Pranav Kuramkote Sudhir

Email:kuramkotesudhir.p@northeastern.edu | Mob: +18574659377 | location: Boston
Linkedin: https://www.linkedin.com/in/pranav-kuramkote-sudhir-a46359168/ | Github: https://github.com/PranavKuramkoteSudhir

---

Professional Summary

Experienced Data Engineer with a Master's degree in Data Analytics and 4 years of industry experience. Proficient in Python, cloud services, and data engineering tools. Skilled in developing scalable data pipelines, implementing machine learning models, and optimizing database performance. Committed to delivering high-quality solutions that drive business value through data-driven insights.

Skills

Languages: Python (Expert), SQL, JavaScript, HTML
Services: AWS (EC2, S3, RDS, Lambda, Redshift), Apache Spark, Kafka, HDFS, Redis, PostgreSQL, MongoDB, Cassandra, Django, Flask, scikit-learn, TensorFlow, Power BI, Tableau
DevOps: Docker, Kubernetes, Jenkins, Git

Professional Experience

Data Scientist | Veeco Instruments, San Jose, CA | Jul 2024 - Present
-Developed and implemented a data ingestion portal using Flask, JavaScript, and Python with a SQL database backend for a semiconductor wafer manufacturing process. The portal enforced data types, implemented value limits, and required all fields to be filled, eliminating missing data and reducing manual entry errors by 95%.
-Optimized ETL processes, resulting in a 40% reduction in daily processing time and a 25% decrease in cloud computing costs. Achievements include:

- Implemented parallel processing for data transformation jobs using Apache Spark, reducing processing time by 30%.
- Introduced data partitioning and columnar storage formats, improving query performance by 25%.
- Implemented incremental data loading, reducing the amount of data processed daily by 50%.

-Developed a machine learning model using TensorFlow to predict optimal cooling intervals for laser spike annealing tools, reducing thermal attrition by 30% and reducing material waste by 30%.

Member Technical Staff |HCL Technologies, Banglaore, India | Feb 2021 - Jun 2023
-Architected fault-tolerant data pipelines using Apache Airflow on AWS ECS, processing 500GB+ daily, with S3 for data lake storage and Glue for ETL jobs, reducing data reprocessing due to failures by 90%.
-Implemented serverless data processing using AWS Step Functions and Lambda, improving scalability.
-Led the migration of on-premise data warehouses to AWS Redshift, improving query performance by 3x and enabling seamless scaling for growing data volumes.
-Developed a Django-based real-time dashboard hosted on AWS Elastic Beanstalk, integrating with AWS Cognito for authentication, serving 200+ daily users.
-Optimized RDS PostgreSQL database queries, implemented AWS DynamoDB for high-throughput operations, reducing query time by 70%.

Trainee Engineer | Quest global Engineering, Bangalore, India| Nov 2019 - Oct 2020
-Engineered an automated data cleaning and preprocessing system using AWS Glue and S3, saving the team 10 hours per week and increasing data scientist productivity by 25%.
-Optimized Amazon Redshift queries for large-scale data retrieval, reducing average query execution time by 40% and improving overall system performance.
-Implemented CI/CD pipelines using AWS CodePipeline and CodeBuild, reducing deployment time by 70% and improving team collaboration efficiency.

Highest Education

Master of Science in Data Analytics | Northeastern University, Boston, MA | Graduation: May 2025 | GPA: 3.89/4.0
Relevant Coursework: Data mining,MLOps, Database Mangement, Data Visualization

---

Projects

News analysis pipeline: Stock market news API with an AWS-based data pipeline orchestrated by Airflow and containerized using Docker.
Most probable character prediction: A machine learning model for next-character prediction using TensorFlow, enhancing text generation capabilities.

Certifications

IBM Data science professional certification