

Supun Nakandala

Phone: (+1) 812-558-6888
Email: snakanda@eng.ucsd.edu
Web: scnakandala.github.io

3232 EBU3B CSE
9500 Gilman Drive
La Jolla, CA 92093

Research Interests My research interest lies broadly in the intersection of Systems and Machine Learning, an emerging area which is increasingly referred to as *Systems for ML*. In this space I operate as a data management researcher. Taking inspirations from classical data management techniques, I build new abstractions, algorithms, and systems to improve efficiency, scalability, and usability of machine learning workloads. I am also interested in large-scale applied ML, which opens new systems challenges.

Education **University of California**, San Diego, CA Sept 2017 - Present
PhD, Computer Science. GPA: 3.96/4.00
Thesis Advisor: Prof. Arun Kumar
Courses: Database System Implementation, Advanced Data Analytics Systems, Data Models in Big Data Era, Advanced Compiler Design, Machine Learning, Recommender Systems and Web Mining, Algorithm Design and Analysis

University of Moratuwa, Sri Lanka Aug 2010 - April 2015
Bachelor of the Science of Engineering, Computer Science & Engineering.
GPA: 4.11/4.20.
Department Topper and Gold Medalist

Publications *Incremental and Approximate Inference for Faster Occlusion-based Deep CNN Explanations*
Supun Nakandala, Arun Kumar, and Yannis Papakonstantinou
ACM SIGMOD 2019
Honorable Mention for Best Paper Award. Invited to ACM TODS.

Compiling Classical ML Pipelines into Tensor Computations for One-size-fits-all Prediction Serving
Supun Nakandala, Gyeong-In Yu, Matteo Interlandi, and Markus Weimer
NeurIPS 2019 MLSys Workshop

Cerebro: Efficient and Reproducible Model Selection on Deep Learning Systems
Supun Nakandala, Yuhao Zhang, and Arun Kumar
ACM SIGMOD 2019 DEEM Workshop

Demonstration of Krypton: Optimized CNN Inference for Occlusion-based Deep CNN Explanations
Allen Ordookhanians, Xin Li, Supun Nakandala, and Arun Kumar
VLDB 2019 Demo | SysML 2019 Demo

Materialization Trade-offs for Feature Transfer from Deep CNNs for Multimodal Data Analytics
Supun Nakandala, Arun Kumar
SysML 2018 Short paper

Gendered Conversation in a Social Game-Streaming Platform
Supun Nakandala, Giovani Cimpaglia, Norma Su, and Yong-Yeol Ahn
AAAI ICWSM 2017

Apache Airavata Security Manager: Authentication and Authorization Implementations for a Multi-Tenant eScience Framework
Supun Nakandala, Hasini Gunasinghe, Suresh Marru, and Marlon Pierce
IEEE e-Science 2016

Anatomy of the SEAGrid Science Gateway

Supun Nakandala, Sudhakar Pamidigantam, Suresh Marru, Marlon Pierce
NSF XSEDE 2016

Pre-Prints

Deep Learning Algorithms for Identifying Sedentary Behavior from Hip Worn Accelerometer Data

Supun Nakandala, Marta Jankowaska, Fatima Tuz-Zahra, John Bellettiere, Arun Kumar, and Loki Natarajan
Under Submission

Resource-Efficient and Reproducible Model Selection on Deep Learning Systems

Supun Nakandala, Yuhao Zhang, and Arun Kumar

Under Review at SysML 2020

https://adalabucsd.github.io/papers/TR_2019_Cerebro.pdf

Vista: Declarative Feature Transfer from Deep CNNs at Scale

Supun Nakandala, Arun Kumar

Under Review at SIGMOD 2020

https://adalabucsd.github.io/papers/TR_2019_Vista.pdf

Work Experience

Research Intern - Systems for ML

June 2019 - Sept 2019

Microsoft Cloud Information Services Lab

Mentors: Matteo Interlandi, Markus Weimer

Designed and implemented a compiler for translating classical machine learning pipelines into tensor computations for unified and faster scoring of machine learning models. The initial workshop paper has been accepted to the Neurips 2019 MLSys Workshop. Full conference publication is in progress.

