Jay Singhvi

+1 (206)660-3609 | jsinghvi09@gmail.com | Portfolio | linkedin.com/in/jay-singhvi/ | github.com/jay-singhvi/

Professional Summary

- Results-driven Senior Data Engineer with 10+ years of comprehensive experience architecting, designing, and optimizing enterprise-scale data solutions across cloud and on-premises environments. Demonstrated expertise in transforming complex business requirements into robust data pipelines and analytics platforms that drive strategic decision-making and operational efficiency.
- Architected and operated large-scale data and analytics solutions on Snowflake Cloud Data Warehouse, implementing complex ETL processes that ensured data consistency and accuracy across multiple source systems.
- Leveraged AWS services (EC2, S3, RDS, Lambda, IAM, CloudFormation) to deliver scalable, secure, and cost-optimized data infrastructure that accelerated digital transformation initiatives.
- Developed and managed Azure Data Platform services including Azure Data Lake (ADLS), Data Factory (ADF), Data Lake Analytics, Stream Analytics, Azure SQL DW, and HDInsight/Databricks, successfully building multiple enterprise data lakes.
- Engineered data processing frameworks using Hadoop ecosystem technologies including HDFS, Hive, HBase, Zookeeper, PIG, Sqoop, and Flume to handle petabyte-scale data workloads.
- Configured and administered Hadoop clusters across major distributions including Apache Hadoop and Cloudera, optimizing performance for high-throughput data operations.
- Developed Spark applications using Spark-SQL in Databricks for complex data extraction, transformation and aggregation from disparate sources, uncovering critical business insights into customer usage patterns.
- Executed strategic database migrations from traditional systems to Snowflake, AWS Redshift, and Azure Data Lake, ensuring data integrity throughout transition processes.
- Implemented and maintained both NOSQL solutions (MongoDB, Cassandra, Redis, Couchbase) and relational databases (Oracle, PostgreSQL, MySQL), selecting optimal technologies based on specific workload requirements.
- Developed REST web services for backend systems, integrating Couchbase DB (NoSQL) with client-side Couchbase Lite technology to enable seamless synchronization between server and device-hosted databases.
- Implemented workflow scheduling and orchestration using Airflow, AWS Data Pipelines, and SSIS, ensuring reliable and timely data processing across environments.
- Built real-time data processing solutions leveraging Apache Flink and Apache Kafka, establishing robust stream processing and event-driven architectures.
- Optimized data processing performance through deep understanding of Spark architecture, including Spark Core, Spark SQL, Data Frames, Spark Streaming, and cluster component interactions.
- Led cross-functional teams in designing and implementing enterprise data solutions, effectively bridging business requirements with technical implementation.
- Implemented comprehensive security best practices including authentication, authorization, and data encryption across platforms, ensuring regulatory compliance and data governance.
- Conducted technical training sessions to enhance team proficiency in Snowflake and other data technologies, accelerating adoption of modern data platforms.
- Operated efficiently in Agile environments using CI/CD methodologies, delivering continuous value through iterative development cycles.

Technical Skills

Big Data & Hadoop Ecosystem	Hadoop 3.3/3.0, Hive 2.3, Apache Flume 1.8, Sqoop 1.4, Kafka 1.0.1, Oozie 4.3, Hue, Cloudera Manager
Cloud Technologies	AWS, Glue, EC2, EC3, EMR, Redshift & MS Azure, Snowflake
Data Modeling Tools	Erwin R9.7, ER Studio v16, Snowflake
Packages	Microsoft Office 2019, Microsoft Project, SAP and Microsoft Visio 2019, Share point Portal Server
Other Tools	VSS, SVN, CVS. Docker, CI/CD, Kubernetes
RDBMS / NoSQL Databases	Oracle 19c, Teradata R15, MS SQL Server 2019, Cosmos DB, Cassandra 3.11, HBase 1.2

Testing and defect tracking Tools	HP/Mercury, Quality Center, Win Runner, MS Visio 2016 & Visual Source Safe
Operating System	Windows 10/8, Unix, Sun Solaris
ETL/Data warehouse Tools	Informatica 9.6, SAP Business Objects XIR3.1/XIR2, Talend, Tableau
Methodologies	RAD, JAD, RUP, UML, System Development Life Cycle (SDLC), Agile, Waterfall Model

