

SOMASEKAR MAKIREDDY

Email: somasekar.makireddy@gmail.com

Contact: +1 408-981-7285

LinkedIn : <https://www.linkedin.com/in/smakireddy/>

PROFESSIONAL SUMMARY

- Seasoned Lead Data Engineer with 14+ years of experience in software and data engineering, transforming fragmented data into scalable, unified platforms that deliver measurable business value.
- Expert in solving complex data challenges across banking, retail, and technology using modern cloud architectures (AWS, GCP), distributed technologies (Spark, Hive, Flink, Kafka, Hudi), enterprise OLAP data warehousing (Snowflake, BigQuery), OLTP (Oracle, PostgreSQL) and infrastructure-as-code.
- Skilled in designing fault-tolerant, high-volume pipelines with strict SLA adherence and regulatory compliance (SOX, CCPA, GDPR). Architected scalable data models, batch and streaming pipelines, data quality frameworks, and delivered actionable dashboards enabling data-driven decision-making.
- Proven technical leader and mentor, scaling data engineering teams, executing cross-functional initiatives, architecting enterprise infrastructure, and establishing operational frameworks to ensure reliability, quality, and continuous improvement.
- Leveraging GenAI tools to enhance engineering efficiency: automated SQL generation, documentation, code scaffolding, and built agentic workflows using LangChain for pipeline monitoring and remediation, reducing manual effort and accelerating delivery timelines.

TECHNICAL SKILLS

Data Engineering	Apache Spark, PySpark, Apache Kafka, Apache Flink, Apache Hive, Apache Airflow, Datalake (Apache Hudi , Iceberg), DBT, Spark SQL, Databricks, Sqoop, Hadoop, MapReduce, System Architecture, Data modelling , Design Patterns
Programming Languages & Scripting	Python, SQL, Java, Scala, Unix/Shell Scripting
Databases & Data Warehouses	Oracle, Snowflake, BigQuery, Redshift, MySQL, PostgreSQL, DynamoDB, HBase
Cloud & Infrastructure	AWS (MWAA, Managed Flink, EMR, Lambda, S3, RDS, Aurora PostgreSQL, Redshift, DynamoDB, CloudWatch, KMS, Secrets Manager), Google Cloud (Dataproc, BigQuery, Cloud Storage, Dataplex), Terraform (IaC) , Cloudformation
Data Visualization & Governance	Oracle Reports, Tableau, Looker, Collibra, Dataplex
DevOps & Tools	Github, Jenkins, Git, Jira, docker, automation, DBeaver, Control-M, Tidal, Cron

PROFESSIONAL EXPERIENCE

Goldman Sachs

01/2025 - Current

VP-Senior Software Engineer

As part of the Transaction Banking (TxB) organization, I led key data engineering initiatives to strengthen the platform's technical foundation, improve architectural and operational excellence across teams using AWS services like Managed Airflow (MWAA), EMR (Spark), Snowflake, and Terraform.

- Built from-scratch Terraform IaC modules, enabling TXB's first-ever Managed Airflow (MWAA) environment, and developed/enhanced reusable modules for Managed Flink, DMS, Aurora PostgreSQL, and DynamoDB.
- Led the collaboration with app teams to migrate pipelines from self-hosted Airflow to Managed Airflow (MWAA), enabling seamless orchestration of 200+ data pipelines with ingestion, transformation, and validation of data while reducing operational overhead by 40%.
- Designed and developed Snowflake ELT pipelines using SQL, Python, and Managed Airflow (MWAA) to support analytics, reporting, and ML workloads.
- Optimized Spark and Flink workloads through performance tuning, intelligent auto-scaling, and non-prod cluster shutdowns, achieving 25% cost reduction and 20% throughput improvement.
- Implemented automated mTLS certificate rotation for Managed Flink using AWS boto3, CDK, Lambda, Secrets Manager, and Terraform, eliminating manual renewals and ensuring continuous compliance.

Technologies: Python, Terraform, JAVA,, AWS Managed Airflow, Spark, Snowflake, DBT, Glue, Secret Manager, Managed Flink, DMS, AWS RDS Oracle, Aurora PostgreSQL, DynamoDB, agile etc.

Walmart Inc. Dallas, TX.
Sr. Data Engineer

08/2021 –01/2025

As part of FDF (Finance Data Factory) team, Led data engineering teams as technical leader and mentor for EComm Finance team in building a fully scalable, resilient data pipeline by establishing best practice standards and guidelines, thus increasing efficiency to reduce cloud costs. This application now fully migrated to GCP from on-prem servers and moved the computing into Data Lake and Lakehouse Architecture with help of Dataproc(Apache Spark) and Apache hudi open source table formats.

- Managing ownership of Sam's Ecomm, Marketplace, Directed Spend, Promotions, AML, CSA, and Glass modules, ensuring resilient data pipelines and delivering high-quality data to cross-functional stakeholders including data science and ML teams.
- Experienced in building high-performance engineering teams, providing technical leadership, and allocating tasks based on project priorities.
- Designed and developed end-to-end data pipelines integrating Kafka, external tables, structured/unstructured files(csv, parquet, orc and avro), external APIs, and Sqoop into the silver layer.
- Led cross-team collaboration across engineering and platform groups to deliver reliable, production-grade data. Built scalable prod/non-prod Lakehouse architectures with ACID features using Spark, Apache Hudi, BigQuery, Hive, and Google Cloud services.
- Developed CI/CD pipelines using Jenkins, Looper, and Proximity. Delivered analytics and reporting with Tableau and Looker, and strengthened data governance, lineage, and quality using Collibra, Dataplex, and Ensure.
- Actively work with users on business requirements to implement defined Data Strategy by proactively finding solutions for points of failure in 4-In-The-Box discussions.

