Ranjana Rajendran

Seattle, WA 98125

Email: ranjana.rajendran@gmail.com
Ph: [Mobile] (831) 332 9892
Skype: ranjana_rajendran
Work Authorization in US: US citizen
Github: ranjana_rajendran, ranjana-altiscale
Linkedin Profile

Education

MS, Computer Science, ***University of California Santa Cruz***, September '10 - March '13

Course Work: Analysis of Algorithms, Advanced Operating Systems, Security in Computer Systems, Computer Graphics, Programming Languages, Data Mining, Machine Learning, Mathematical Statistics, Design & Implementation of Database Systems

BTech, Computer Science & Engineering, ***Cochin University of Science and Technology, India***, September '02 - May '06 (First Class with Distinction & Honours)

Publications

Ranjana Rajendran, Aditya D Telang, Deepak P, Prasad M Deshpande, "Detecting localized homogeneous anomalies over spatio-temporal data", (Data Mining and Knowledge Discovery, Volume 28), *Internship project with IBM India Research Lab*

Ranjana Rajendran, Ethan L Miller, Darrell D E Long, "Horus: Fine-Grained Encryption-Based Security for High Performance Petascale Storage",*Proceedings of the 6th Parallel Data Storage Workshop (PDSW '11)*, Nov '11

Patent

US-20250053985-A1 : Method and System for Cryptocurrency Fraud Detection, result of a hackathon while at Mastercard. Pub. Date: Feb 13, 2025

Awards & Recognitions

Spot Bonus Award at Cloudera for best performance in team during quarter Jan-March 2014.
UC **Regents Graduate Fellowship** 2010-2011, Spring 2012
UC Full Non-Resident Tuition Fees Waiver 2010-2011
District Merit Scholarship, 2000-2006 by Department of Collegiate Education, Kerala, India
Ranked **17 among ~ 500000** students in Secondary School Leaving Certificate Exam with 95.67% marks
Ranked **770 among more than ~ 40000** students in Kerala State Engineering Entrance Exam, India
GRE Test Score: (**Quantitative Analysis :800/800** Analytical Reasoning: 5.0/6.0 Verbal Analysis: 450/800)

Work Experience

Hit-Reset August '23 - Present

Strategic pause to refresh ML Engineering skills and complete personal projects.

- Traditional ML Projects:
  - Supervised: Predict time to eruption and pressure at eruption of volcano. Regression models / Correlation heatmap / Principal Component Analysis (PCA) / Exploratory Data Analysis (EDA) / Data Visualization / Data Normalization / Automated model benchmarking using LazyPredict, scikit-learn, and Python with residual analysis
  - Unsupervised: Detect clusters of users in Facebook social network dataset. EDA, Data Visualization, Clustering algorithms (KMeans, Agglomerative / DBScan / Birch / Gaussian Mixture Models / Affinity Propagation / Self Organizing Map / OPTIC), Graph-based embeddings (node2vec with networkx), Correlation heatmap / PCA
- Deep Learning Projects:
  - Demonstration of variational auto encoder using Fashion MNIST dataset
  - Supervised, Natural Language Processing (NLP): Detect spoilers among reviews about movies. EDA / Visualization / Histograms / Text normalization and processing / lemmatization / BERT / Bi-Directional LSTM / Longformer / Low-Rank Optimization / Early stopping / Batch processing / Two-tower model / Accuracy, Precision vs Recall, F1 score / Calibration plots.
- Generative AI Projects:
  - NLP : LLM-based chatbot for answering questions regarding the NVidia SDK enabling accurate multi-turn question answering with latency under 1s. Web crawling / Text pre-processing / Retrieval Augmented Generation (RAG) / Vector Database (FAISS) / Langchain / agents
  - NLP : LLM-based conversational AI chatbot resume coach (in-progress). End to End MLOps pipeline / React + AWS S3 web hosting / Langchain / agents / Deployment using FastAPI + Kubernetes

Lead Software Engineer, *Ekata-Mastercard (Identity Verification)*, Seattle August '22 - August '23

- Collaborated closely with product managers to align identity workflows and PII suppression systems, powering ML-based identity scoring, with compliance, privacy, and product requirements.
- Engineered distributed ETL pipelines over terabytes of identity data using Spark on Databricks & EMR, orchestrated via Airflow, enabling large-scale feature extraction, and integrated observability with automated data quality checks.
- Managed a Redis-backed identity graph with dynamic suppression, stateful updates, and low-latency digest lookups supporting real-time ML fraud scoring.
- Co-invented a patented fraud detection system (US-20250053985-A1), enabling multi-signal reasoning and real-time anomaly detection, integrating ML models with high-scale identity pipelines for fraud prevention.

