

# Karthik S

Email: karthikshivashankar654@gmail.com | Phone: +91 7019597321 | LinkedIn: [linkedin.com/in/karthik-s-b6346635a](https://www.linkedin.com/in/karthik-s-b6346635a)

---

## Professional Summary

AWS Data Engineer with 2.8 years of experience in designing, developing, and optimizing data pipelines using Python, PySpark, and AWS Data Analytics services. Skilled in implementing data migration, transformation, and integration workflows using AWS DMS, Glue, Lambda, Redshift, Athena, and RDS. Proficient in SQL (Oracle & Postgres) for querying, manipulating, and analyzing large datasets. Adept at leveraging AWS services like S3, IAM, CloudWatch, and VPC to ensure secure, scalable, and high-performance data solutions. Strong collaborator experienced in working with DevOps teams using Azure DevOps for CI/CD pipelines and infrastructure management. Passionate about delivering reliable, efficient, and automation-driven data engineering solutions.

---

## Technical Skills

- **Cloud Platforms:** AWS (S3, Redshift, Glue, Lambda, RDS, DMS, IAM, CloudWatch, Athena, VPC)
- **Data Processing:** PySpark, Apache Spark, Airflow
- **Databases:** Oracle, Postgres, Redshift, SQL
- **Programming Languages:** Python, SQL
- **Other:** ETL/ELT Development, Data Modeling, Data Quality, CI/CD (Azure DevOps), Performance Optimization

## Experience

---

### Data Engineer | KPMG | Oct 2022 – Present

- Designed, developed, and optimized data pipelines using Python, PySpark, and AWS services (Glue, Lambda, Redshift, Athena, RDS).
- Implemented data migration and transformation workflows using AWS DMS and Glue.
- Created and maintained ETL/ELT processes for ingestion, transformation, and analytics.
- Utilized SQL (Oracle & Postgres) to query, manipulate, and analyze large datasets.
- Configured AWS services like S3, IAM, CloudWatch, and VPC for secure, efficient data operations.
- Developed automation scripts in Python for data processing and workflow orchestration.
- Collaborated with DevOps teams using Azure DevOps for CI/CD pipelines.
- Monitored and troubleshooted data workflows to ensure high availability and performance.

## Projects

---

### 1. Real-Time Sales Data Pipeline (AWS)

- Built an ETL pipeline to ingest sales data from APIs, flat files, and streaming sources into S3.
- Processed data with AWS Glue + PySpark and loaded into Redshift for analytics and dashboards.
- Improved processing speed by 40% through optimized transformations, partitioning, and compression.

### 2. Financial Transactions ETL (AWS)

- Automated ingestion of financial data into S3 from multiple structured and semi-structured sources using DMS and Glue.
- Developed PySpark transformations for cleansing, schema validation, and aggregations.
- Stored curated data in Redshift for compliance reporting and analytics.

### 3. Log Data Processing & Monitoring Pipeline (AWS)

- Designed a pipeline to process and monitor large volumes of server logs using S3, Glue, PySpark, and CloudWatch.
- Reduced raw log size by 55% through filtering and aggregation.
- Enabled real-time monitoring dashboards for operational insights.

## Education

---

**Bachelor of Engineering** Dayananda Sagar Academy of Technology and Management (DSATM), Bangalore  
Aug 2018 – Jun 2022