

VAMSHI SHIRAGANI

DATA ENGINEER

vamshi.s@mycvtalent.com | 660-541-2919 | Riverview, Florida

SUMMARY

- Data Engineer with 5+ years of experience across the data pipeline, from acquiring and validating large datasets (structured and unstructured) to building data models, developing reports, and utilizing visualization tools for impactful insights.
- Expert in developing enterprise solutions utilizing batch processing with Data Bricks and streaming frameworks, including Spark Streaming, Apache Kafka, Apache Airflow and Apache Flink, ensuring high throughput and low latency in data processing.
- Proficient in SQL, database design, and ETL pipeline implementation. Proven ability to leverage a wide range of cloud services, including Azure (Azure DevOps, Azure Data Lake, Azure Data Factory, Azure Databricks) and AWS (EC2, S3, RDS, Lambda, Glue, Athena, AWS Pipeline, Redshift) to design, build, and manage data solutions.
- Experience in developing and deploying continuous integration/continuous delivery (CI/CD) pipelines for Snowflake objects, ensuring streamlined deployments and minimizing the risk of errors.

SKILLS

Programming Language:	Python, R, SQL
IDE's:	PyCharm, Jupyter Notebook
Big Data Ecosystem:	Hadoop, Hive, Apache Airflow, Apache Kafka, Apache Spark, Apache Flink, DataBricks
Cloud Technologies:	Azure (Azure DevOps, Azure Data Lake, Azure Data Factory, Azure Databricks),AWS (EC2, S3, RDS, Lambda, Glue, Athena, AWS Pipeline, Redshift),
Visualizations:	Tableau, Power BI, Excel
Packages & Data Processing:	NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, PySpark, Data Pipelines, Jenkins
Version Control & Database:	GitHub, Git, SQL Server, PostgreSQL, MongoDB, DynamoDB, MySQL, Snowflake
Operating Systems:	Windows, MacOS

EDUCATION

Masters in Computer and Information Science

Apr 2024

Northwest Missouri State University, MO

EXPERIENCE

CVS Health, MO | Azure Data Engineer

Jan 2024 – Present

- Optimized Spark jobs by leveraging in-memory processing and data partitioning techniques to achieve faster processing times.
- Applied robust Kafka-based streaming pipeline to ensure reliable and timely delivery of real-time data to diverse applications, enabling faster insights and decision-making.
- Employed Azure Data Lake Storage, leading to a 30% improvement in data storage and retrieval efficiency for large-scale datasets.
- Created a Power BI dashboard to visualize key business KPIs, resulting in a weekly time savings of 2 hours on manual reporting.
- Implemented and deployed Spark applications on Databricks clusters for data processing, achieving a 10% improvement in data processing speed.

Dixon Technology, India | Azure Data Engineer

Aug 2020 – Nov 2022

- Designed and managed a large-scale data warehouse on Hive, storing and querying petabytes of historical sales data to support business intelligence and ad-hoc analysis.
- Developed and executed notebooks in Azure Databricks for interactive data exploration, visualization, and collaborative data analysis with data scientists.
- Employed Snowflake materialized views, data masking, and optimization to deliver efficient and secure data solutions for seamless data management and advanced big data analytics.
- Crafted and modified complex SQL queries, leveraging indexing and query optimization techniques to enhance database efficiency, leading to a 30% improvement in application response time.
- Achieved a 10% improvement in data processing efficiency by optimizing data flows within Azure Data Factory.

Oval SoftTech, India | Data Engineer

Jan 2018 – Jul 2020

- Implement a robust orchestration framework using Apache Airflow to automate daily data ingestion pipelines from various data sources (relational databases, CSV files) into Hive.
- Leveraged Databricks as a cloud-based platform for deploying and managing Apache Spark workloads, enabling efficient resource allocation and scalability for big data processing.
- Calculated AWS Data Pipeline workflows for efficient ETL Process, which reduced data processing time by 20%.
- Conducted comprehensive architecture and implementation assessments for AWS services, including Amazon EMR, Redshift, and S3, ensuring optimal cloud solutions.
- Monitored and maintained the entire data pipeline infrastructure (Kafka, Snowflake, MongoDB) to ensure high availability and real-time data processing for fraud detection.

CERTIFICATION

- **Certified Microsoft Azure Data Engineer Associate (DP-203)**
- **Certified Microsoft Azure fundamentals (AZ-900)**