

Raza Ahmad

☎ +1 (650) 509-0502 • ✉ sra0nasir@gmail.com

Education

DePaul University, Chicago IL <i>PhD in Computer & Information Sciences</i>	2024–Present
DePaul University, Chicago IL <i>Master in Computer Science</i>	2021–2023
FAST - National University of Computer and Emerging Sciences, Lahore Pakistan <i>Bachelor in Computer Science</i>	2011–2015

Experience

DePaul University, Chicago IL <i>Research Assistant</i> Research in the area of reproducibility in large-scale systems. ○ Develop systems and interfaces for auditing and executing computer programs in a portable, containerized, and reproducible environment. ○ Apply state-of-the art machine learning models for efficient solutions to intractable computer systems problems. ○ Contribution to several open source tools and projects to improve re-execution and repeatability of computational programs. ○ Developing techniques to improve computational reproducibility using machine learning. ○ Work on designing and improving software for computation reproducibility. ○ Collaborate with domain scientists from High Energy Physics for improving reproducibility in computational experiments.	Sep 2021–Present
DePaul University, Chicago IL <i>Research Engineer</i> Research in the area of computational reproducibility. ○ Develop novel techniques and algorithms to containerize and reproduce scientific workflows in interactive notebook environments. ○ Develop systems and interfaces for auditing and executing computer programs in a portable, containerized, and reproducible environment. ○ Lead the development and managed several open-source system used by research scientists to improve re-execution and repeatability of computational programs ○ Integrating reproducible software to share batch scripts and interactive notebooks across multiple programming languages and compute environments. ○ Work on designing and improving software for computation reproducibility. ○ Collaborate with domain scientists from Hydrology for improving reproducibility in computational experiments.	Mar 2020–Mar 2021
SRI International, Menlo Park CA <i>International Research Fellow</i> Research in the area of provenance management. ○ Revised provenance querying architecture in a distributed peer-to-peer setting.	Oct 2016–Mar 2020

- Designed and developed query interfaces for querying provenance from several relational and graph databases.
- Created flexible integration of Neo4j, PostgreSQL, MySQL, H2, BerkeleyDB, LevelDB, SSDB, Redis databases for SPADE storage.
- Designed methods to improve real-time ingestion of provenance metadata in PostgreSQL.
- Optimized compression of provenance graphs in a streaming setting with improvements to performance and compression over the baseline web compression techniques.
- Designed and implemented the system for measuring provenance response integrity in distributed environments through discrepancy detection.
- Introduced cache management of provenance responses in peer-to-peer architectures.
- Added privacy enhancements to provenance querying using sanitization and attributed-based encryption.

Lahore University of Management Sciences (LUMS), Lahore

Aug 2013–Sep 2016

Research Associate

Research in areas of applied machine learning, data analytics, NLP and information retrieval.

- Performed aspect based sentiment analysis on the coverage of major US news papers on the candidates of US Presidential election campaign 2016 to identify reporting trend and differential coverage for various demographics.
- Designed novel techniques to detect automatically spun content on the web using stylometric analysis of text.
- Designed a tree-based algorithm to detect Tuberculosis patients at an early stage on the basis of MFI values of various antibodies collected from their blood serum.
- Analyzed data from nine traffic exchange networks and analyzed them as a means for spreading malware on the internet.
- Studied browsing behavior patterns of college students and analyzed several correlations between internet usage and academic performance.
- Studied interaction effects on charge transport in simulations of heterogeneous networks of carbon nanotubes. Applied machine learning techniques to predict the bulk DC conductance of the network with high accuracy.
- Measured politeness in text using deep recursive neural networks on the Stanford politeness corpus.
- Developed and analyzed several deep learning techniques to classify tau tau decay of a Higgs Boson particle from the background, as part of the Higgs Boson machine learning challenge by CERN.

Deezon Ltd., Lahore

May 2012–Dec 2013

Project Manager & SEO Consultant

Managed projects in search engine optimization, online marketing, blog and social media development, link building, keyword analysis, vertical search and search analytics.

Publications

- T. Azaz, R. Ahmad, A. Adibifar, D. Thain, and T. Malik, "Fidelis: Exploiting Task Similarity for Low-overhead Distributed Workflow Audit and Replay," Submitted to the 27th ACM International Middleware Conference, 2026.
- T. Azaz, R. Ahmad, M. S. Islam, D. Thain, and T. Malik, "Efficiently Reproducing Distributed Workflows in Notebook-based Systems," Proceedings of The 26th IEEE International Symposium on Cluster, Cloud, and Internet Computing (CCGrid), 2026, pp. 74-83, 10.1109/CCGrid68966.2026.00017.

