SATHISH KUMAR

+1 (314) 325-9624 | dvskr.333@gmail.com | www.linkedin.com/in/dvskr | https://github.com/dvskr

PROFESSIONAL SUMMARY

Data Engineer experienced in designing scalable data pipelines in Azure and hybrid cloud environments, with expertise in tools like Azure Data Factory, Snowflake, and Apache Spark. Focused on real-time data streaming and batch processing, with accomplishments in optimizing OLAP/OLTP systems and implementing secure data workflows compliant with industry standards. Aims to leverage technical skills and cross-functional collaboration to enhance data solutions and support business decisions.

TECHNICAL SKILLS

- **Languages**: Python (Pandas, NumPy, PySpark), SQL, Bash
- **Big Data**: Apache Spark, Hadoop, Hive, Kafka, Delta Lake, BigQuery
- **Warehousing**: Snowflake, Redshift, PostgreSQL, SQL Server, MongoDB
- **ETL Tools**: Airflow, dbt, AWS Glue, Azure Data Factory, Apache NiFi, Databricks
- **Cloud Platforms**: AWS (S3, Glue, Redshift, Lambda), Azure (ADF, Synapse), GCP (BigQuery, Dataflow)
- **DevOps**: Terraform, Jenkins, GitHub Actions, Docker, Kubernetes
- **Modeling**: Star/Snowflake Schema, OLAP/OLTP, Partitioning, Indexing
- **Security**: RBAC, Data Encryption, CloudWatch, Audit Logging, HIPAA, GDPR
- **BI & Reporting**: Tableau, Power BI, Looker
- **Soft Skills**: Agile, Problem-Solving, Communication, Debugging, Team Collaboration

PROFESSIONAL EXPERIENCE

Propper International **May 2023 - Present**
Big Data Engineer *Saint Louis, Mo, USA.*

- Developed and orchestrated over 100 data pipelines using Apache Airflow and Azure Data Factory to ingest, transform, and load multi-terabyte datasets from manufacturing systems, supplier integrations, retail POS, and eCommerce channels into centralized data lakes and warehouses.
- Enabled real-time visibility into production, inventory, and fulfillment operations using Apache Kafka and Spark Structured Streaming, processing events from factory floors, logistics platforms, and ERP systems.
- Migrated legacy ETL workflows to Snowflake using dbt, implementing flexible data models to support use cases such as production tracking, inventory optimization, and order management.
- Deployed Azure Functions for automation of key manufacturing events like order status updates, quality inspection flags, return logistics, and exception handling across distributed systems.
- Built scalable, secure RESTful APIs to expose near real-time inventory, throughput, and sales data to analytics platforms and decision support systems for manufacturing leadership.
- Partnered with production, supply chain, and operations teams to define KPIs and build Power BI dashboards for monitoring plant utilization, order cycle times, downtime incidents, and supplier performance.
- Integrated enterprise platforms such as ERP and CRM systems using standardized APIs, ensuring consistent, synchronized data across product catalogs, work orders, and customer histories.
- Implemented CI/CD pipelines using Terraform and GitHub Actions for infrastructure provisioning, release management, and continuous delivery of data services and transformations.
- Created a reusable data quality framework to monitor critical KPIs such as inventory variance, order fulfillment accuracy, and supply chain disruptions, supporting proactive quality control.
- Tuned performance in Azure Synapse Analytics and Snowflake through strategic use of partitioning, materialized views, and clustering, ensuring efficient querying of large-scale manufacturing datasets.
- Optimized ADF job scheduling, data lake storage formats, and parallelization strategies to reduce processing time and improve overall data pipeline reliability and scalability.

Globus Medical Inc **Jan 2021 - Dec 2022**
Data Engineer *Hyderabad, IND.*

- Built and maintained over 50 batch and real-time pipelines using Airflow, Spark, and Apache NiFi to process clinical, claims, and financial data.
- Developed secure, token-based APIs with AWS Lambda to expose PHI-compliant datasets for downstream analytics, ensuring HIPAA compliance.
- Ingested over 1TB of healthcare data daily via Spark, supporting both real-time and scheduled processing for clinical and financial workflows.
- Created executive dashboards in Power BI and Tableau to visualize operational metrics, cost trends, and patient outcomes.
- Applied RBAC, audit logging, and data encryption across AWS S3 and Glue to maintain strict compliance with HIPAA and GDPR.
- Deployed predictive models for patient risk and monitoring workflows using Spark and Lambda for low latency scoring and alerting.

- Integrated third-party APIs and external health data streams using NiFi, standardizing input for centralized analytics.
- Implemented CI/CD pipelines with Jenkins, Docker, and GitHub Actions to support continuous integration and production-grade releases.
- Designed audit trails and data lineage frameworks for traceability and regulatory audits.
- Led Agile ceremonies, mentored new engineers, and contributed to internal documentation and reusable component libraries.

PROJECTS

Real-Time Patient Monitoring & Alerting System

•Ingested live data from wearable devices via Apache Kafka into Spark Streaming for near real-time processing.
•Implemented threshold-based and ML-driven anomaly alerts using PySpark, deployed to AWS Lambda.
•Built a resilient pipeline in Airflow to aggregate, transform, and store alert data in Snowflake for historical analysis.
•Developed interactive dashboards in Power BI for ICU teams and emergency response staff.
•Integrated with EMR systems via FHIR API and ensured full compliance with HIPAA.

Azure Databricks & Formula 1 Racing Data

- Designed and implemented a modular, scalable Lakehouse Architecture using Azure Databricks, Delta Lake, and Azure Data Factory to process Formula 1 racing data end-to-end.
- Engineered robust PySpark transformation logic leveraging Spark Core APIs for efficient ingestion, cleansing, and enrichment of historical and real-time race data.
- Utilized Delta Lake features include schema enforcement, ACID transactions, and time travel for consistent and performant data processing.
- Built and orchestrated complex data pipelines with Azure Data Factory, handling incremental loads, dependency chaining, and error handling mechanisms.

EDUCATION

Southeast Missouri State University – Cape Girardeau, MO **Present**
M.S., Computer Science

Karunya Institute of Technology and Sciences – Coimbatore, India **Present**
B.Tech, Computer Science

CERTIFICATIONS

•**Neural Networks and Deep Learning:**DeepLearning.Al
•**Python for Everybody & Python Data Structures:**University of Michigan