Kiran Gorijala
Santa Clara, California | *+1 949-431-2642* | *kirangorijala2014@gmail.com* | *@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 | ***Capital One*** | ***California, United States*** ***August 2023 – Present***

- Built and developed scalable ETL workflows using AWS Glue, AWS Lambda, and AWS EMR with PySpark to process millions of finance records per day, achieving 99.9% reliability in high-volume data processing through automation and proactive monitoring with Linux and shell scripting.
- Implemented Hadoop and HDFS for distributed storage and parallel processing, improving data retrieval speed by 35% and reducing query execution time from 15 seconds to 5 seconds.
- Utilized Amazon Kinesis and AWS Lambda for real-time data ingestion and processing, reducing data latency by 60%, enabling near-instantaneous insights for finance analytics.
- Enforced Agile methodologies, optimizing sprint planning and iterative development, resulting in a 25% increase in data pipeline efficiency and a 15% boost in team productivity.
- Containerized data processing applications using Docker, reducing environment setup time by 20% and improving deployment consistency across multiple environments.
- Optimized Spark SQL queries and implemented partitioning/caching in EMR, reducing data transformation time by 50%.
- Designed and implemented a full-cycle data warehouse using SQL and Amazon Redshift for large-scale Financial data storage and retrieval, enhancing query performance by 40% and reducing report generation time.
- Stored processed financial data in PostgreSQL, streamlining operational queries and reducing financial analysis processing time by 45%, from 12 minutes to 6.5 minutes.
- Automated CI/CD deployments with Jenkins and Git, improving development efficiency by 40% and reducing infrastructure overhead by 30%, cutting deployment time.
- Harnessed Amazon CloudWatch and AWS X-Ray to build real-time monitoring dashboards, ensuring 99.9% system availability and reducing incident resolution time from 2 hours to 30 minutes.
- Developed Amazon QuickSight reports on key finance metrics, accelerating executive decision-making by delivering real-time insights 20% faster.

Data Engineer | ***Airbus*** | ***Bengaluru, India*** ***September 2020 – January 2023***

- Engineered an Apache Spark pipeline processing millions of flight records daily, reducing refresh time by 30% through iterative Agile improvements and optimized resource allocation.
- Automated real-time data streaming with Apache Kafka and NiFi, boosting data ingestion efficiency by 25% and enabling near real-time dashboard updates for operational insights.
- Orchestrated ETL workflows using SQL, Azure Data Lake, and Data Factory, uncovering key borrowing patterns and improving predictive analytics for financial decision-making.
- Enhanced ETL reliability by automating workflows with Azure Functions and Key Vault, reducing manual intervention by 40% while strengthening security and monitoring.
- Developed a Python-based data validation library, improving dataset consistency by 30% and reducing query latency by 25% in Azure Synapse Analytics for large-scale data processing.
- Optimized ETL in Azure Synapse for scalable integration and faster processing, ensuring efficient data workflows.
- Implemented GDPR-compliant data governance frameworks for secure, auditable data management.
- Created SQL-driven Tableau dashboards with self-service access to 200+ variables, enabling dynamic analysis.
- Extracted key datasets via APIs and batch processing, streamlining data-driven decision-making.
- Developed robust ETL validation and reconciliation, ensuring data accuracy 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.
- Enhanced the analytics pipeline by integrating Apache Kafka with Snowflake, enabling high-performance, low-latency, real-time business tracking of customer behaviors and trend analysis
- 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 | *python, Google Cloud Storage, Mage, Jupyter Notebook,Looker* *Nov 2023*
- 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 | *Flask, CNN, Yolov8, OpenCV, Data Visualization, CVAT, Albumentations* *Fall 2024*
- 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 | *Data Factory, Data leak, DataBricks, Pyspark, powerBi, Tableau, ETL* *Spring 2024*
- 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 *January 2023 – December 2024*

Publications and Certifications

Open Data Visualizer | *Paper Publication* *September 2021*
Azure Certified Data Engineer | *Certificate* *August 2024*
google Advanced Data Analytics | *Certificate* *april 2024*