

# Saiprem V

Data Engineer Specializing in Azure and AWS

📞 334-843-1429 saipremvm@gmail.com 🏠 Texas, 75071  [saiprem](#)

## Professional Summary

---

- Expert Data Engineer with around 4 years of hands-on experience in developing, deploying, and optimizing big data solutions using Microsoft Azure and AWS, aligning with modern cloud and data engineering practices.
- Proficient in Apache Spark, Hadoop, and Hive, focusing on building scalable data processing and analytics pipelines, key requirements in contemporary data engineering roles.
- Demonstrated expertise in cloud migrations, including data lake constructions and transformations within Azure and AWS ecosystems, ensuring data integrity and compliance.
- Strong foundation in ETL development and optimization, frequently utilizing Azure Data Factory and AWS Glue, aligning with industry standards for data movement and transformation.
- Skilled in developing data warehousing solutions on Azure Synapse and AWS Redshift, catering to complex analytical queries and large-scale data storage needs.
- Proven capability in employing advanced analytical techniques, including real-time data streaming and machine learning model implementation to drive strategic business decisions.
- Experienced in enhancing data accessibility and system performance through comprehensive cloud service integrations and database optimizations.
- Adept at implementing robust data security measures and compliance protocols, ensuring data protection and adherence to governance standards.
- Engaged in continuous learning and application of the latest data engineering technologies and methodologies, maintaining relevance in a fast-evolving field.
- Effective collaborator and communicator, often bridging the gap between technical teams and business stakeholders to ensure alignment of data initiatives with business goals.

## Technical Skills

---

- Cloud Platforms: Azure (Databricks, Data Factory, SQL Database, Cosmos DB), AWS (Redshift, S3, Lambda, RDS)
- Big Data Technologies: Spark, Apache Hive, Hadoop
- Containers and Orchestration: Docker, Kubernetes, AWS ECS, AWS EKS
- Programming Languages: Python, R, Scala, Bash
- Data Modeling and Database Management: SQL, NoSQL, PostgreSQL, DynamoDB, Snowflake
- Data Analysis and Machine Learning: Pandas, NumPy, ML Flow, Spark MLlib
- Reporting and Visualization Tools: Power BI, Tableau
- Web Development Frameworks: Django, Flask
- Version Control: Git

## Education

---

[George Mason University](#) - MSc Data Analytics Engineering, Fairfax, VA, USA Grade: 3.73/4

## Professional Experience

---

[East West Bank](#) - Azure Data Engineer, USA [Nov 2022 - Present]

- Worked on development and deployment of complex ETL pipelines for high-volume data transfer and transformation, achieving significant improvements in data processing efficiency and accuracy.

- Spearheaded the integration of Azure Data Factory with SQL Data Warehouses, automating data flows and significantly reducing manual data handling errors.
- Designed custom data extraction templates and transformation modules in Azure Databricks, facilitating seamless data integration from multiple data sources.
- Engineered real-time data ingestion systems using Apache Kafka and Azure Stream Analytics, optimizing data throughput and latency for time-sensitive analytics.
- Implemented advanced error handling and data validation mechanisms within ETL workflows, enhancing data quality and reliability.
- Developed performance tuning strategies for ETL processes, significantly reducing load times and improving overall system responsiveness.
- Automated the deployment of ETL jobs using Azure DevOps pipelines, ensuring continuous integration and delivery across multiple environments.
- Conducted detailed data analysis and profiling to identify bottlenecks in data pipelines, proposing and implementing optimizations that improved processing speeds by up to 40
- Collaborated with data scientists to embed machine learning models into ETL pipelines, enabling predictive analytics directly within data flows.
- Documented ETL design and operational procedures, providing clear guidelines and training materials for system users and fellow engineers.

#### [RSA India](#) - AWS Data Engineer, India [May 2019 - Aug 2021]

- Developed scalable ETL pipelines in AWS Glue to handle complex data transformations and load processes, supporting business intelligence and reporting needs.
- Utilized AWS Lambda and Step Functions to orchestrate and automate ETL tasks, reducing manual oversight and increasing process reliability.
- Integrated AWS Kinesis for real-time data streaming into ETL pipelines, enabling immediate data availability for analytics platforms.
- Optimized data extraction techniques using advanced SQL queries and Python scripting in AWS environments, ensuring efficient data manipulation.
- Implemented comprehensive logging and monitoring of ETL processes using AWS CloudWatch, providing real-time insights into pipeline performance.
- Designed data redundancy and backup strategies within AWS S3, ensuring data durability and high availability.
- Leveraged AWS Redshift Spectrum for querying exabytes of data directly in S3 from Redshift, enhancing the scalability of data analysis tasks.
- Conducted regular reviews and updates of ETL codebase, applying best practices and new technologies to improve pipeline efficiency and maintainability.
- Led cross-departmental initiatives to streamline data collection and integration processes, significantly enhancing data-driven decision-making capabilities.
- Provided expert troubleshooting and support for ETL systems, rapidly resolving issues to maintain high levels of system availability and user satisfaction.