Saikiran Reddy

Texas | 7628475063 | Saikiranreddy1044@gmail.com | LinkedIn

SUMMARY

- Accomplished **Data Engineer & Data Analyst** with over **5 years** of expertise in implementing **Big Data technologies** such as **Hadoop**, **Map Reduce**, **Spark** and **Google Cloud Platform (GCP)** services including BigQuery, Dataflow, and Cloud Storage, resulting in a 25% boost in data processing efficiency. Proficient in creating **Power BI** dashboards for enhanced **data visualization** and delivering critical insights to senior management.
- Experienced in engineering **ETL pipelines** on Azure Databricks and AWS, optimizing data processing within Azure Data Factory. Leveraged Azure Synapse Analytics, **Hive**, and AWS services to streamline operational data flows, while also authoring **PowerShell** scripts for efficient cloud resource management.
- Demonstrated success in developing modern **ETL pipelines** in **Azure Data Factory**, integrating real-time data ingestion with Azure Event Hubs and Stream Analytics. Reduced Jira tickets by 18% through automated email alerts and provided valuable recommendations for improving job efficiency and **SQL** query performance.
- Designed and optimized complex **ETL** workflows in **Informatica PowerCenter,** enhancing data integration and processing efficiency across multiple data sources, leading to a 30% reduction in data processing time.

WORK EXPERIENCE

Molina Healthcare   **TX, USA**
Data Engineer   *Aug 2024 - Present*
- Implemented Big Data technologies such as **Hadoop**, **Map Reduce**, **HBa**se, and Hive, processing over 2 million data records monthly, resulting in streamlined data ingestion and enhanced efficiency.
- Transformed over 50 complex **Hive/SQL** queries into Spark transformations using **Spark RDD**, **Scala**, and **Pytho**n, boosting data processing efficiency by 25%.
- Built and maintained **BigQuery datasets and tables** for healthcare data, enabling fast analytics and reporting across multiple departments.
- Created and maintained Power BI dashboards for sales reporting, enhancing data visualization and providing critical business insights to senior management.
- Engineered **ETL** pipelines on **Azure Data bricks**, facilitating efficient **data ingestion** and processing from diverse sources, and managed environments to optimize analytics.
- Utilized Azure Synapse Analytics and Hive to manage and streamline data processing within Azure Data Factory, significantly improving operational data flows.
- Developed **Dataflow pipelines** for processing and transforming large-scale healthcare data, improving ETL efficiency by 20–25%.
- Leveraged Cassandra as the backend for a customer-facing healthcare application, managing 6 TB of user profile data across 5 nodes and handling 10,000 queries per minute to maintain seamless performance.
- Authored PowerShell scripts to monitor and manage **Azure cloud resources**, optimizing infrastructure for substantial data volumes and improving resource efficiency.

Dell Technologies   **India**
Data Engineer   *Jul 2020 – Jun 2022*
- Developed and maintained scalable **ETL** pipelines in **Azure Data Factory** and **Databricks** to process large volumes of operational and enterprise data across global systems.
- Built streaming data solutions using **Azure** Event Hubs and Stream Analytics for IoT telemetry from manufacturing sites, enabling near real-time insights into production performance.
- Implemented Medallion Architecture **(Bronze, Silver, and Gold)** in **Databricks** with **Delta Lake** and **Delta Live Tables**, improving data quality, consistency, and accessibility for analytics teams.
- Designed robust ingestion frameworks to handle data from **SAP, Oracle, SQL Server, Parquet,** and **REST APIs**, ensuring smooth integration into Teradata Vantage.
- Optimized **ETL** workflows and **SQL** queries, reducing data pipeline execution time from over 3 hours to under 60 minutes with zero job failures.
- Integrated Logic Apps with **ADF** to automate job monitoring and email notifications, cutting manual tracking effort and reducing **Jira** incidents by 18%.
- Collaborated with data modelers and business stakeholders to design efficient data models for manufacturing operations, financial analytics, and customer delivery tracking.
- Developed automation scripts in **Python, PowerShell,** and **VB** to clean, transform, and standardize raw datasets, reducing manual effort by 85%.
- Delivered **Power BI** and **Tableau** dashboards connected to Azure Synapse and Databricks datasets, improving visibility into production metrics and business **KPIs**.
- Enhanced **Informatica PowerCenter** and **Informatica Cloud (IICS)** workflows to enable seamless **data migration** and synchronization across hybrid cloud environments.
- Implemented **data governance** best practices with row-level security and access control, ensuring compliance and protecting

