

Marmik Patel

Spring Hill, KS 66083 | +1 816-456-7598 | marmikpatel262@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

SUMMARY

Senior Data Engineer with 8+ years designing scalable data platforms using Hadoop, Spark, and Snowflake. Developed migration factories handling 700+ databases and 1500+ ETL jobs per hour, achieving 1.2B row ingests and up to 80% query latency reduction. Skilled in Python, SQL, and Terraform, with a strong record mentoring teams and building robust, compliant data pipelines.

SKILLS

- **Data Engineering & Big Data:** Hadoop, HDFS, YARN, Spark 2/3, PySpark, Hive, Snowpark, Snowflake, Databricks, Kafka, Livy, Sqoop
- **Cloud & DevOps:** AWS (S3, EMR), Azure, Docker, CI/CD, Git, Ranger, Knox, Kerberos, TLS/SSL
- **Programming & Scripting:** Python, Scala, Java, SQL, Bash
- **Orchestration & ETL:** Apache Airflow, Snowflake Tasks/Snowpipe, Apache Oozie, IBM DataStage, ETL/ELT Pipelines
- **Databases & Warehouses:** Snowflake, PostgreSQL, MySQL, Teradata, Data Warehousing
- **Analytics & BI:** Power BI, Tableau
- **Tools & IDEs:** Maven, SBT, IntelliJ IDEA, PyCharm, Eclipse, Unix/Linux, JIRA

PROFESSIONAL EXPERIENCE

Squadron Data Inc.

May 2023 - Present

Senior Solution Architect | Senior Data Engineer

Overland Park, Kansas

- Architected cloud-scale data platform modernizations utilizing Snowflake and Snowpark, facilitating the migration and optimization of over 1500 ETL jobs per hour while upholding data warehousing best practices.
- Established a secure Azure-Snowflake foundation with Private Link, RBAC, and multi-cluster warehouses, integrating CI/CD and monitoring processes that enhanced release speed and maintained a 99.8% SLA, aligning with scalable and secure data platform requirements.
- Mentored and conducted code reviews for a five-engineer team, incorporating automated CI/CD pipelines that reduced release cycles by 40%, and supported infrastructure as code practices.

Client Project | Snowflake Migration Factory & Real-Time Analytics Enablement

Nov 2024 - Present

Senior Solution Architect

Overland Park, Kansas

- Migrated over 700 databases and 4500 tables from Hive to Snowflake using Snowpark-driven automated schema-transformation logic, reinforcing robust ETL pipelines and data warehousing design.
- Designed an Azure-Snowflake network topology deploying Private Endpoints and DNS forwarding to ensure data-in-flight remained secure and off the public Internet, meeting data lake and storage security standards.
- Provisioned multi-cluster warehouses with auto-suspend and auto-resume features, driving 20% credit savings and ensuring burst-load SLAs, while optimizing data processing workflows.
- Implemented role-based access control, integrating Azure AD SSO and automation with SnowSQL and GitHub Actions to streamline user onboarding and support secure data formatting.
- Configured Snowflake Tasks, Streams, and Snowpipe for near-real-time data ingestion and orchestrated cross-database dependencies via Airflow DAGs, in line with modern data stack practices.
- Optimized file formats and micro-partition clustering to sustain over 1500 ETL jobs per hour with sub-second query latency for more than 40 downstream teams, resulting in efficient data retrieval

Client Project | Spark 2 to Spark 3 Upgrade & Cluster Optimization

Dec 2023 - Sep 2024

Senior Data Engineer

Overland Park, Kansas

- Refactored over 400 Spark2 jobs to Spark3, leveraging Adaptive Query Execution, dynamic partition pruning, and skew-aware shuffle to streamline ETL processes and support data platform scalability.
- Optimized cluster configurations by adjusting executor/core ratios, shuffle partitions, GC flags, and I/O buffer sizes, reducing median job memory by 30% and enhancing overall data processing performance.
- Curated the dependency set and resolved JAR conflicts, which accelerated container start-up times by 20%, further supporting agile deployment and performance tuning.
- Implemented advanced query optimizations including bucketed tables, predicate push-down, and rewriting wide joins with Bloom filters, contributing to robust data warehousing solutions.
- Reduced small-file counts by 80% through compaction and dynamic repartitioning strategies, achieving 1.5 to 2x faster end-to-end runtime on daily workloads and aligning with data engineering best practices.
- Authored a Spark 3 performance playbook that included configuration templates and code patterns, facilitating knowledge sharing and mentoring for improved deployment and testing automation.

Squadron Data Inc.

Big Data Engineer

Jan 2018 - May 2023

Overland Park, Kansas

- Engineered the Hive Query Stats Spark application in Scala to surface YARN metrics for capacity planning, directly saving clients approximately \$200K per year by optimizing data warehousing processes
- Led in-place upgrades from HDP to CDP for seven enterprise clusters (managing over 150K tables), implementing critical security measures such as Kerberos, Ranger, TLS/SSL, and developing ACID adoption playbooks to ensure data integrity and compliance.

Client Project | Enterprise Hadoop Modernization (HDP to CDP)

Data Engineer

May 2020 - Jan 2023

Overland Park, Kansas

- Assessed 2,000-node HDP estates by profiling services, evaluating Hive metastore health (for over 150K tables), and auditing security practices to determine optimal migration strategies, supporting scalable data storage solutions.
- Developed detailed 'code-delta' playbooks to flag syntax breaks and adjustments needed for Spark/Hive, Sqoop, and Oozie jobs, ensuring seamless transitions and continuous data pipeline operations.
- Co-authored a modular runbook framework generating environment-specific instructions and automated validation scripts across DEV, UAT, and PROD environments, reinforcing consistent data platform practices.
- Implemented essential security protocols including Ranger policies, Kerberos, Knox gateway, and TLS/SSL, and enabled ACID compliance on Hive/Spark tables, maintaining high standards for data privacy and governance.

Client Project | Teradata to Hadoop Migration Framework

Data Engineer

Mar 2019 - Feb 2020

Overland Park, Kansas

- Analyzed historical Teradata query logs with Spark to identify access trends, recommending optimal HDFS layouts, clustering keys, and partition strategies to enhance the data lake architecture.
- Generated Sqoop-ready table manifests and developed an auto-builder to produce parallel Sqoop scripts for over 600 databases and over 3500 tables, streamlining migration workflows.
- Developed a companion module to scaffold Spark ETL templates for table migrations, facilitating rapid developer onboarding and embedding performance best practices in data pipeline construction.

Client Project | Telecom ETL Optimization & Small File Reduction

Data Engineer

Jan 2018 - Sep 2018

Overland Park, Kansas

- Evaluated and audited legacy Spark and Hive code to identify inefficient data shuffling, hard-coded partitions, and file-layout issues, leading to improved ETL pipeline performance.
- Redesigned the Hive partitioning scheme and file-format strategy with optimized bucket configurations and larger ORC stripes for Tez, significantly mitigating small-file problems and enhancing data retrieval.
- Re-architected the Spark ingest framework to handle 1.2 billion rows per hour, incorporating checkpointing and dynamic repartitioning, achieving a tenfold increase in ingest throughput and reducing Hive query latency from 6 minutes to under one minute.

EDUCATION

University of Missouri – Kansas City

Master of Science, Computer Science

•GPA: 3.95

Jan 2016 - Dec 2017

Kansas City, Missouri

Gujarat Technological University

Bachelor of Engineering, Computer Engineering

•GPA: 3.7

May 2011 - Jun 2015

Ahmedabad, India

CERTIFICATIONS

•SnowPro Core Certified: (Snowflake)