Arun George Zachariah

Address: MU School of Medicine, MA215, One Hospital Dr, Columbia, MO 65212
Mob: +1 (816) 694-6537 | **Email**: azachariah@mail.missouri.edu
Website: https://arun-george-zachariah.github.io/Personal-Page/

EDUCATION

University of Missouri-Columbia
Ph.D., Computer Science (Transferred)   Spring 2020 – Summer 2022 (Expected)
Dissertation Title: *A System for Large-Scale Image and Video Retrieval on Everyday Scenes*
Advisor: *Dr. Praveen Rao*

University of Missouri-Kansas City
Interdisciplinary Ph.D., Computer Science   Fall 2018 – Fall 2019
Advisor: *Dr. Praveen Rao*

Cochin University of Science and Technology
Bachelor of Technology, Electrical and Electronics Engineering   2008 – 2012

PUBLICATIONS

Praveen Rao, **Arun Zachariah** – "Accelerating Variant Calling on Human Genomes Using a Commodity Cluster". In the 9th International Workshop on Computer and Networking Experimental Research using Testbeds (Collocated with IEEE INFOCOM), 2022. (To Appear)

Arun Zachariah, Praveen Rao, Brian Corn, and Dominique Davison - Zero Shot Learning for Predicting Energy Usage of Buildings in Sustainable Design. In AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE), 4 pages, Canada, 2022.

Arun Zachariah, Maha Alrasheed – "Private-Share: A Secure and Privacy-Preserving De-Centralized Framework for Large Scale Data Sharing". In Proc. of the *3rd ACM International Conference on Multimedia in Asia* (ACM MM Asia 2021), 3 pages, Australia, 2021.

Praveen Rao, **Arun Zachariah**, Deepthi Rao, Peter Tonellato, Wesley Warren and Eduardo Simoes - "Accelerating Variant Calling on Human Genomes Using a Commodity Cluster". In the *30th ACM International Conference on Information and Knowledge Management* (CIKM), pages 3388-3392, Australia, 2021 (Nominated for Best Short Paper Award)

Suveen Angraal, **Arun Zachariah**, Raaisa Raaisa, Rohan Khera, Praveen Rao, Harlan M Krumholz, and John A Spertus – "Evaluation of Internet-based Crowdsourced Fundraising to Cover Healthcare Costs in the United States". In *JAMA Network Open*, 4(1), 2021

Arun Zachariah, Mohamed Gharibi, Praveen Rao – "A Large-Scale Image Retrieval System for Everyday Scenes". In Proc. of the *2nd ACM International Conference on Multimedia in Asia* (ACM MM Asia 2020), 3 pages, Singapore, 2021

Nouf Alrasheed, **Arun Zachariah**, Shivika Prasanna, Deepthi Rao, and Praveen Rao. "Deepfakes for Histopathology Images: A Myth or Reality?" In *49th Annual IEEE Applied Imagery Pattern Recognition (AIPR) Workshop 2020: Trusted Computing, Privacy, and Securing Multimedia*, pages 1-7, Washington, D.C., 2020

Arun Zachariah, Mohamed Gharibi, Praveen Rao – "QIK: A System for Large-Scale Image Retrieval on Everyday Scenes with Common Objects". In the *10th International Conference on Multimedia Retrieval* (ICMR 2020), pages 126-135, Dublin, Ireland, 2020

Mohamed Gharibi, **Arun Zachariah** and Praveen Rao – "FoodKG: A Tool to Enrich Knowledge Using Machine Learning Techniques". In *Frontiers in Big Data*, Volume 3, 12 pages, 2020

Daniel E. Lopez Barron, Praveen Rao, Deepthi Rao, Ossama Tawfik, **Arun Zachariah** – "Large-Scale Storage of Whole Slide Images and Fast Retrieval of Tiles Using DRAM". In *2020 SPIE Defense + Commercial Sensing: Big Data II: Learning, Analytics, and Applications Conference* (11395), 6 pages, Anaheim, CA, 2020

Arun Zachariah, Praveen Rao, Anas Katib, Monica Senapati, Kobus Barnard – "A Gossip-Based System for Fast Approximate Score Computation in Multinomial Bayesian Networks". In the *35th IEEE International Conference on Data Engineering* (ICDE), pages 1968-197, Macau, China, 2019

DATASETS

Praveen Rao, **Arun Zachariah**, Deepthi Rao, Peter Tonellato, Wesley Warren, Eduardo Simoes, February 27, 2021, "Variant Analysis of Human Genome Sequences for COVID-19 Research", IEEE Dataport, doi: https://dx.doi.org/10.21227/b0ph-s175.