Professional Experience

Atlantic Union Bank, VA **Oct 2024 - till date**

Sr. Data Engineer

Responsibilities:

- Architected and engineered enterprise-scale data solutions by leading the migration of on-premises data warehouses to Snowflake, resulting in 30% reduction in operational costs while implementing robust data models, schemas, and views to support critical business intelligence initiatives.
- Designed and implemented high-performance real-time data processing pipelines utilizing Apache Flink for fraud detection systems, which decreased fraudulent transactions by 25% through low-latency data ingestion from multiple sources including specialized connectors for databases, cloud storage, and third-party APIs.
- Orchestrated comprehensive cloud infrastructure solutions by developing scalable AWS architectures utilizing EC2, RDS, Lambda, and implementing Infrastructure as Code (IaC) with Cloud Formation templates, reducing manual provisioning effort by 60% while optimizing application responsiveness and infrastructure costs.
- Engineered sophisticated data transformation processes leveraging Python, PySpark, shell scripting, and Azure Databricks for large-scale distributed computing, incorporating advanced libraries such as pandas and NumPy for complex data aggregations and comparative analytics across substantial datasets.
- Developed and maintained full-stack web solutions by constructing interactive database models and views with Python/Django, collaborating with cross-functional teams to resolve complex deployment challenges, and implementing responsive front-end interfaces using HTML, CSS, jQuery, and JavaScript.
- Established seamless data integration ecosystems across diverse storage technologies including RDBMS (MySQL, SQL Server) and NoSQL databases (MongoDB, Cassandra, Redis), demonstrating expertise in SQL query optimization and Hibernate framework implementation within Spring ORM architecture.
- Automated end-to-end data pipeline workflows through Azure Data Factory (ADF) with custom JSON deployment scripts and Azure Synapse SQL Activity integration, while implementing Luigi task workflow management for streamlined reporting processes and dependency handling.
- Facilitated robust DevOps practices by collaborating on automated deployment and monitoring for Snowflake environments using Jenkins for multi-environment code deployment and scheduling, alongside industry-standard tools including Jira, Confluence, Git, and GitLab for version control and issue tracking.

Environment: Python, Django, Luigi, windows, Linux, MySQL, SQL, Cassandra, AWS RDS, AWS S3, AWS EC2, Kafka, JSON, Restful API, MVC architecture, GitLab, Agile, Enterprise Scheduler, Bitvise SSH Client, Scrum, JIRA, GIT.

Doordash, CA **Mar 2023 - Sep 2024**

Data Engineer

Responsibilities:

- Architected and executed full-scale migration from on-premises infrastructure to Google Cloud Platform (GCP), implementing a comprehensive data ecosystem leveraging GCP BigQuery, Cloud Storage, and BigTable to establish a robust data lake architecture.
- Spearheaded the successful transition of legacy data warehouses to Snowflake, defining optimal virtual warehouse sizing configurations for diverse workloads while integrating and automating critical data flows.
- Designed and deployed real-time data processing pipelines utilizing Apache Kafka and Spark Streaming for ingesting time-sensitive operational data, enabling near-instantaneous business insights.
- Engineered high-performance ETL pipelines using a sophisticated combination of Python, PySpark, and Snowflake's SnowSQL, implementing rigorous data transformation, cleansing, and quality control protocols.
- Constructed and optimized distributed data processing frameworks by successfully installing Apache Spark and

Apache Flink on Google Kubernetes Engine (GKE), significantly enhancing parallel processing capabilities for large-scale datasets.
- Developed automated data extraction workflows from diverse sources including data lakes and enterprise data warehouses using SQL queries and PySpark, generating actionable business intelligence.
- Implemented comprehensive AWS Identity and Access Management (IAM) policies adhering to least privilege principles, configuring robust security groups, encryption protocols, and CloudTrail logging to ensure regulatory compliance.
- Established and enforced data governance standards across multiple data layers, maintaining consistent business element nomenclature and enhancing overall data integrity.
- Conducted systematic security assessments including vulnerability scans and penetration testing to proactively identify and remediate potential security risks.
- Designed sophisticated monitoring and alerting infrastructure for Apache Flink jobs utilizing Prometheus and Grafana, enabling proactive issue identification and resolution to maintain system reliability.
- Orchestrated complex data workflows through Apache Airflow implementation, automating Change Data Capture (CDC) services and ensuring seamless data synchronization.
- Created comprehensive reporting dashboards for monitoring GCP data load processes, driving site-level reliability metrics and operational excellence.