System Development Engineer, *Amazon Web Services*, Seattle July '17 - August '22

Release engineer for *Elastic Map Reduce* (EMR). Specifically, the activities included :

- Monthly build and release of EMR : Resolve integration issues between ML and big data frameworks(Hive, Spark, Presto, etc.) and the operating system; develop automated tools in Java for deployment of release artifacts into production. Write ad-hoc scripts to deploy new releases or correct bugs in existing releases in production.
- Implement and maintain an integration testing framework (200+ automated tests) in Java covering all EMR-supported data processing applications, ensuring reliability for ML feature engineering and data pipelines.
- Design and integrate new compute types (EC2) and applications into EMR, significantly improving distributed model training, large-scale data preprocessing, and ML workload performance.
- Provide technical documentation, collaborate with internal and external stakeholders, including product managers to highlight new EMR features, ensuring usability for data scientists and ML engineers leveraging Spark MLlib and custom ML pipelines.

Senior Developer Support Engineer, *Qubole*, Santa Clara April '16 - May '17

Optimized large-scale data and feature engineering pipelines on Qubole's Hadoop-as-a-Service platform using MapReduce, Spark, Hive, and SparkSQL, enabling downstream ML model training and analytics. Achieved up to 70% faster runtimes and 90% fewer pipeline failures, improving reliability of ML workflows. Collaborated with engineering and product teams to drive platform enhancements based on real-world data science and ML challenges.

Hadoop Engineer, *Altiscale*, Palo Alto February '15 - March '16

Handled customer feature requests and support tickets for production *Map-Reduce / Spark jobs*, ML data pipelines, ETL, and security (*Kerberos*). Specifically, the activities included:

- Troubleshoot, optimize, and tune customer's Hive, Spark, and Spark SQL applications.
- Java performance and memory profiling using Eclipse memory analyzer tool for running Java processes.
- Convert customer ML pipeline needs to feature requests. Author knowledge-base articles with best practices.

Solution Architect, *Cloudera*, San Francisco September '13 - January '15

ML Ops Engineer – Hadoop Ecosystem Solutions

- Deployed and managed Cloudera dist. of Hadoop(CDH) & Cloudera Manager(CM) on up to 45 nodes (provisioned on AWS EC2 instances or on-premise private cloud such as CloudStack or bare-metal compute nodes), supporting large-scale feature engineering, ML training pipelines, and inference workloads.
- Setup back-end MySql databases with master-slave replication for the components of CM & CDH.
- Setup kerberos with CDH (Local KDC (with back-up KDC), local KDC with one-way cross realm trust to Active Directory (AD), & direct to AD)
- Designed batch & real time data pipelines using Spark, Hive, & Impala on HDFS, involving minimal duplication of data, enabling automated data ingestion and transformation for ML model training under strict SLAs.
- Integrated real-time data sources (e.g., Twitter) using Flume and RDBMS data ingestion via Sqoop, enriching data warehouses for continuous ML retraining and inference pipelines.
- Developed and maintained Hive-based data warehouses with role-based access control (Sentry) for secure storage and retrieval of ML training datasets.
- Automated end-to-end ML pipeline orchestration using Oozie, from data ingestion to model-ready datasets, with integrated monitoring and dashboarding (Tableau with Hive, Impala & Spark SQL) for performance tracking.

Certifications: Cloudera Certified Developer for Apache Hadoop (CCDH), Cloudera Certified Administrator for Apache Hadoop (CCAH), Cloudera Certified Specialist in Apache HBase (CCSHB)

Software Engineer, *Gap Inc via Nisum Technologies*, San Francisco June '13 - September '13

- Developing testing frameworks in Ruby using *Watir & Taza frameworks*.
- Release testing and debugging for an order management system implemented in Java with several tools including SoapUI.

Research Intern - IBM India Research Lab, Bangalore, July '12 - December '12

Researched, developed, and implemented an ML-based anomaly detection system (Java, R, Unix shell) for spatial and spatio-temporal datasets, resulting in a publication in Data Mining and Knowledge Discovery (Vol. 28).