PROFESSIONAL EXPERIENCE

Graduate Research Assistant
University of Missouri-Columbia   02/2020 – Present
University of Missouri-Kansas City   09/2018 – 01/2020

Responsibilities
- Build a comprehensive evaluation framework for benchmarking Blind Learning against other state-of-the-art privacy-preserving deep learning techniques as a part of the industry-sponsored research project with TripleBlind.
- Develop a system to enable democratizing genome sequence analysis on CloudLab as a part of NSF-RAPID (CNS-2034247) and MU Center for Biomedical Informatics.
- Develop a deep learning model to predict energy usage metrics of buildings by considering a large number of building parameters, as a part of the NSF Center for Big Learning (CNS-1747751).
- Develop and test a large-scale image retrieval system as a part of the NSF Center for Big Learning(CNS-1747751).
- Develop a scalable system for storage and fast retrieval of Whole Slide Images, for enabling next generation image analytics, as a part of the NSF I-Crops program.
- High-performance networking with RDMA and DPDK to enable lightning-fast storage and retrieval.
- Collect, process, and analyze data from various sources using complex techniques and procedures as required, to help prepare grant proposals or funding applications.
- Conduct customer discovery interviews, as a part of the NSF I-Crops program, to understand industry requirements and challenges.
- Supervise students working on research projects.
- Perform other duties as assigned by the supervisor.

Intern, Engineering
Arm Ltd.   06/2021 – 08/2021

Responsibilities
- Analyze code to identify bottlenecks across software, micro-architecture, and architecture.
- Perform extensive experiments on real and simulated platforms.
- Develop innovative tools for architectural exploration and performance analysis.
- Autonomously queue and execute benchmarks on a pool of machines from MAAS and AWS with Ansible and Juju.

Technology Analyst
Infosys (Consultant at Apple Inc.)   05/2013 – 08/2018

Achievements
- Implemented innovative systems for data collection, storage and management of customer data.
- Built a custom product for Customer person data standardization, matching, de-duplication, linking and feeding to enterprise data warehouse systems.
- Successfully set up Hadoop over multiple nodes and migrated data from Oracle 10g to HBase using Sqoop.
- Designed and implemented a locking and buffering mechanism using an in-memory database, resulting in a 10 times increase in processing bandwidth with the same hardware configuration.
- Transformed regular Normalized RDBMS apps to De-Normalized structure for scaling and improved response times from 200-300ms to 20-30ms.

Responsibilities
- Translate business goals and customer needs into prioritized product requirements and use cases.
- Application tuning – JVM Performance tuning, Database tuning, Profiling, Capacity planning and health check monitoring during peak business days.
- Develop multiple MapReduce jobs in for data cleaning and pre-processing.
- Writing Hive queries to read from HBase.
- Identify new technologies and tools for enhancing product value and increasing team productivity.
- Train, mentor and manage a team of software engineers.

TECHNICAL SKILLS

- Programming Languages: C, C++, Java, Python, Scala, Gremlin
- J2EE Technologies: Servlets, JSP, Web Services
- Web Technologies: HTML5, Javascript, jQuery, Angular JS
- Database Technologies: SQL, SPARQL, Mongo DB, PostgreSQL, Greenplum, MySQL, TinkerPop, Neo4j
- Software Libraries & Tools: PyTorch, TensorFlow, Keras, Caffe, OpenCV, scikit-learn, SciPy, pandas, NumPy, Hadoop, Spark, Hive, Kafka, HBase, genism, Deepwalk
- Version Control Tools: GIT, SVN
- Build Tools: Ant, Maven, Gradle, sbt

ACADEMIC SERVICES

External Reviewer
- NeurIPS Workshop on Human and Machine Decisions   2021
- 23rd International Conference on Big Data Analytics and Knowledge Discovery   2021

- 30th International Conference on Computer Communications and Networks — 2021
- 4th International Conference on Connected and Autonomous Driving (MetroCAD) — 2021
- Cluster Computing Journal – Springer — 2020
- 8th International Conference on Big data and Cloud Computing — 2020
- 8th IEEE International Conference on Healthcare Informatics — 2020
- 25th International Conference on Pattern Recognition — 2020
- 22nd International Conference on Big Data Analytics and Knowledge Discovery — 2020
- 21st IEEE International Conference on Mobile Data Management — 2020
- 32nd Conference on Graphics, Patterns and Images SIBGRAPI — 2019
- 7th International Conference on Big data and Cloud Computing — 2019
- 21st International Conference on Big Data Analytics and Knowledge Discovery — 2019

Conference Workshops Conducted
- Upsilon Pi Epsilon - After School Program — 2021
- ACM/Code.org Hour of Code at Centro Latino — 2021
- ICBCC - Fundamentals of Deep Learning for Computer Vision — 2019

Student Volunteer
- 35th Conference on Neural Information Processing Systems (NeurIPS) — 2021
- 34th Conference on Neural Information Processing Systems (NeurIPS) — 2020
- 7th International Conference on Big data and Cloud Computing — 2019
- Deep Learning Fundamentals for Computer Vision. — 2019

AWARDS AND GRANTS

- 35th NeurIPS Student Volunteer Registration — 2021
- Upsilon Pi Epsilon Scholarship Award — 2021
- ACM SIGMM Travel Grant — 2021
- VLDB SPEND grant — 2021
- MU Graduate Professional Council Research Development Award. — 2021
- 34th NeurIPS Student Volunteer Registration — 2020
- UMKC School of Graduate Studies Travel Grant — 2019
- Student Activity Fee Committee Travel Grant — 2019
- CSEE Balaji Krithikaivasan Memorial Travel Grant — 2019
- IEEE Standards Education Grant — 2012