Environment: Hadoop, GCP, Big Query, Big Table, Spark, Sqoop, ETL, HDFS, Snowflake DW, Oracle Sql, MapReduce, Kafka and Agile process.

Seattle University, WA **Sep 2022 – Dec 2023**
Data Scientist (Research Assistant)

Responsibilities:

- Architected and implemented HIPAA-compliant data pipelines for asthma patient research, leveraging transfer learning methodologies to engineer personalized ensemble prediction models achieving 88% accuracy in asthma onset forecasting - a 20% improvement over traditional classifiers and 12% enhancement compared to established neural network architectures.
- Secured comprehensive CITI Program certifications in research ethics and human subject protocols, establishing governance frameworks for sensitive medical data handling while authoring peer-reviewed research publications on novel machine learning approaches for healthcare analytics.
- Research Projects under review for publishing:
  - **Asthma Patient Research Project (South Korean Hospital Collaboration):**
    - Spearheaded advanced clustering analysis implementing diverse algorithms (K-means, DBSCAN, Affinity Propagation, BIRCH, Mean-Shift, OPTICS) to identify intricate patient cohort relationships while developing a comprehensive environmental analysis framework integrating weather and air quality data.
    - Implemented and optimized open-source Lag-Llama foundation model within a customized prediction framework to reconstruct missing values in time-series patient data, conducting rigorous comparative analysis against traditional imputation methods using multiple statistical metrics (MSE, RMSE, MAE, R^2, MAPE).
    - Performed systematic hyperparameter optimization for Lag-Llama context window configurations, effectively expanding usable training data segments by 37% while maximizing prediction accuracy on missing time-series values.
  - **Agricultural Computer Vision Project (Washington State Farmers Collaboration):**
    - Engineered sophisticated 3D visualization pipeline transforming 2D drone imagery into comprehensive volumetric models while implementing state-of-the-art YOLOv10 object detection architecture for automated plant counting with 89.8% accuracy.
    - Orchestrated comprehensive aerial data collection strategy utilizing drone technology and developed extensive training datasets through meticulous manual annotation of 2,000+ images with 7,000+ unique bounding boxes using open-source platforms (CVAT, Roboflow).
    - Conducted comparative performance analysis between YOLO (v8, v11, and v12) architectures, achieving superior 93.6% accuracy with YOLOv12 while integrating ML-Depth-Pro libraries for precise distance

measurement and size estimation of detected fruits.

Environment: PySpark, Kafka, Snowflake, AWS, Python, SQL, Kubernetes (EKS), ETL/ELT, YOLOv8-12, Machine Learning, Transfer Learning, Time-Series Analysis, HIPAA compliance frameworks.

Yardi Systems, Dubai, UAE **Apr 2019 - Jul 2022**
Data Engineer

Responsibilities:

