

MOHIT SAI GUTHA

mohit.sai6@gmail.com • (617) 415-3812 • <https://mohitsai.github.io> • LinkedIn: www.linkedin.com/in/mohitsaigutha

EDUCATION

Boston University | Boston, USA | GPA: 3.6/4.0 Sep 2023 - Jan 2025

Master of Science in Computer Science

Relevant Coursework: Data Engineering @ Scale, Cloud Computing with GCP, Data Science Tools, Database Architecture

PES University | Bangalore, India | GPA: 7.7/10.0 Aug 2018 - May 2022

Bachelor of Technology, Computer Science and Engineering

Relevant Coursework: Machine Learning, Python Application Programming, DevOps and Site Reliability

SKILLS

Programming & Big Data: Python, Java, SQL, C++, R, Apache Spark, Hadoop, Kafka, Delta Lake

Cloud & DevOps: Azure Data Factory, Google Cloud Dataflow, AWS, Docker, Kubernetes, Terraform, Jenkins, GitHub Actions

Data Modeling & Storage: SQL, Cloud SQL, Amazon Aurora, Snowflake, PostgreSQL, NoSQL, Redis

Machine Learning & Analytics: PySpark ML, Azure Databricks ML, Scikit-learn, Power BI, Tableau, Google BigQuery

PROJECTS

Epidemic Engine – Healthcare Data Pipeline Jan 2024 - May 2024

- Engineered an ETL pipeline for healthcare event data, enabling real-time ingestion via Apache Kafka and batch processing with Hadoop MapReduce, handling more than 1 million records
- Enabled fully containerized execution with Docker, ensuring modularity and easy deployment, while Apache Airflow managed data workflows with automated task scheduling and event-driven orchestration
- Built predictive models using Spark ML and PySpark, leveraging KMeans clustering for health risk detection, achieving 93.7% accuracy in identifying potential epidemic outbreaks and improving public health monitoring

Future of Hiring – Job Market Pipeline with Azure Databricks & Power BI Jan 2025 - Feb 2025

- Implemented scalable data extraction and ingestion using Azure Functions to fetch job market data from Adzuna API, securely storing raw data in Azure Data Lake Storage Gen2, with Azure Key Vault managing sensitive credentials
- Built an ETL pipeline with Azure Data Factory to automate workflow execution, leveraging Delta Lake for incremental data loading and Azure Databricks with PySpark for efficient data transformation and structured analysis
- Developed predictive analytics models using Azure Databricks ML, training on historical job data to forecast salary trends and hiring demand, and integrated Power BI to create interactive dashboards for real-time workforce insights

Google Analytics Simulation – Google Cloud Platform Sep 2024 - Dec 2024

- Deployed 10,000 simulated webpages on Google Cloud Storage, served via a Virtual Machine, tracking 100,000+ user requests in Cloud SQL with metadata like location, age, and gender for traffic analysis and behavioral insights
- Computed PageRank for 10,000 webpages using Google Cloud Dataflow for real-time streaming analytics, identifying high-authority pages, optimizing web performance metrics with a cloud-native big data approach
- Trained an ML model to predict user demographics based on webpage requests, achieving 99.7% accuracy, and deployed the system in Google Kubernetes Engine using Google Deployment Manager for scalable orchestration

WORK EXPERIENCE

Genpact | Bangalore, India Oct 2022 - Jun 2023

Software Development Consultant

- Architected and deployed Java-based Spring Boot microservices, modernizing legacy monolithic architecture for a top healthcare client, integrating with AWS SQS for reduced latency, driving 60% faster transaction processing
- Collaborated with a team of 12 on migrating legacy mainframe systems to AWS, leveraging AWS DMS, Amazon Aurora, and EC2 Spot Instances, achieving a 60% reduction in compute costs
- Automated CI/CD pipelines with Jenkins and GitHub Actions, decreasing deployment time by 50%, improving software reliability, and supporting Agile-driven iterative releases

Johnson Controls Inc | Bangalore, India Jan 2022 - Jun 2022

Business Intelligence Analyst Intern

- Developed a React-based Power BI dashboard to visualize the company's network infrastructure transition, integrating real-time analytics and improving project efficiency by 20% through data-driven insights
- Led a globally diverse team of 6 interns for the Future Leaders Program, designing sustainability-focused financial solutions and reaching the finals of the Sustainable Innovation Competition