sensitive business information.
- Migrated legacy SQL and Spark workloads to **BigQuery**, reducing query execution time and cloud compute costs.
- Implemented a **Snowflake**-based data warehouse architecture integrated with **Azure Synapse** and **Cosmos DB**, enhancing cross-platform queries, enabling scalable data access, and reducing compute costs for enterprise-wide predictive and operational reporting.
- Implemented **real-time streaming pipelines** on GCP using **Pub/Sub** and **Dataflow**, enabling faster IoT data processing from manufacturing sites.

KPMG   **India**
Data Analyst   *May 2018 – May 2020*
- Performed exploratory data analysis using **Python (NumPy, Pandas, and Matplotlib)** to uncover insights and anomalies in large financial and operational datasets supporting audit and advisory projects.
- Wrote optimized **SQL** queries in **SQL Server** and **PostgreSQL** to retrieve, clean, and transform multi-source client data, improving reporting accuracy and reducing turnaround time.
- Developed interactive **Tableau dashboards** to visualize audit metrics, risk indicators, and financial **KPIs** for business stakeholders and client reporting.
- Applied data wrangling and preprocessing techniques in **Python** and **SQL** to handle missing values, outliers, and inconsistencies in transactional and compliance data.
- Leveraged **AWS** services (S3, Glue, Redshift, Athena, and EMR) to design secure, **scalable data pipelines** for financial analytics and regulatory reporting.
- Implemented AWS Lambda functions to automate **ETL workflows** and **data validation**, reducing manual intervention and improving data reliability.
- Conducted financial impact assessments and forecasting using advanced **Excel tools (PivotTables, Power Query, INDEX-MATCH, and VLOOKUP)** to support budgeting and strategic decisions.
- Enhanced **SQL** and **Python** workflows to improve data aggregation and query performance, cutting report generation time by over 40%.
- Designed and implemented **ETL pipelines** on GCP using **Cloud Storage, Glue, and BigQuery** to consolidate financial and operational datasets.
- Supported due diligence and **risk analysis** initiatives by integrating and analyzing large datasets from banking, insurance, and consumer industries.

CERTIFICATIONS
- AWS Certification (AWS Certified Cloud Practitioner)
- Talend Data Integration Certified Developer

EDUCATION

Texas A&M University - Commerce   *Aug 2022 – May 2024*
Master's, Computer Science

Osmania University   *Jul 2016 – May 2019*
Bachelors in Mathematics, Statistics & Computer Science

SKILLS & INTERESTS

Programming Languages: Python, SQL, Scala, PowerShell, VBA, JavaScript
Data Engineering & ETL Tools: Apache Airflow, Spark, PySpark, Talend, Informatica (PowerCenter & IICS), AWS Lambda, AWS Glue, AWS EMR, Azure Data Factory, Azure Stream Analytics, IBM DataStage
Cloud Platforms & Databases: Microsoft Azure (Data Lake, Databricks, Synapse, Blob Storage), Amazon Web Services (S3, RDS, Lambda, EventBridge, Athena, Redshift), Google BigQuery, Oracle, PostgreSQL, MySQL, Snowflake, Teradata, MongoDB, Cassandra, MS SQL Server
Big Data & Streaming Technologies: Apache Hadoop, HDFS, YARN, Hive, Pig, Kafka, MapReduce, Azure Databricks
Data Analytics & Visualization: Power BI, Tableau, QlikView, Excel (Power Query, PivotTables), Seaborn, Matplotlib, ArcGIS
Machine Learning & Data Science: Azure Machine Learning, Forecasting Models, Statistical Analysis, SciPy, OpenCV
Business Intelligence & Reporting: SSIS, SSRS, SSMS, Tableau, Power BI, Data Reporting Automation
Software Development & Tools: Git, GitLab CI/CD, Jenkins, JIRA, Visual Studio, PyCharm, Jupyter Notebook, RESTful APIs, Selenium, SharePoint
Methodologies & Compliance: Agile (Scrum), Waterfall, HIPAA Compliance, Collaboration with Cross-Functional Teams
Operating Systems: Linux/Unix, Windows