

Mansi Borole

mansi.borole@rutgers.edu | New Brunswick, NJ | +1 7323745143 | [linkedin/mansi-borole](https://www.linkedin.com/in/mansi-borole) | [github/MansiGit](https://github.com/MansiGit) | mansigit.github.io

EDUCATION

Rutgers University, New Brunswick, New Jersey

Master of Science in Computer Science

Aug 2022 - May 2024

GPA: 3.91/4.0

University of Pune, India

Bachelor of Engineering in Computer Science

Aug 2015 - Jul 2019

GPA: 8.61/10

TECHNICAL SKILLS

Languages: Python, SQL, HTML, CSS, Java, JavaScript, UNIX / Linux Shell Scripting, YAML

Cloud & Databases platforms: PostgreSQL, MySQL, MongoDB, AWS (S3, EC2, Lambda, RDS, API Gateway, CloudFormation)

Tools & Frameworks: Git, Serverless, Boto, Node.js, React, Pytest, RESTful, Tableau, Grafana, Microsoft Suite, JIRA, Agile, CI/CD

PROFESSIONAL WORK EXPERIENCE

Michigan Health Information Network | Software Development Engineer Intern

May 2023 - Present

- Contributed to backend dev of a state-sponsored project, "Social Determinants of Health." Developed and deployed **AWS Lambda** functions using Python Boto library for processing, validating, and storing Electronic Health Records in **AWS RDS**.
- Designed and developed **unit tests** and set up mock objects to ensure effective **Test-driven development**; improved code coverage from 20% to 80%.
- Developed REST API using **AWS API Gateway**, **Chalice framework** to manage file submissions and patient screening records.
- Optimized PostgreSQL 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 showcased it to the leadership.

HSBC Bank | Software Engineer (Big Data Analytics)

Jul 2019 - Aug 2022

- Built "BigData360", a real-time alerting, monitoring system and designed dashboards for 80+ Hadoop dev and prod clusters using **Grafana** and **Prometheus** by collecting metrics from **Cloudera APIs** and **Node Exporters**.
- Generated monthly cost reports to support Big Data Capacity Planning and forecasting demand. Built a self-serving data catalog tool to analyze real-time tenant usage metrics and trends using **Airflow** and **MySQL**, saving ~150 work hours.
- Enhanced cluster efficiency by **virtualizing control nodes**, migrating critical Hadoop Master services (NameNodes, YARN, Solr, Zookeeper, Hive) to **ESXi virtual machines**. This saved \$700k on new server costs and reduced maintenance hours by 400.
- Developed ETL pipelines and scripts in python to archive cold data. Set up HDFS Erasure coding policies to achieve savings of 50% in total storage space.
- Created **automation using Ansible** to streamline the provisioning of RedHat servers with Cloudera(CDP) Hadoop capabilities.
- Mentored and onboarded new SWEs. Conducted knowledge sharing sessions with my team.

RELEVANT PROJECTS

"MovieSync"

AWS, Node.js, React, JavaScript, MongoDB

- Created a MERN stack app for synchronized movie playback using **Dynamic Adaptive Streaming over HTTP (DASH)** protocol.
- Deployed application on AWS EC2 and used S3 and CloudFront for storage and latency improvements.
- Utilized Selenium for comprehensive testing among 50 concurrent users to study buffering rate, bitrate, latency metrics. [\[link\]](#)

Groceries Recommendation system evaluation

Pytorch, Tensorflow, Keras, Pandas

- Designed Groceries Recommendation Systems using collaborative filtering techniques, including Matrix Factorization, Slope One, and SVD, to analyze a dataset of 5 million reviews from the Amazon Groceries and Gourmet Foods dataset. [\[link\]](#)
- Evaluated system accuracy using metrics like RMSE and MAE, and assessed performance through Precision, Recall, and F1 Scores. Report: [\[link\]](#).

Data analysis and Price Prediction of Airbnb listings data

Sklearn, Pandas, Matplotlib

- [Airbnb Price Prediction](#) Used linear regression, XGBoost and multivariate adaptive regression splines (MARS) to predict prices based on city, availability, reviews per month and room type.

Airflow Batch Data Processing

Apache Airflow, Docker, Python, DAGs

- Created a batch data processing pipeline to transform and load Hotel booking data stream into an SQLite database using Apache Airflow and Python. Containerized the application using Docker to enable easy deployment. [\[link\]](#)

CERTIFICATIONS & ACHIEVEMENTS

- **AWS Certified Cloud Practitioner (2023)** - **Certified Associate in Python Programming (2021) PCAP**
- **HSBC Star Performer Award (2021,2022)** - **Smart India Hackathon (2018) Finalist**