Graduate Student Researcher, *Storage System Research Centre, UC Santa Cruz* January '11 - December '11

Researched, developed, and implemented an encryption model for the Linux based distributed file system (Ceph), which resulted in a publication at a prestigious workshop.

Teaching Assistant, *UC Santa Cruz* Fall '10, Fall '11, Spring '12

Teaching assistant for undergraduate courses in Computer Programming & Unix systems, Mathematics (Precalculus), and Business Information Systems. Examination grader and invigilator for graduate and undergraduate courses. Assessed students' comprehension, prepared examinations accordingly, gave feedback, and evaluated their performance.

MTS - Software Engineer, Lucent Technologies, *R & D Centre, Bangalore* Dec '06 - Dec'08

Performed duties as module leader, engineered innovative solutions to meet customer requirements, enhancing system efficiency by 30% of an OA & M GUI for Radio Access Networks, and trained team members in various J2EE technologies.

Software Engineer, *Wipro Technologies, Cochin* August '06 - December '06

Designed and implemented fixes for middleware components in digital television set-top boxes.

Selected Seminar Projects

Data Mining: Performed an empirical analysis on *frequent sequence mining algorithms* for datasets both dense and sparse with existential uncertainty.

Machine Learning: Speech recognition by training a back propagation artificial neural network by extracting features from sound. Achieved 60% speed up by parallelizing the algorithm on 8 nodes of a Beowulf Linux cluster.

Information Flow Security and Programming Languages: Enhanced a monadic library in a Haskell web browser and performed analysis on various information flow security vulnerabilities.

Operating Systems: Experimented with a scalable key management scheme for eCryptfs. Modified the kernel modules for ecryptfs to pick up keys generated from a root key while encrypting a large file in blocks.

Technical Skills

Programming: Python, Java, C, C++, Scala, R, Ruby, UNIX shell scripting, SQL, assembly language, object-oriented programming

Machine Learning: Tensorflow, PyTorch, Sci-Kit Learn, HuggingFace library, OpenAI, Langchain, agents, RAG, vector databases(FAISS), Transformers, LLM, Pytorch Lightning, Multi-turn dialogue, Reinforcement Learning

Data Visualization & Interactive Apps: Streamlit, Plotly, Dash, Matplotlib, FastAPI, Flask

Web & API Development: FastAPI, Flask, Spring MVC, JSP, REST, SOAP, MySQL, Apache Tomcat, NginX

Continuous Integration Services (Build & Release): Apache Jenkins, AWS CodeBuild

Containerization & Orchestration: Docker, Kubernetes, AWS ECS

Databases: MySQL, OracleDB, NoSQL distributed databases (HBASE, Cassandra), TOAD tool, SQL, PIG, HIVE, Impala, Spark SQL, Redis

Distributed File Systems: HDFS, Ceph

Distributed Systems: Multithreaded programming in Java and concurrency management.

Networking: TCP, UDP, ARP, DNS, Dynamic routing (OSPF & BGP), OSI model, Service (*Apache HTTP, SNMP, application-specific daemon design*).

Workflow and job scheduling: Oozie and Airflow (for Big Data applications), Crontab

Version Control System: Git, Clearcase, SVN

Open Source project contributions: HBase, Hive, Pig

Testing Frameworks: JUnit, MRUnit, TAZA on Ruby, BDD based Cucumber, Ruby on Rails, Rake, Apache Bigtop

Security: Kerberos (*perimeter security*), Sentry (*role based access control*), Firewalls (*IPTable rules*) etc.

Software Engineering: OOAD using UML and class diagrams, design patterns, SOA, Waterfall, Agile and BDD methodologies.

Computer Applications: HTML, CSS, XML, JSP, J2EE, web application development (servlets, MySQL, Spring MVC, JDBC, Hibernate, SOA based high volume scalable web services (SOAP), REST, Apache Tomcat and nginX), SoapUI, IBM WebSphere MQ, Weka datamining tool, ipython, most common productivity packages (for Windows and Linux platforms), Vim, application specific IDEs (RStudio, Eclipse, Netbeans, QT-OpenGL, Blender), Hadoop / Mapreduce, SAS, Excel for analytics TEX (LATEX, BIBTEX).

Language Skills

English : Fluent in reading, writing, and speaking. My education was in English. (TOEFL: 112/120)
Hindi : Fluent in reading, writing, and speaking.
Malayalam (native speaker) : Fluent in reading, writing, and speaking.