

Kiran Gorijala

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

PROFESSIONAL SUMMARY

Experienced Data Engineer with a strong background in designing scalable and efficient data architectures. Skilled in developing secure, high-performance data platforms and enabling real-time analytics. Expertise in AWS, Azure, big data processing, machine learning, CI/CD, and Agile methodologies, with a focus on optimizing infrastructure and driving business insights

TECHNICAL SKILLS

Programming 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)

Cloud Technologies : 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 : Machine Learning, Data Warehousing, Data Visualization, ETL, CI/CD pipelines, Agile Development

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

EXPERIENCE

Data Engineer | *Experian* | California, United States

August 2023 – Present

- Built and developed scalable ETL workflows using AWS Glue, AWS Lambda, and AWS EMR with PySpark to process more than 50 million finance records per day. Used Linux operating systems and shell scripting to automate and monitor pipeline execution, ensuring reliability in high-volume data processing.
- Implemented Hadoop and HDFS for distributed storage and parallel processing, improving data retrieval by 35%.
- Utilized Amazon Kinesis and AWS Lambda for real-time data ingestion and processing.
- Enforced Agile methodologies, optimizing sprint planning, iterative development, and data pipeline efficiency.
- Containerized data processing applications using Docker, reducing environment setup time by 20%.
- Optimized Spark SQL queries and implemented partitioning/caching in AWS EMR, 50% reduction in transformation time.
- Designed and implemented a full-cycle data warehousing using SQL and Amazon Redshift for efficient storage and retrieval of large-scale flight data. Modernized query performance in flight allocation and 40% in scheduling reporting.
- Stored processed Financial data in PostgreSQL for efficient operational queries and reporting.
- Automated CI/CD deployments with Jenkins and Git, enhancing efficiency by 40% and 30% reduction of infrastructure.
- Harnessed Amazon CloudWatch and AWS X-Ray to build real-time monitoring dashboards on pipeline health with alerting to ensure 99.9% system availability.
- Developed Amazon QuickSight reports on key finance metrics, and improving decision-making.

Data Engineer | *Larsen & Toubro Technology Solutions* | Bengaluru, India

September 2020 – January 2023

- Engineered an Apache Spark-based data processing workflow that is scalable for over 10 million flight 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 Flight dashboards in near real-time.
- Orchestrated ETL transformations using SQL, BLOB Storage, Data Lake and Azure Data Factory, identifying key factors in customer borrowing patterns.
- Automated ETL workflows with Azure Functions, Azure Monitor, and Azure Key Vault for security and observability.
- Developed a custom Python library for data cleaning and validation, improving dataset consistency by 30%. Utilized TDD principles in developing robust ETL workflows in Azure Synapse Analytics and reduced query latency by 25%.
- Optimized ETL workflows in Azure Synapse Analytics for scalable data integration and enhanced processing.
- Established compliance with GDPR and other regulatory standards by implementing robust data governance frameworks, ensuring secure and auditable data management.
- 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 extract key datasets using APIs and batch processing.
- Developed and implemented robust Data Validation and Data Reconciliation processes within ETL workflows, ensuring data accuracy and integrity across all datasets.

Associate Data Engineer | *Accenture* | Hyderabad, India

July 2019 – August 2020

- Formulated ETL workflows using Informatica, SSIS, and Talend, reducing data processing time by 40%.
- Designed and optimized a real-time data streaming pipeline using Apache Kafka and Apache NiFi to ingest data from REST APIs and process high-velocity datasets.
- Deployed containerized ETL jobs with Kubernetes and Talend, enhancing scalability and fault tolerance for large-scale data pipelines.

- Amplified Analytics pipeline with the addition of Apache Kafka along with Snowflake provided the necessary high-performance, low-latency, real-time use cases for business tracking of customer behaviors/analyses of trends.
- Integrated data lineage and governance mechanisms to ensure transparency and accountability in data usage using tools like Apache Atlas and Snowflake Information Schema.
- Automated the ingestion and transformation of streaming data into Snowflake and AWS S3 for real-time availability to stakeholders, minimum latency, and quick decision-making regarding key business metrics.
- Designed and optimized NoSQL databases MongoDB, Cassandra for high-throughput applications, improving query performance by 30% and ensuring scalability for large datasets.
- Created Predictive Models in customer behavioral analysis with the strengthened data supplied through the AWS Glue pipeline, using these to underpin action and enhancements in Customer retention and mapped marketing strategy on emergent patterns.

PROJECTS

- | | |
|---|--------------------|
| Coronavirus Data Analytics <i>python, Google Cloud Storage, Mage, Jupyter Notebook, Looker</i> | <i>Nov 2023</i> |
| <ul style="list-style-type: none"> - Enhanced COVID-19 predictive accuracy by 20% and pipeline efficiency by 30% with machine learning and GCP. - Produced Looker dashboards to provide stakeholders with real-time access to critical coronavirus metrics. | |
| Parking Spot Finder <i>Flask, CNN, YOLOv8, OpenCV, Data Visualization, CVAT, Albumentations</i> | <i>Fall 2024</i> |
| <ul style="list-style-type: none"> - 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 <i>Data Factory, Data leak, DataBricks, Pyspark, powerBi, Tableau, ETL</i> | <i>Spring 2024</i> |
| <ul style="list-style-type: none"> - Reduced processing time by 30% and Boosted data accuracy by 25% using Azure Data Factory, Databricks, and PySpark. | |

EDUCATION

University of Maryland baltimore county	MPS in Data Science GPA : 3.86	<i>January 2023 – December 2024</i>
---	---	-------------------------------------

PUBLICATIONS AND CERTIFICATIONS

Open Data Visualizer <i>Paper Publication</i>	<i>September 2021</i>
Azure Certified Data Engineer <i>Certificate</i>	<i>August 2024</i>
google Advanced Data Analytics <i>Certificate</i>	<i>april 2024</i>