engineer, WSO2 Inc. Sri Lanka *Jan 2019 - July 2019*
- Contributed to a research project focused on leveraging ML techniques to dynamically auto-tune server configurations. Co-authored a paper published in ISCC '19, which presented novel findings on the effectiveness of the approach.

Research Intern, StatNLP Lab, SUTD Singapore *Jun 2017 - Dec 2017*
- Was responsible for testing and fixing bugs of the StatNLP framework by implementing several traditional ML models.

SELECTED PROJECTS

1. **Optimizing End-to-end Data Science Pipelines:** Leveraging generative programming techniques to accelerate combined data processing and ML workloads by constructing common intermediate (IR) layer integrations. Observed **speedups up to 20x** in end-to-end performance. **(Scala, C++, CUDA)**
2. **Efficient Incrementalization of SQL Queries with Nested Aggregates:** Building novel tree-based index structures to improve the incrementalization efficiency of nested-aggregate queries by **up to 1000x** over the state-of-the-art.
3. **Building Efficient and Expressive Datalog Systems:** A Datalog compiler for program analysis with support for a variety of features like user-defined aggregates, UDFs, stratified negations, user-defined aggregates and so on, built based on generative programming techniques. Flan achieves competitive performance (and sometimes significantly outperform) with state-of-the-art Datalog compilers. **(Scala, C++, Datalog)**
4. **Server Parameter Auto-tuning using Machine Learning (2019):** Leveraging ML based optimization techniques to dynamically auto-tune server parameters to enhance runtime performance. **(Java, Python, PyTorch)**
5. **Social Media Analytics Platform (2018):** A platform for automatically extracting information related to a particular entity (e.g., restaurant reviews from multiple sources) and performing various types of analysis (emotion detection, aspect-based sentiment analysis, trending topics and evolution of them, etc.). **(Python, PyTorch)**

TECHNICAL SKILLS

- **Programming Languages:** Scala, Java, Python, C, C++
- **Systems and Libraries:** Spark, Flink, PyTorch, Tensorflow, CUDA, LLVM, MLIR;
  Prior Experience - Spring Boot, Angular, Node.js

SELECTED PUBLICATIONS

1. **Efficient Incrementialization of Correlated Nested Aggregate Queries using Relative Partial Aggregate Indexes (RPAI)**
   Supun Abeysinghe*, Qiyang He, Tiark Rompf*
   (SIGMOD '22)
2. **Architecting Intermediate Layers for Efficient Composition in End-to-End Data Science Pipelines**
   Supun Abeysinghe*, Fei Wang, Gregory Essertel, Tiark Rompf*
   (in submission)
3. **Flan: An Expressive and Efficient Datalog Compiler for Program Analysis**
   Supun Abeysinghe*, Anxhelo Xhebraj, Tiark Rompf*
   (in submission)
4. **Graph IRs for Impure Higher-Order Languages - Making Aggressive Optimizations Affordable with Precise Effect Dependencies**
   Oliver Bračevac, Guannan Wei, Songlin Jia, ***Supun Abeysinghe****, Luke Jiang, Yuyan Bao, and Tiark Rompf*
   (OOPSLA '23)
5. **ADAPT-T: An Adaptive Algorithm for Auto-Tuning Worker Thread Pool Size in Application Servers**
   Nilushan Costa, Malith Jayasinghe, Ajantha Atukorale, ***Supun Abeysinghe****, Srinath Perera, Isuru Perera*
   At International Symposium on Computers and Communications (ISCC), Barcelona, Spain (2019)
6. **Sentylic at IEST 2018: Gated Recurrent Neural Network and Capsule Network Based Approach for Implicit Emotion Detection**
   Prabod Rathnayaka, ***Supun Abeysinghe****, Chamod Samarajeewa, Isura Manchanayake, Malaka Walpola*
   At Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (held with EMNLP) (2018)

SELECTED COURSE PROJECTS

1. **Graph Query Compilation:** Extending relational query compilation techniques based on generative programming to support compilation of graph queries, achieving **an order of magnitude speedup** compared to interpreted engines. **(Scala, C++)**
2. **Sharded, Distributed Key-Value Store:** Implemented a sharded, distributed KV store using Paxos algorithm for replication and support for transactions using Two-phase Commit. **(Java, Distributed Systems)**
3. **University C Compiler:** A compiler for a C-like programming language that uses LLVM as an IR. Written multiple LLVM passes for compiler optimizations. **(C++, LLVM)**

SELECTED OPEN SOURCE CONTRIBUTIONS

Lantern - Deep Learning Framework
- Implemented custom CUDA kernels for several Deep Learning operators (Softmax, Embedding layer, etc.), added support for modern deep learning models like Transformers, etc.

Lightweight Modular Staging (LMS) - A Compiler Framework
- Several key feature additions (e.g., lambda lifting support for top level functions) and general maintenance.

SELECTED AWARDS

- **Competitive Programming:** IEEEXtreme v12.0 (2018) - Country Rank 1st Global Rank 60th (top 2%), IEEEXtreme v10.0 (2016) - Country Rank 3rd Global Rank 106th (top 5%)
- 5th place - Implicit Emotion Detection Shared Task, Organized as part of WASSA 2018 at EMNLP 2018
- **Dean's List Award**: Dean's list award for all semesters (should achieve a semester GPA of 3.8 or above to be included in the Dean's List) (in undergraduate)