Technologies: Scala, Python, SQL , Apache Spark, Apache Hudi, Apache Hive, Big Query, GCP, Airflow, Automic, Tableau, Jenkins, Agile etc.

SRS Consulting Inc. Client: Experian PLC.
Sr. Data Engineer

08/2019 – 08/2021

- Led a data engineering team to design and develop end-to-end data pipelines and analytical solutions for Experian Ascend Marketing using AWS Big Data cloud technologies.
- Designed and developed data pipelines for CCPA data load, search, and file-reuse functionalities using AWS EMR (Spark) and Lambda functions.
- Designed and developed end-to-end ELT pipelines on Snowflake using SQL, Spark, and Airflow, including automated ingestion, staging, core, and consumption layers supporting analytics, reporting, and ML workloads.
- Implemented Airflow workflows with advanced branching and custom operators to copy data from S3 to Snowflake/Redshift and unload from Snowflake to S3.
- Developed encryption solutions for PII data using AWS KMS (at rest) and Base64 (field-level), and built Java REST APIs to fetch OVPIN from Omni View API client.

Technologies: Python, Scala, Java, Apache Spark, Snowflake, Hive, Kafka, SQL, Airflow, Jenkins, AWS..etc.

SRS Consulting Inc. Client: Cisco Systems

08/2019-08/2020

- Led a team of data engineering in the designing and developing the Cisco Customer Success – Health Index Centralized Rule Engine to help users/CSE/ Business owners to change rules/Formulae to Calculate Health Index Metrics.
- Designed and developed the data pipelines for variety of business applications and implemented data warehousing solutions using Hadoop, hive and spark
- Performing ETL process on HDFS (Hadoop, Spark), Snowflake/Teradata DB's and creating workflows. Worked with stakeholders to gather requirements and define project scope.

Technologies: Python, Scala, Apache Spark, Hive, Teradata, Snowflake, SQL, AWS, Json, orc, parquet, Unix Scripting. Etc.

MapR Technologies, Santa Clara, USA, CA

01/2019 - 08/2019

Solutions Developer

- Provided professional services to MapR customers by designing and building custom big data solutions using following technologies used on on-premises or cloud environments.
- Designed and developed distributed solutions using Big Data Technologies and its eco systems like HDFS, HIVE, Sqoop and Spark Core API's and Spark DF/SQL, Kafka and NoSQL DB (HBase/MapR-DB).

Technologies: Java/J2EE, Python, Apache Spark, Sqoop, Hive Kafka, SQL, Hadoop MapR Distribution...etc.

Tata Consultancy Services, Client: Cisco Systems Ltd., Bangalore (India), San Jose, CA

08/2011 - 01/2019

Team Lead/ETL Developer

- Designed, Development, conversions, Customizations, implementation of business applications using Oracle E-Business Suite (EBS) R12 (OM, AR) and related supporting OLTP applications like Revenue recognition (NGCCRM) and reporting (RRR) applications.
- Led a team into designing and developing the data pipelines for a variety of business applications and implemented data warehousing solutions using Informatica, Teradata and Oracle.
- Designed distributed, reliable, and scalable data pipelines to ingest and process data in near real-time using various technologies.
- Designed and modified ETL (Informatica) mappings and loading data. Implemented Apache Kafka Consumers for various topics to receive header and booking information
- Responsible for requirements gathering, data modeling and developing new components and enhancing existing components.
- Responsible for loading that data into HDFS using Sqoop based on the business requirements and data cleansing and input missing data.
- Responsible for handling different optimization techniques in HIVE using partitioning, bucketing.
- Designed and developed the SQL/PLSQL ETL components
- Used job scheduling tool (\$U) to Create, Launch, Monitor

Technologies: Java, Python, Oracle ERP(AR,OM), Apache, Hive, Kafka, SQL,PLSQL, Unix Scripting...etc.

Oracle Financial Services, Chennai, India, Client: Citi Bank

11/2010 - 07/2011

Staff Consultant

- **Project Name:** TPS 2(Treasury Product Systems)
- Designed and developed the database modelling and ETL pipelines.Migrated UNIX shell scripts (from TPS1) to Oracle packages into TPS2 system. Wrote database PL/SQL procedures, Functions, Packages and Creating/modifying database tables, triggers, and collections. Wrote complex SQL queries and tuning SQL's and PLSQL code.

Technologies: Oracle SQL,PLSQL, Unit testing, Unix Scripting...etc.

3i Infotech Ltd, Chennai, India, Client: AIG MEMSA

06/2008 - 11/2010

Software Engineer

- **Project Name:** PREMIA 9 and Premia 10 – KENINDIA Data migration.
- Developed PL/SQL Procedures, Packages, Sub Programs, Cursors and Triggers for developing new requirements. Wrote SQL queries to pull data from vario

Technologies: Oracle SQL, PLSQL, unit testing, Forms, Reports, Unix Scripting...etc.

EDUCATION

- **Master of Computer Applications (MCA):** Computer Applications, Sri Venkateshwara University - Tirupati, AP. India
- **Bachelor of Science (BSc. Computer Science),** Sri Venkateshwara University - Tirupati, AP. India