



[Solutions Architect](#)



[Data Engineer](#)

Sathvika Kolisetty

Data Engineer | Cloud Data Engineer | Cloud Solutions Architect

Email: sathvika.kolisetty09@gmail.com

Phone: +1 9374009988

LinkedIn: linkedin.com/sathvikakolisetty

Portfolio: portfolio.com/sathvikakolisetty

Professional Summary:

- Highly skilled Data Engineer **with 4+ years of experience** in designing, developing, and deploying **scalable AWS and GCP cloud-based data solutions** to drive business intelligence, process automation, and cost optimization.
- Extensive expertise in AWS services, including **S3, Glue, Lambda, Redshift, Athena, DynamoDB, RDS, EMR, Step Functions, QuickSight, Kinesis, AWS DMS, CloudWatch, Lake Formation, IAM**, ensuring efficient data ingestion, transformation, and storage.
- Hands-on experience with GCP services, including **BigQuery, Dataflow, Cloud SQL, Cloud Run, Cloud Composer, Vertex AI, and Pub/Sub**, enabling scalable and efficient data processing.
- Expertise in Big Data technologies, including **Apache Spark (Core, SQL, Streaming, MLlib), Kafka, Flink, Snowflake, dbt, Hive and Pig** for scalable data solutions.
- Designed and built data lakes using **AWS S3, Glue, and Lake Formation**, enabling structured and unstructured data storage, governance, and retrieval, eliminating manual interventions and **reducing data retrieval time by 35%**.
- Developed and optimized highly scalable ETL/ELT pipelines using **AWS Glue, PySpark, Redshift, dbt, and GCP Dataflow**, improving data integration, transformation, and analytics efficiency.
- Built and optimized real-time streaming data pipelines using **Apache Kafka, AWS Kinesis, Flink, and GCP Dataflow**, **reducing processing latency by 60%** and enabling event-driven architectures.
- Migrated large-scale datasets from on-premise databases to **AWS Redshift, Snowflake, and BigQuery**, ensuring seamless cloud transition with minimal downtime, while enabling real-time reporting that improved **data accessibility by 45%**.
- Optimized petabyte-scale **Redshift & BigQuery workloads**, cutting **query execution time by 40%** and **reducing compute costs by 25%** through fine-tuned SQL optimizations, indexing strategies, partition pruning, and materialized views.

- Developed and deployed **FastAPI-based microservices** with **AWS API Gateway**, ensuring **low-latency API responses** for real-time data processing. Integrated CI/CD pipelines with **GitLab, Jenkins, and AWS CodePipeline** for seamless deployments.
- Automated cloud infrastructure provisioning using **Terraform and AWS CloudFormation**, improving **deployment speed by 70%** and ensuring consistency across environments.
- Implemented **security best practices** using **AWS IAM, KMS, Secrets Manager, GCP IAM**, ensuring data protection and compliance with **GDPR, HIPAA, and SOC 2 regulations**.
- Designed AI/ML-driven analytics solutions using **AWS SageMaker, Bedrock, and Vertex AI**, enabling predictive insights and intelligent automation that enhanced **business forecasting accuracy by 20%**.
- Built interactive dashboards using **AWS QuickSight and GCP Looker**, providing **real-time** business insights and reporting, leading to a **30% improvement** in data-driven decision-making.
- Developed containerized data processing applications using **AWS ECS, Fargate, and Kubernetes**, ensuring scalability and high availability while **reducing operational costs by 15%**.
- Orchestrated batch and streaming data workflows using **AWS Step Functions, Apache Airflow, and EventBridge**, automating end-to-end processes for optimized efficiency.
- Strong expertise in **Data Modeling**, including Schema Design, Normalization & Denormalization, **Star & Snowflake Schema**, Fact-Dimension Tables, Indexing Strategies, and Data Partitioning, ensuring optimized database performance and query efficiency.
- Experienced working in **Agile/Scrum teams**, collaborating cross-functionally with data scientists, business analysts, and engineers to deliver scalable, high-performance data solutions.

Technical Skills:

Big Data & Data Processing	HDFS, MapReduce, Sqoop, Flume, Spark, Kafka, Airflow, Flink, Hive, Pig.
ETL & Workflow Orchestration	AWS Glue, Step Functions, EMR, Apache Airflow DAGs, SQLAlchemy, AWS Data Pipeline, GCP Dataflow
Data Modeling & Query Optimization	Star & Snowflake Schema, Fact-Dimension Tables, Normalization, Partitioning, Query Execution Plans, Materialized Views, Indexing Strategies
SQL & NoSQL Databases	PostgreSQL, MySQL, Microsoft SQL Server, Oracle, Teradata, MongoDB, Cassandra, DynamoDB

DataLake & Warehousing	AWS S3, Apache HDFS, Lake Formation, Glacier, EBS, Redshift, Snowflake, BigQuery
Data Visualization & Reporting	QuickSight, Athena GCP Looker, Tableau,Power BI.
Programming Languages	Python, SQL, Scala, PL/SQL, Bash Scripting.
Cloud Ecosystem	AWS and GCP
Machine Learning & AI	AWS SageMaker, Bedrock, GCP Vertex AI
Security & Compliance	AWS IAM, KMS, Secrets Manager, GCP IAM, GDPR, HIPAA, SOC 2
APIs & Web Frameworks	FastAPI, Flask, Swagger UI, OpenAPI, AWS API Gateway, Postman
Infrastructure, DevOps & CI/CD	Terraform, AWS CloudFormation, GitLab CI/CD, AWS CodePipeline, Jenkins, GCP Cloud Build, Docker, Kubernetes

Professional Experience:

Client: Swift Role: Cloud Data Engineer	May 2023 - present
--	---------------------------

Project Summary:

As a **Cloud Data Engineer at Swift**, designed and implemented **high-performance, scalable AWS data pipelines** for **real-time and batch data processing**, enabling seamless **data transformation, analytics, and automation**. Developed **serverless and containerized architectures** to **reduce operational overhead** and **enhance data accessibility**. Focused on **cost efficiency, security compliance, and infrastructure automation** using **Terraform, CI/CD, and observability tools**, leading to **improved system reliability and faster deployments**.

