Mansi Borole

mansi.borole@rutgers.edu | +1 7323745143 | linkedin/mansi-borole | github/MansiGit | mansigit.github.io

Seeking full-time opportunities in Software Engineering / Data Engineering starting May 2024

EDUCATION

Rutgers University, New Brunswick, New Jersey — **Aug 2022 - May 2024**
Master of Science in Computer Science — *GPA: 3.91/4.0*

University of Pune, India — **Aug 2015 – Jul 2019**
Bachelor of Engineering in Computer Science — *GPA: 8.61/10*

SKILLS

Languages: Python, SQL, HTML, CSS, JavaScript, Shell Scripting, YAML
Cloud & Databases platforms: PostgreSQL, MySQL, MongoDB, AWS (S3, EC2, Lambda, RDS, IAM, API Gateway, CloudFormation)
Big Data Hadoop: HDFS, Apache Spark, Pyspark, Scoop, ETL, AWS, Enterprise Data Ingestion and Warehousing
Tools & Frameworks: Git, Serverless, Boto, Node.js, React, pytest, Tableau, Grafana, Microsoft Excel, JIRA
Other Skills: Agile Workflow, CI/CD, Problem-solving, Technical writing, Presentation

PROFESSIONAL WORK EXPERIENCE

Michigan Health Information Network (MiHIN) | *Software Development Engineer Intern* — *May 2023 – Present*

- Worked as a backend developer on a state sponsored project “Social Determinants of health”. Developed and deployed AWS Lambda functions using **Python** and **Boto** library for processing, validating, and storing Electronic Health Records in AWS RDS.
- Created a mock database interface and implemented **unit tests** to validate the functionality of the database drivers. Took a **test-driven design** approach to ensure code quality and coverage.
- Optimized **Postgres** database performance using indexes and views, resulting in **80% faster processing**.
- Created a PoC Project to evaluate **Serverless architecture** using AWS Lambda, SNS, CloudFormation and presented it to the team.

HSBC Bank | *Software Engineer (Big Data Analytics)* — *Jul 2019 – Aug 2022*

- Built a real-time alerting, monitoring, data cataloguing system "BigData360" for 58 Hadoop dev and production clusters using **Grafana** and **Prometheus** by collecting metrics from **Cloudera APIs**. Achieved 15% reduction in mean time to repair.
- Developed ETL pipelines and scripts in python to archive cold data. Leveraged Erasure coding to achieve savings of 50% in total storage space.
- Built a pipeline to feed live tenant usage metrics from the big data estate to aid in Capacity Planning and Billing. Generated automated cost reports using **Tableau, MySQL** and saving ~150 man-hours.
- **Cluster Virtualization:** Automated migration of Hadoop Master services (NameNodes, YARN, Solr, Zookeeper, Hive) from on-premise servers to virtual machines in **ESXi hypervisor**, achieving a **cost avoidance of $700k** on buying new expensive servers while saving 400 man hours in maintenance upgrades.
- **Ansible Automation**: Created Ansible playbooks and python scripts to build RedHat servers with Hadoop capabilities.

RELEVANT PROJECTS

MovieSync — **AWS, Node.js, React, JavaScript, MongoDB**

- Watch movies in sync with up to 50 friends! Created a MERN Stack project using EC2, S3, CloudFront for synchronized playback of video files. It utilized Express.js as the application server, MongoDB for the database, and incorporated Node.js for server-side JavaScript execution, and static assets in S3.
- Utilized **Selenium** for comprehensive testing of 50 concurrent users, evaluating video synchronization, buffering rate, and bitrate latency. Project Link

Exploring Business Intelligence - Data Visualization — **Tableau Public Portfolio**

- Analyzed and created interactive dashboards for case studies in Credit Card Complaints Trends, Netflix movies analysis, Salary Insights in NY State showcasing the Trends, Predictions, Density maps, Scatter Plots and Timeseries Charts.
- Used features such as filters, drilldowns, and tool tips to enhance the UX and make it easy for them to interact with the data.

Machine Learning / Data Mining — **Pytorch, Tensorflow, Keras, Pandas**

- Developed a Groceries Recommendation System with collaborative filtering techniques (Matrix Factorization, Slope One, SVD). Evaluated accuracy (RMSE, MAE) and performance (Precision, Recall, F1 Scores). Report: [link].
- Image Similarity Search using a pretrained Convolutional Neural Network (CNN) model and then using KNN classifier to find similar images. Achieved an accuracy of 98%.

CERTIFICATIONS & ACHIEVEMENTS

- **Grace Hopper Scholar (2023), AWS Cloud Practitioner (2023), Tableau Certified Desktop Specialist (2022), Certified Associate in Python Programming (2021) PCAP**