Marmik Patel

Spring Hill, KS 66083 | marmikpatel262@gmail.com | +1 816-456-7598

SUMMARY

- Senior Data Engineer & Solution Architect with **8+ years** building petabyte-scale data platforms on Hadoop, Spark, and Snowflake
- Built migration factories moving **700+ databases / 4500+ tables** and orchestrating **1500+ ETL jobs per hour** with sub-5-minute SLAs
- Expert in performance tuning and cost governance - achieved **1.2 B rows/hour** ingests and cut query latency by up to **80%**
- Hands-on coder (Python, Scala, SQL) who pairs deep technical skills with client-facing leadership for Fortune 500 transformations

WORK EXPERIENCE

Squadron Data Inc., Overland Park, KS

Senior Solution Architect   **Jan 2024 - Present**

- Lead architect for cloud-scale data-platform modernizations, delivering a **Snowflake + Snowpark migration factory (1500+ ETL jobs per hour)** and a Spark 3 upgrade that cut cluster cost by 20%.
- Built out secure Azure - Snowflake foundation (Private Link, RBAC, multi-cluster warehouses) and rolled out CI/CD + monitoring that trims release times and holds **99.8% SLA**.
- Mentor and code-review lead for a five-engineer team; introduced CI/CD pipelines that trim release cycles by 40%

Client Projects:

Snowflake Migration Factory & Real-Time Analytics Enablement

- Migrated **700+ databases / 4500+ tables** from Hive to Snowflake with Snowpark-driven automated schema-transformation logic.
- Designed Azure–Snowflake **network topology**: deployed **Private Endpoints (PE / PEP)**, routed traffic through **Azure Private Link** and DNS forwarding rules to keep all data-in-flight off the public Internet.
- Provisioned **multi-cluster warehouses** (XS → L) with auto-suspend / auto-resume and resource monitors; drove 20% credit savings while meeting burst-load SLAs.
- Implemented **role-based access control (RBAC)**: created custom roles for Consumer Teams vs Platform Ops, integrated Azure AD SSO, and built user-onboarding scripts with SnowSQL + GitHub Actions
- Configured **Snowflake Tasks & Streams + Snowpipe** for near-real-time ingestion; orchestrated cross-database dependencies via Airflow DAGs.
- Tuned file formats & micro-partition clustering to sustain **1500+ ETL jobs per hour** while keeping query latency sub-second for 40+ downstream teams.
- **Impact:** pipeline execution **3× faster**, **99.8% SLA** compliance, and audited zero egress to public endpoints for compliance.

Spark 2 to Spark 3 Upgrade & Cluster Optimization

- **Refactored 400+ Spark2 jobs** to Spark3, exploiting **Adaptive Query Execution (AQE)**, dynamic partition pruning, and skew-aware shuffle.
- **Fine-tuned cluster configs:** adjusted executor/core ratios, shuffle partitions, GC flags, and I/O buffer sizes; trimmed median job memory by 30%.
- **Curated dependency set** (Spark, Hadoop, Hive, Scala libs) and removed shaded JAR conflicts, shrinking container start-up time by 20%.
- Implemented **bucketed tables + predicate push-down**, optimized broadcast vs. shuffle joins, and rewrote wide joins with Bloom filters.
- Slashed small-file count **80%** via compaction strategy and dynamic repartitioning; end-to-end runtime **1.5 - 2× faster** on daily workload.
- Authored a **Spark 3 performance playbook** with config templates and code patterns adopted by 6 peer teams.
- **Impact: 20% reduction** in Yarn-cluster CPU / memory cost while sustaining tighter SLAs for downstream analytics.

Squadron Data Inc., Overland Park, KS

Big Data Hadoop Developer   **Jan 2018 - Dec 2023**

- Designed the Hive Query Stats Spark application (Scala) to surface YARN metrics, informing capacity planning and saving clients >$200 K/yr

- Led HDP → CDP in-place upgrades for seven enterprise clusters (150 K tables), implementing Kerberos, Ranger, TLS/SSL, and ACID adoption playbooks

Client Projects:

Enterprise Hadoop Modernization (HDP → CDP)
- **Assessed 2,000-node HDP estates**: profiled services, Hive metastore health (150K+ tables), and security posture to choose **in-place upgrade vs. side-by-side migration** per cluster.
- Created **"code-delta" playbooks**—flagged Spark/Hive syntax breaks, Sqoop/OOZIE job tweaks, and storage-path changes— so application teams could refactor **before** cut-over.
- Authored and co-owned a **modular runbook framework**; generated environment-specific runbooks and automated validation scripts for DEV → UAT → PROD waves.
- Implemented **Ranger policies, Kerberos, Knox gateway, TLS/SSL,** and enabled Hive/Spark **ACID** tables; handed over post-go-live hardening checklist to platform ops.

Telecom ETL Optimization & Small File Reduction
- **Audited legacy Spark & Hive codebase** to surface inefficient shuffle patterns, hard-coded partitions, and file-layout anti-patterns.
- **Redesigned Hive partition scheme + file-format strategy** (multiple buckets, larger ORC stripes) for Tez; eliminated small-file explosion.
- **Re-architected Spark ingest framework** to sustain **1.2 B rows/hour** with checkpointing and dynamic repartitioning.
- **Result:** 10× higher ingest throughput and Hive query latency cut from 6 min to **< 1 min**.

Teradata to Hadoop Migration Framework
- **Parsed historical Teradata query logs** with Spark to profile access patterns and recommend optimal **HDFS layout, clustering keys, and partition strategy**.
- Generated Sqoop-ready table manifests and an **auto-builder that emits parallel Sqoop scripts** for 600+ databases / 3,500+ tables.
- Produced a companion module that scaffolds **Spark ETL templates** (read, transform, write to Hive) for each migrated table - accelerating developer onboarding and ensuring performance best practices.

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
Databases & Warehouses:	Snowflake, PostgreSQL, MySQL, Teradata
Analytics & BI:	Power BI, Tableau
Tools & IDEs:	Maven, SBT, IntelliJ IDEA, PyCharm, Eclipse; Unix/Linux, JIRA

EDUCATION

Master of Science in Computer Science, University of Missouri – Kansas City, GPA 3.95	*Jan 2016 - Dec 2017*
Bachelor of Engineering in Computer Engineering, Gujarat Technological University – India, GPA 3.5	*May 2011 - Jun 2015*

CERTIFICATION

SnowPro Core Certified (Snowflake)