

SIDDHARTHA KUMAR

Data Engineer

Bengaluru, India | +91 9116155496 | shivsiddhartha187@hotmail.com | [linkedin.com/in/siddhartha-kumar](https://www.linkedin.com/in/siddhartha-kumar)

PROFESSIONAL SUMMARY

Data Engineer with 5+ years of experience architecting high-throughput distributed data systems and cloud-native pipelines on AWS, Snowflake, and Databricks. Expert in designing event-driven architectures handling 8,000+ concurrent executions with sub-second latency. Proven track record of reducing deployment cycles by 85%, optimizing system performance by 60%, and delivering scalable solutions that process millions of transactions across global markets. Passionate about building reliable, fault-tolerant systems at scale.

TECHNICAL SKILLS

- **Languages & Frameworks:** Python, SQL, PySpark, Pandas, NumPy, Boto3, Flask, REST APIs, Shell Scripting
- **Distributed Systems:** Apache Spark, Apache Airflow, Event-Driven Architecture, Microservices, Message Queues
- **AWS Services:** Lambda, Step Functions, DynamoDB, S3, Athena, Glue, EMR, EC2, API Gateway, CloudWatch, SNS, SQS, EventBridge, IAM
- **Data Platforms:** Snowflake, PostgreSQL, DynamoDB, Hive, Data Lakes, Data Warehouses
- **Data Engineering:** ETL/ELT Pipelines, Dimensional Modeling, OLAP, SCD Type 2, Data Quality, Schema Design, Partitioning, Indexing
- **DevOps & Tools:** Docker, Git, Azure Devops, CI/CD, Terraform(limited exposure), Grafana, JIRA, Agile/Scrum
- **Analytics & BI:** Power BI, Tableau, Databricks, Dataiku DSS

PROFESSIONAL EXPERIENCE

Data Engineer

Tiger Analytics India Consulting Pvt. Ltd. | Bengaluru, India | March 2022 – Present

- Architected developed and deployed high-throughput POS messaging pipeline on AWS, orchestrating 8,000+ concurrent Step Functions executions while maintaining strict 3-minute SLA for processing restaurant transaction data across 15+ international markets.
- Led database migration from PostgreSQL to DynamoDB, implementing optimized partition key strategies and GSI designs that reduced query latency by 95% and improved throughput capacity by 10x for high-volume transactional workloads.
- Designed fault-tolerant Lambda functions with idempotency tracking, circuit breaker patterns, and soft rejection strategies using AWS Lambda Powertools, achieving 99.9% message processing reliability.
- Built multi-tenant shipment tracking portal with sub-minute data freshness serving 50+ enterprise clients, reducing support escalations by 40% and stabilizing compute costs through intelligent query optimization.
- Engineered Model Repository platform with Snowflake-backed microservices, cutting ML model deployment time from 14 days to 2 days for 50+ data scientists and establishing auditable lineage tracking for regulatory compliance.
- Automated marketing-mix research workflows through production-grade APIs, achieving 80% improvement in operational efficiency and reducing manual intervention by 90% using Python, SQL, and Snowflake.

Trainee Data Engineer

Futureense Technologies Pvt. Ltd. | Bengaluru, India | November 2021 – February 2022

- Developed distributed ETL POC pipelines for clients using AWS EMR, HDFS, Hive, and PySpark, processing 500GB+ data volumes with optimized partitioning strategies for analytical workloads.
- Created interactive Tableau dashboards with real-time data refresh, enabling data-driven decision making and reducing report generation time by 65%.

Graduate Apprenticeship – Data Analytics

Bosch Limited (Robert Bosch India) | Jaipur, India | September 2019 – December 2020

- Digitized executive KPI reporting using Power BI with automated data pipelines, reducing reporting cycle time by 70% and improving operational visibility for plant leadership.
- Conducted root cause analysis on manufacturing IoT sensor data, reducing Mean Time Between Failures (MTBF) by 20% and Mean Time To Repair (MTTR) by 25%, saving \$150K+ annually in maintenance costs.

KEY PROJECTS

POS Messaging Pipeline – QSR Client (Subway)

October 2025 – Present | AWS Step Functions, Lambda, DynamoDB, S3, CloudWatch, EventBridge, ODS(AuroraDB PostgreSQL)

- Architected and developed serverless event-driven pipeline partially migrating from PostgreSQL to DynamoDB, handling 8,000 concurrent Step Functions executions with Express Workflows optimized for high-throughput, low-latency transaction processing.
- Implemented atomic DynamoDB counter systems for distributed key generation, ensuring globally unique identifiers across concurrent executions without race conditions.
- Designed comprehensive error handling with exponential backoff, dead-letter queues, and soft rejection strategies, achieving 99.9% successful message processing within strict 3-minute SLA requirements.

Customer Portal – Logistics Client

September 2025 – October 2025 | AWS Athena, DynamoDB, Lambda, API Gateway, S3, CloudWatch

- Optimized Athena query patterns with intelligent partitioning and caching strategies, reducing query costs by 60% and improving P95 response times from 8 seconds to under 2 seconds.

Model Repository – Pharmaceutical Client

October 2024 – September 2025 | Python, SQL, Flask, Snowflake, Dataiku DSS, REST APIs

- Built centralized ML model registry with automated lineage tracking, version control, and metadata management, reducing model deployment time from 14 days to 2 days (85% improvement).
- Developed RESTful microservices orchestrated via Dataiku, enabling 50+ data scientists to self-serve model artifacts while maintaining compliance with pharmaceutical regulatory requirements.

Data Fabric Platform – Internal Asset

AWS, Databricks, Airflow, PySpark, Grafana, PostgreSQL

- Engineered reusable data platform with Airflow-orchestrated PySpark pipelines, standardized data contracts, and automated schema evolution, achieving 60% faster processing for downstream ML applications.
- Implemented real-time observability using Grafana dashboards with custom metrics, enabling proactive incident detection and reducing MTTR by 50%.

EDUCATION

Bachelor of Technology in Electronics & Communication Engineering Jaipur National University | 2019

CERTIFICATIONS

AWS Certified Cloud Practitioner | Microsoft Azure Fundamentals (AZ-900) | Microsoft Azure Data Fundamentals (DP-900) | Microsoft Azure AI Fundamentals (AI-900) | Astronomer Airflow 3 Fundamentals & DAG Authoring | Databricks Generative AI Fundamentals | Google Generative AI Fundamentals

HONORS & ACHIEVEMENTS

Astronomer Champions Program for Apache Airflow (2024) – Recognized for expertise and contributions to the Airflow community

Languages: English (Fluent), Hindi (Native)