Key Responsibilities:

- Designed and **automated batch & streaming data pipelines** using **AWS Glue** for ETL, **Apache Airflow** for orchestration, and **Kinesis** for real-time ingestion, improving **data processing efficiency by 50%**.
- Developed **real-time streaming solutions** using **Kafka & Flink**, reducing **event-processing latency by 60%** and enabling near-instant analytics.
- **Optimized analytical workloads** in **Redshift & Snowflake**, leveraging **partitioning, materialized views, and caching**, cutting **query execution time by 35%** and **compute costs by 25%**.
- Led an **on-prem to cloud migration** using **AWS DMS & Glue**, ensuring **99.99% uptime** with **minimal downtime**.
- **Automated ETL workflows** with **Airflow & AWS Step Functions**, eliminating **90% of manual interventions** and improving **pipeline reliability**.
- **Built a secure, cost-efficient data lake** using **AWS S3, Lake Formation, and Databricks (Apache Spark)**, enhancing **structured & semi-structured data management**.
- Developed **RESTful data APIs** using **FastAPI & API Gateway**, enabling real-time access to **1M+ records/day** with **40% faster response times**.
- **Implemented infrastructure automation** using **Terraform & AWS CloudFormation**, reducing deployment time by **60%** and ensuring consistency across environments.
- Developed **CI/CD pipelines** for data pipeline deployments using **AWS CodePipeline, CodeBuild, and GitHub Actions**, improving **deployment speed by 50%**.
- **Enhanced security & compliance** by enforcing **IAM-based access control, AWS Secrets Manager, and KMS encryption**, ensuring **GDPR & HIPAA compliance**.
- Orchestrated **data workflows** with **Airflow & Step Functions**, reducing **pipeline failures by 50%** and improving workflow automation.
- **Optimized AWS costs** via **Auto Scaling & S3 lifecycle policies**, reducing cloud expenditure by **20%**.
- Configured **monitoring & alerting** using **AWS CloudWatch, Prometheus, and Grafana**, reducing incident response time by **40%**.
- Designed **real-time analytics dashboards** using **QuickSight & Athena**, improving **business decision-making by 30%**.

Client: WTW
Role: Data Engineer

Apr 2020 - Jul 2022

Project Summary:

As a Data Engineer at WTW, I built and optimized **scalable AWS-based data solutions** for **real-time and batch data processing**. The project involved developing a **centralized data lake, automating ETL/ELT workflows, and designing real-time streaming architectures** to improve data accessibility, analytics, and

cost efficiency. Focused on data reliability, **workflow automation**, and **infrastructure optimization** using **Terraform, Apache Airflow, and Snowflake**, ensuring fault tolerance, high availability, and performance tuning for large-scale data pipelines.

Key Responsibilities:

- **Developed & automated ETL/ELT pipelines** using **Apache Airflow, AWS Glue, and dbt**, increasing **data integration speed by 50%** while reducing manual intervention.
- **Optimized large-scale data transformations** with **Glue & Spark**, improving processing speed by **45%** and reducing ETL failures by **30%**.
- Designed and **implemented schema evolution** using **AWS Glue Data Catalog**, ensuring seamless data ingestion across structured & semi-structured sources.
- Built a **real-time data streaming architecture** using **Apache Kafka & AWS Kinesis**, reducing **event-processing latency by 60%** and ensuring near-instant analytics.
- Enabled event-driven processing using **AWS Lambda & Flink**, reducing **manual data validation efforts by 40%**.
- Implemented deduplication & enrichment in real-time pipelines, **improving data accuracy for analytics by 35%**.
- **Led migration on-prem datasets to AWS S3, Snowflake, and Redshift**, ensuring zero downtime and a **45% boost in query performance**.
- **Optimized Redshift & Snowflake queries** using partitioning, materialized views, and result caching, **reducing compute costs by 25%**.
- Built and maintained **analytical data marts** using **AWS Athena, Redshift, and Snowflake**, **improving business reporting efficiency by 50%**.
- **Automated infrastructure provisioning** for data pipelines using **Terraform & CloudFormation**, **reducing deployment time by 70%**.
- **Developed CI/CD pipelines** for ETL workflows using **GitLab, Jenkins, and AWS CodePipeline**, **improving release cycle efficiency by 50%**.
- Integrated logging & monitoring with **AWS CloudWatch & Datadog**, **reducing incident response time by 40%**.
- Implemented **fine-grained access control** using **AWS IAM, Lake Formation, and KMS encryption**, ensuring GDPR & SOC 2 compliance.
- Developed **automated data validation & auditing workflows**, improving **compliance tracking by 30%**.