

K SRIHARSHA

Aubrey, TX | +1 (913)-201-6393 | harshas4411@gmail.com | [Portfolio](#)

PROFESSIONAL SUMMARY

Data Engineer with 3 years of experience designing and optimizing scalable ETL pipelines, data lakes, and cloud architectures across AWS, Azure, and hybrid environments. Expert in Spark, PySpark, Scala, Kafka, Airflow, Databricks, SQL, and Python for real-time and batch data processing, warehouse migrations (Synapse, Redshift, Snowflake), and BI integration with Power BI and Tableau. Skilled in data modeling, governance, quality, and high-throughput, secure data solutions, enabling actionable insights for financial, operational, and enterprise analytics.

TECHNICAL SKILLS

Programming & Scripting: Python, SQL, T-SQL, PySpark, Bash, Shell Scripting

Big Data & ETL Tools: Apache Spark, Hadoop, Hive, Pig, Sqoop, Flume, Oozie, Kafka, Airflow, Delta Lake, dbt, Informatica, SSIS, Talend

Cloud & Data Platforms: AWS (S3, EMR, Redshift, Glue, Lambda, RDS, DynamoDB, Athena, CloudWatch, SNS, SQS, Kinesis), Azure (Data Factory, Synapse Analytics, Databricks, Data Lake Gen2, Azure SQL, Azure Functions, Logic Apps, Key Vault), Google Cloud Platform (BigQuery, Dataflow, Pub/Sub, Cloud Storage)

Databases & Data Warehousing: MySQL, PostgreSQL, MongoDB, HBase, SQL Server, Snowflake, Teradata, Oracle, Cosmos DB

Data Modeling & Analytics: Star/Snowflake Schema, OLAP/OLTP, Data Lakes, Data Warehousing, Incremental & Batch ETL, Data Governance, Data Quality, Metadata Management

Business Intelligence & Visualization: Power BI, Tableau, Looker, Qlik Sense, Matplotlib, Seaborn, Plotly, Pandas Profiling

Machine Learning & Data Science: Pandas, NumPy, SciPy, Scikit-learn, TensorFlow, PyTorch

DevOps & Orchestration: Docker, Kubernetes, Jenkins, Git, GitHub Actions, CI/CD Pipelines, Terraform, Helm

Operating Systems & Tools: Linux, Windows, MacOS, VS Code, PyCharm, Jupyter Notebook

Workflow & Scheduling: Apache Airflow, Oozie

PROFESSIONAL EXPERIENCE

Data Engineer

Aug 2023 - Present

Credit Suisse | New York, USA

- Engineered cloud-based data pipelines using Azure Data Factory and Databricks, successfully migrating on-premise warehouses to Synapse and reducing query latency by 35%.
- Streamlined ETL processes to load API and transactional data into Azure SQL Database, enabling analytics teams to access refreshed datasets 40% faster.
- Enhanced Spark workloads in PySpark and Scala for batch and real-time processing, increasing throughput by 50% while minimizing cluster resource consumption.
- Implemented monitoring dashboards and automated triggers via Airflow and Log Analytics, cutting manual pipeline interventions and reducing downtime by 30%.
- Built Kafka streaming applications to ingest high-volume real-time events, achieving consistent sub-2-second latency and supporting timely business insights.
- Developed and optimized T-SQL queries, stored procedures, and JSON scripts, ensuring accurate Power BI reporting for finance and operational teams.
- Strengthened enterprise data security and governance by deploying Azure Key Vault for secrets and collaborating on physical and logical data models, mitigating compliance risks.
- Performed comprehensive data validation, profiling, and anomaly detection using PySpark and Python libraries, increasing downstream dataset accuracy to 99%.

Data Analyst

Oct 2020 - Nov 2021

Hewlett Packard | India

- Transformed raw operational data into interactive Power BI dashboards, enabling management to track KPIs and trends, which accelerated decisionmaking by 30%.
- Streamlined SQL queries and subqueries on Microsoft SQL Server, reducing report generation time by 25% and increasing overall analytics efficiency.
- Validated and mapped data across multiple source systems to warehouses, ensuring 99% accuracy in BI reports and reliable insights for business users.
- Leveraged Hadoop and Spark to process large datasets, shortening batch analytics time by 40% and supporting timely report delivery.
- Automated ETL and reporting processes using Python and SQL, improving data pipeline reliability and cutting manual effort by 30%.
- Collaborated with business and IT teams under Agile SDLC, improving BI development speed by 20% while maintaining requirement fidelity.
- Conducted data quality and anomaly detection checks, ensuring dashboards delivered actionable insights and governance-compliant datasets.
- Generated ad-hoc and scheduled reports to highlight operational and financial trends, enabling teams to quickly identify gaps and opportunities.

PROJECTS

Real-Time Loan Analytics Platform

- Engineered ETL pipelines with Azure Data Factory, Databricks, and Synapse, migrating on-premise data warehouses and reducing query latency by 35%.
- Built Kafka streaming applications and Spark (PySpark & Scala) jobs to process high-volume real-time loan events, enabling sub-2-second data availability for analytics.
- Developed T-SQL queries, stored procedures, and JSON scripts for Power BI dashboards, improving finance and operations decision-making speed by 40%.

Cloud-Based Data Warehouse Migration & Optimization

- Implemented AWS Glue, Lambda, S3, and EMR pipelines to process large-scale web server logs, accelerating ETL workflows by 35%.
- Converted Hive and SQL queries into SparkSQL and PySpark jobs, enhancing transformed dataset accessibility for analytics teams by 40%.
- Optimized HDFS and relational databases with partitioning, indexing, and compression, lowering query runtimes and storage costs by 25%.

Enterprise Data Lake & Streaming Analytics Solution

- Built Spark (Scala/PySpark) and Hive pipelines for structured and semi-structured data, increasing ETL throughput by 35% and enabling timely analytics delivery.
- Orchestrated batch and streaming ETL workflows using Apache Airflow DAGs, reducing scheduling errors by 30% and improving pipeline reliability.
- Developed Kafka producers and Spark streaming applications for real-time data ingestion, supporting low-latency insights across business operations.

EDUCATION

Master of Computer Science

University of Central Missouri | Warrensburg, MO

CERTIFICATIONS

- IBM AI Practitioner - **Coursera**
- Google Data Analytics Professional Certificate - **Coursera**
- Microsoft Azure Data Engineer Associate - **LinkedIn Learning**
- Data Engineering with Google Cloud Professional Certificate - **Coursera**