

# Mansi Borole

[mrborole@gmail.com](mailto:mrborole@gmail.com) | 732-374-5143 | [linkedin.com/in/mansi-borole](https://www.linkedin.com/in/mansi-borole) | [github.com/MansiGit](https://github.com/MansiGit) | New Brunswick, NJ

## EDUCATION

**Rutgers University, Master of Science, Computer Science**, New Brunswick, NJ, USA  
Coursework: Algorithms, Mathematical foundations of Data Science, Artificial Intelligence

Sep 2022 - May 2024  
**GPA: 3.83/4.0**

**University of Pune, Bachelor of Engineering, Computer Science**  
Coursework: Data Structures, Database Management, Web Technology, Machine Learning

Aug 2015 – Jul 2019  
**CGPA: 8.61/10**

## SKILLS

**Programming Proficiency:** Python, SQL

**Big Data Hadoop:** HDFS, Apache Spark, Pyspark, NiFi, YARN, Enterprise Data Warehousing

**Database platforms:** PostgreSQL, MySQL, Apache Hive

**Data Visualization:** Tableau, Grafana, QlikSense, Microsoft Excel (Pivot Tables, Pivot Charts, V lookups)

**Skills:** Hadoop Administration, DevOps, Agile Development, Problem solving, Big Data Analytics, Stakeholder Management

## PROFESSIONAL WORK EXPERIENCE

**HSBC Bank | Software Engineer (Big Data)**

July 2019 – Aug 2022

- **Cluster Virtualization:** Migrated Hadoop Master services (Name nodes, YARN, Solr, Zookeeper, Hive) from on-prem servers to Virtual Machines, achieving a **cost avoidance of \$0.7M** on buying new expensive servers while saving 400 man hours in maintenance upgrades using Ansible Automation.
- **Capacity Management:** Performed Hadoop/Linux administration provisioning space on clusters by building and decommissioning servers, Node repurposes from one cluster to another and installing Hadoop components and services on the nodes
- **Nifi Dataflow:** Set up a one-stop dashboard BigData360 and data-flows to monitor Hadoop services and alerts on clusters using Ambari Metrics/ Cloudera Metrics APIs that **reduced the mean time to repair failures by 1 hour**
- **Ansible Automation:** Created Ansible playbooks and **python scripts** to get daily FS image from clusters and find the daily organic growth of client's HDFS data on the clusters
- Created a MySQL tool to generate Chargeback reports to determine the cost of Physical Hardware, Licenses, and Supporting Staff for each of **2000+ servers** in the estate based on resource usage and used Qlik Sense to publish the reports.
- Wrote housekeeping scripts that would reduce replication factor of cold data on the cluster effectively saving 45 TB/month

**HSBC Bank | Communications Leader**

July 2019 – August 2022

- Lead a team of 4 colleagues delivered high-quality, impactful quarterly newsletters highlighting project success stories and KPIs from the 'Group Data Technology' Department, closely working with Ex Co members in US, UK and Asia-Pacific regions
- Published team's quarterly achievements and highlights via Newsletters to the entire department

## CERTIFICATIONS & ACHIEVEMENTS

- **Tableau** Certified Desktop Specialist 2022
- Certified Associate in **Python** Programming 2021
- IBM Foundation Blockchain Developer 2019

## RELEVANT PROJECTS

**Fast Trajectory Replanning using AI Agent**

2022

- Implemented and analyzed paths generated by various flavors of A-star search algorithms like Repeated A-Star, forward/backward A-star, Adaptive A-star.

**Tableau Data Visualization**

2021

- Developed COVID 19 Running bar chart for demonstrating country wise positive cases over the span of last year using Tableau Desktop
- Incorporated live data sources from data.world and published the chart on [Git](#)

**Machine Learning**

2022

- [Image Classifier](#) Built a naive Bayes classifier, a perceptron classifier and a KNN classifier to classify handwritten digits (OCR) and recognize faces. Achieved an accuracy of 96%
- [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%