

Kiran Gorijala

+1 443 858 5184 | kgorija1@umbc.edu | [LinkedIn](#) | [Portfolio](#)

Education

University of Maryland Baltimore County
Masters of professional studies in data Science

Baltimore, US
01/2023 – 12/2024

Technical Expertise

Programming and scripting languages: Python, C, R, C++, Java, Rust, Scala

Big Data Technologies: Hadoop, Apache Spark, HDFS, Hive, Databricks, Kafka, NiFi, HBase, Airflow, Snowflake

Database Technologies: Oracle, PostgreSQL, PL/SQL, DB2, MySQL, NoSQL Databases (MongoDB, Cassandra, DynamoDB)

Data Engineering: AWS (IAM, S3, EC2, RDS, Athena, Glue, Redshift, Lambda, CloudWatch, Kinesis, EMR) Azure (Data Lake, Cosmos DB, ADF, Azure Stream Analytics, HDInsight)

Utilities/Tools: SSIS, Informatica, Talend, Docker, Kubernetes, Jenkins, Terraform, Git, GitHub, SSRS, SSAS, Grafana

Machine Learning: Linear Regression, Logistic Regression Decision Tree, CNN, KNN, SVM, TensorFlow, PyTorch

Other Technical Skills: Data Quality and Governance, Data Modeling, Machine Learning, Data Mining, Data Warehousing, Data Visualization, ETL & CI/CD pipelines, Agile Development

Dev Tools / Applications: Django, Tableau, Linux, Looker, Power BI

Professional Experience

LARSEN & TOUBRO TECHNOLOGY SOLUTIONS,

Bengaluru, India

Data Engineer

09/2020 – 01/2023

- Built and developed scalable ETL workflows using **Azure Data Factory** and **Databricks** with **PySpark** to process more than 50 million flight records per day. Used **Linux operating systems** and **shell scripting** to automate and monitor pipeline execution, ensuring reliability in high-volume data processing.
- Improved **Spark SQL** queries and implemented partitioning/caching strategy in **Azure Databricks** for up to a 50% decrease in large-scale data transformation and flight allocation calculation times.
- Designed and implemented a full-cycle **data warehousing** using SQL and **Azure Synapse Analytics** for efficient storage and retrieval of large-scale flight data. Modernized query performance in flight allocation and scheduling reporting by 40%.
- Implemented **CI/CD pipelines** with **Jenkins** and **Git** to automate the deployment of real-time streaming workflows and advanced analytics, improving deployment efficiency by 40% and enhancing data quality frameworks while decreasing infrastructure costs by 30%.
- Innovated automation scripts within the **Azure platform** that reduced manual intervention time during daily updates by 6 hours per week while ensuring seamless integration with existing systems managing aircraft schedules.
- Achieved intuitive reports within **Power BI** focusing on key performance metrics such as passenger load factors and allocation efficiencies; directly improved decision-making processes for airlines managing over 500 daily flights.

ACCENTURE

Hyderabad, India

Associate Data Engineer

07/2019 – 08/2020

- Engineered an **Apache Spark**-based data processing workflow that is scalable for over 100 million customer records per day by applying the principles of **Agile** to enhance refresh timing and resource efficiency iteratively.
- Established and Automated a real-time data streaming pipeline using **Apache Kafka** and **Apache NiFi**. The improvement in data ingestion efficiency stood at 25% and orchestrated workflows using **Apache Airflow**, updating financial dashboards in near real-time.
- Orchestrated end-to-end transformations of diverse datasets through tailored ETL solutions via **SQL** and **AWS Glue**. This effort directly contributed to identifying three major factors driving fluctuations in customer borrowing patterns.
- Developed a **custom Python library** for data cleaning and validation, improving dataset consistency by 30%. Utilized TDD principles in developing robust ETL workflows in **Snowflake** and reduced query latency by 25%. Optimized ETL workflows using Snowflake to take advantage of scalable processing and storage to manage larger datasets for enhanced data integration.
- Produced tailored visualization tools using expertise with **SQL** queries feeding directly into the **Tableau** environment, achieving a significant automation milestone allowing self-service access covering over 200 different variables relevant to ongoing analysis initiatives.
- Collaborated with cross-functional teams to identify key data sources and extract relevant datasets using APIs and batch processing techniques to achieve comprehensive data coverage for analytics.
- Planned and coordinated data engineering tasks in a global delivery environment. Partnered with stakeholders to deliver robust data warehousing solutions, ensuring scalability and compliance with industry best practices.

Projects

Coronavirus Data Analytics: | Tech: python, Google Cloud Storage, Mage, Jupyter Notebook, Looker

12/2023 – 01/2023

- Enhanced COVID-19 predictive accuracy by 20% and pipeline efficiency by 30% with machine learning and Google Cloud Storage.
- Produced Looker dashboards to provide stakeholders with real-time access to critical coronavirus metrics.

Parking Spot Finder: | Tech: Flask, CNN, YOLOv8, OpenCV, Data Visualization, CVAT, Albumentations.

08/2023 – 12/2023

- Outlined parking detection by a hybrid CNN-YOLOv8 with 97.9% accuracy and a 20% reduction in search time.
- Sharpened performance using frame-skipping and state-change detection, reducing the load on processing by 40%.

Olympic Data Analysis: | Tech: Data Factory, Data Lake, Databricks, PySpark, Power BI, Tableau, ETL process

11/2023 – 12/2023

- Reduced processing time by 30% and Boosted data accuracy by 25% using Azure Data Factory, Databricks, and PySpark.
- Crafted engaging dashboards in Power BI and Tableau, increasing user engagement by 40% and supporting data-driven decisions.

Publications & Certificates

<http://www.ijcrt.org/papers/IJCRT2105661.pdf>

[Azure Data Engineer](#)

[Google Advanced Data Analytics](#)