Research Software Developer

Oct 2015 - Aug 2017

Science Gateways Research Center - Indiana University

Manager: Marlon Pierce

Developed APACHE AIRAVATA, which is a software framework to compose, manage, execute, and monitor large scale applications and workflows on distributed computing resources such as local clusters, supercomputers, computational grids, and computing clouds.

Teaching Experience

Teaching Assistant - Advanced Data Analytics Systems

UCSD - Spring 2019

Scholarships and Awards

NSF travel award to attend ACM SIGMOD 2019

NSF - 2019

Gold Medal for the Best Academic Performance

University of Moratuwa - 2015

Travel award to attend 4th South Asia Workshop on Research

Frontiers in Computing

National University of Singapore - 2014

Mahapola Higher Education Merit Scholarship

Govt. of Sri Lanka - 2010

Research Impact

Ideas from project CEREBRO integrated into MADlib/Greenplum by Pivotal 2019
CEREBRO system is being used by behavioral science researchers at UC San Diego medical school 2019

“Gendered Conversation in a Social Game-Streaming Platform” paper gains lot of media attention and creates awareness about the bleak issue of sexism in online game streaming platforms 2017

APACHE AIRAVATA science gateways middleware and the SEAGRID science gateway are widely used by computational science researchers to execute and manage computational jobs on university clusters and national supercomputing infrastructure 2017

| | |
|-------------------------|---|
| Ongoing Projects | <div> <div>Project: Medical Data to Knowledge</div> <div>Started September 2018</div> </div> <p>In this joint project between UCSD CS department and UCSD Medical School, I develop new deep learning-based techniques for predicting human activity (e.g., sitting, standing, and stepping) from accelerometer data. The data is collected from a large cohort of patients who wore accelerometer devices for seven days of free living. The goal is to develop accurate methods to predict human activity from these accelerometer data and then use them in downstream human activity and metabolic health correlation analysis. The challenges of this project include working with large volumes of training data (1 TB) and performing extensive <i>model selection</i> such as neural architecture search and hyperparameter tuning.</p> |
| | <div> <div>Project: Cerebro</div> <div>Started September 2018</div> </div> <p>Deep neural networks are revolutionizing many ML applications. But there is a major bottleneck to wider adoption: the pain of <i>model selection</i>. This empirical process involves exploring the deep net architecture and hyper-parameters, often requiring hundreds of trials. Alas, most ML systems focus on training one model at a time, reducing throughput and raising costs; some also sacrifice reproducibility. We are developing Cerebro, which is a system to raise deep net model selection throughput at scale and ensure reproducibility. CEREBRO uses a novel parallel execution strategy we call <i>model hopper parallelism</i> which is inspired by the multi-query optimization technique. Experiments on <i>Criteo</i> and <i>ImageNet</i> datasets show CEREBRO offers up to 10X speedups and improves resource efficiency significantly compared to existing systems like Parameter Server, Horovod, and task-parallel tools.</p> |
| Technical Talks | <div> <div><i>Resource-Efficient and Reproducible Model Selection on Popular Deep Learning Systems</i></div> <div>UCSD CNS Research Review 2019</div> </div> <div> <div><i>Compiling Classical ML Pipelines into Tensor Computations for One-size-fits-all Prediction Serving</i></div> <div>Microsoft Cloud Information Services Lab 2019</div> </div> <div> <div><i>Incremental and Approximate Inference for Faster Occlusion-based Deep CNN Explanations</i></div> <div>ACM SIGMOD 2019</div> </div> <div> <div><i>Cerebro: A System for Efficient and Reproducible Model Selection on Deep Learning Systems</i></div> <div>ACM SIGMOD 2019</div> </div> <div> <div><i>Materialization Trade-offs for Feature Transfer from Deep CNNs for Multimodal Data Analytics</i></div> <div>UCSD CNS Research Review 2018</div> </div> |