- M.S. Islam, T. Azaz, R. Ahmad, A S M S. Hossain, F. Baig, S. Wang, K. Lannon, T. Malik, D. Thain, 2025, Backpacks for Notebooks: Enabling Containerized Notebook Workflows in Distributed Environments, *Proceedings of The 21st IEEE International Conference on eScience (eScience)*, 2025, pp. 169-177, <http://doi.org/10.1109/eScience65000.2025.00028>.
- R. Ahmad, N. N. Manne, and T. Malik, 2025, Improving Reproducibility of Interactive Notebooks using Application Virtualization, *Future Generation Computer Systems*, vol. 175, p. 108043, 10.1016/j.future.2025.108043
- R. Ahmad, H. Jung, Y. Nakamura, and T. Malik, 2024, Accurate Path Prediction of Provenance Traces, *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM)*, 2024, p. 3617–3621, 10.1145/3627673.3679872.
- Y. Choi, B. Roy, J. Nguyen, R. Ahmad, I. Maghami, A. Nassar, Z. Li, A. M. Castronova, T. Malik, S. Wang, J. L. Goodall, 2023, Comparing Containerization-based Approaches for Reproducible Computational Modeling of Environmental Systems, in *Environmental Modelling & Software*, vol. 167, p. 105760, 10.1016/j.envsoft.2023.105760.
- R. Ahmad, A. Modi, E. Jung, C. Garcia, H. Irshad, and A. Gehani, 2023, Analyzing Discrepancies in Whole Network Provenance, *Seventeenth IFIP Working Group 11.10 International Conference on Critical Infrastructure Protection (ICCIP)*, March 2023.
- R. Ahmad, Y.D. Choi, J.L. Goodall, D. Tarboton, A. Nassar, and T. Malik, 2023, Improving reproducibility of geoscience models with Sciunit, in X. Ma, M. Mookerjee, L. Hsu, and D. Hills, eds., *Recent Advancement in Geoinformatics and Data Science: Geological Society of America Special Paper 558*, p. 1–12, 10.1130/2022.2558(07).
- R. Ahmad, N. N. Manne and T. Malik, "Reproducible Notebook Containers using Application Virtualization," 2022, *IEEE 18th International Conference on e-Science (e-Science)*, 2022, pp. 1-10, doi: 10.1109/eScience55777.2022.00015. Best Paper Award.
- A. Gehani, R. Ahmad, H. Irshad, J. Zhu, and J. Patel, "Digging into Big Provenance (with SPADE)", in *Communications of the ACM vol. 64 no. 12, December 2021*
- Y. Choi, J. Goodall, R. Ahmad, T. Malik, and D. Tarboton, "An Approach for Open and Reproducible Hydrological Modeling using Sciunit and HydroShare", in *EGU General Assembly 2021*
- A. Gehani, R. Ahmad, and H. Irshad, "Practical Provenance Privacy Protection", in *13th International USENIX Workshop on Theory and Practice of Provenance (TaPP)*, *ProvenanceWeek 2020*
- Y. Nakamura, R. Ahmad, T. Malik, "Content-defined Container Delivery", in *International Conference On High Performance Computing, Data, And Analytics (HiPC) 2020*
- R. Ahmad, Y. Nakamura, N. N. Manne, T. Malik, "PROV-CRT: Provenance Support for Container Runtimes", in *12th International USENIX Workshop on Theory and Practice of Provenance (TaPP)*, *ProvenanceWeek 2020*
- R. Ahmad, E. Jung, C. S. Garcia, H. Irshad and A. Gehani, "Discrepancy Detection in Whole Network Provenance", in *12th International USENIX Workshop on Theory and Practice of Provenance (TaPP)*, *ProvenanceWeek 2020*
- R. Ahmad, M. Bru and A. Gehani, "Streaming Provenance Compression", in *International Provenance and Annotation Workshop (IPAW) 2018: Lecture Notes in Computer Science vol 11017. Springer, Cham.*
- U. Shahid, S. Farooqi, R. Ahmad, Z. Shafiq, P. Srinivasan and F. Zaffar, "Accurate Detection of Automatically Spun Content via Stylometric Analysis", in *IEEE International Conference on Data Mining (ICDM) 2017.*
- R. Ahmad, A. Pervaiz, H. Mannan and F. Zaffar, "Aspect Based Sentiment Analysis for Large Documents with Applications to US Presidential Elections 2016", in *The Americas Conference on Information Systems (AMCIS)*, 2017.

- S. F. Hussainy, R. Ahmad, A. Khaliq, A. Zaffar, F. Zaffar and I. H. Khan, in "Decision-tree Inspired Classification Algorithm to Detect Tuberculosis (TB)", in *Pacific-Asia Conference on Information Systems (PACIS) 2017*.
- S. Yousuf, U. Iqbal, R. Ahmad, S. Farooqi, F. Zaffar and Z. Shafiq, "Malware Slums: Measurement and Analysis of Malware on Traffic Exchanges", in *Dependable Systems and Networks (DSN) 2016*.

Professional Services

- Reviewer for Journal Distributed and Parallel Databases
- Reviewer for Journal Future Generation Computer Systems Journal