- Architected end-to-end Big Data processing solutions on AWS, implementing multi-node clusters on EC2 and establishing strict security controls that achieved full compliance with PCI DSS and HIPAA standards, resulting in zero findings during security audits.
- Engineered robust data pipelines that extracted information from heterogeneous sources within the Enterprise Data Lake (MapR Hadoop distribution), employing complex Hive queries, custom UDFs, and optimized partitioning/bucketing strategies to enhance query performance.
- Developed and deployed Python-based incremental and full load processes that transformed data from Hive into Elasticsearch indexes, while implementing reconciliation mechanisms to ensure data integrity between source and target systems.
- Designed and implemented RESTful services using Python Flask to facilitate efficient data writing into Elasticsearch, creating a seamless integration between various data platforms.
- Orchestrated automated data workflows through Oozie, establishing coordinators and schedulers that ensured timely data availability for downstream analytical applications.
- Collaborated extensively with business stakeholders and global architecture teams to gather requirements, document processes, and design data warehouse solutions that aligned with organizational objectives.
- Created sophisticated Tableau visualizations and dashboards for analyzing POS data, implementing action filters, parameters, and calculated sets to deliver actionable business insights.
- Optimized database performance through PL/SQL query tuning and implemented automated backup systems using RMAN with recovery catalog script maintenance, significantly reducing stored procedure execution times.
- Conducted technical orientation sessions for team members and provided expert guidance on implementing analytical and reporting solutions for clients using agile (SCRUM) methodologies.

Environment: AWS, Python, Agile, Hive, Oracle 12c, Tableau, HDFS, PL/SQL, Sqoop, Flume

Yardi Systems, Pune, India **Nov 2016 - Mar 2019**
Data Engineer

Responsibilities:

- Spearheaded multiple client projects with complex Python, SQL, and Tableau components, while developing comprehensive data models that transformed raw information into actionable business intelligence, resulting in enhanced decision-making capabilities across organizational hierarchies.
- Engineered serverless applications utilizing AWS Lambda, API Gateway, and DynamoDB that dramatically reduced infrastructure costs while achieving superior scalability, complemented by implementing robust CI/CD pipelines through AWS Code Pipeline with automated testing frameworks to ensure rapid, error-free deployments.
- Developed sophisticated ETL processes leveraging Python scripts to extract, transform, and load data from diverse source files (HTML, Excel, PDF, Word) into standardized CSV formats and databases, while implementing rigorous data merging, cleaning, and quality control procedures through programmatic data object rules.
- Created detailed source-to-target mapping documentation establishing precise relationships between data structures and transformation rules, while actively participating in logical and physical data modeling discussions that shaped enterprise architecture decisions.
- Constructed advanced Tableau dashboards and visualizations that translated complex datasets into comprehensible business insights, enabling stakeholders to identify trends, anomalies, and opportunities that would otherwise remain obscured in raw data.
- Executed performance optimization initiatives by identifying and refactoring resource-intensive queries,

significantly reducing execution times and improving overall system responsiveness for mission-critical applications.

- Designed comprehensive master data workbooks documenting ETL requirements, mapping rules, and physical data element structures, establishing definitive reference materials that streamlined development processes and enhanced cross-team collaboration.
- Delivered detailed daily and weekly progress reports on data processing activities, meticulously documenting statistics, data quality issues, and implemented resolutions, ensuring complete transparency throughout project lifecycles

Environment: Oracle 10g, UNIX Shell Scripts, MS Excel, MS Power Point, Python, SQL.

JJIT Fintech Pvt. Ltd, Pune, India **Feb 2015 - Oct 2016**
Jr Data Engineer

Responsibilities:

- Architected and implemented sophisticated SSIS packages to extract, transform, and load data from diverse sources (Flat Files, Excel, SQL Server, Raw Files) using SQL Server Data Tools, incorporating robust error handling and advanced logging mechanisms to ensure data integrity throughout processing pipelines.
- Engineered a comprehensive Data Mart utilizing enterprise data warehousing techniques that served as the foundation for downstream reporting solutions, while developing a User Access Tool that empowered business users to create ad-hoc reports and execute analytical queries against the implemented Cube structure.
- Designed and deployed complex ETL packages that leveraged SQL Server partitioning strategies to efficiently load data from staging environments to partitioned tables, implementing incremental load methodologies that significantly reduced processing times and resource utilization.
- Developed intricate database objects including stored procedures, user-defined functions, triggers, and optimized views that enforced data consistency rules and surfaced critical information to application interfaces, strictly adhering to established SQL coding standards.
- Executed comprehensive performance tuning of SQL queries by analyzing execution plans and implementing optimization techniques (indexing strategies, query rewrites, table partitioning), resulting in dramatically improved data load times and enhanced overall system responsiveness.
- Managed complete SDLC processes from system analysis and design through development, testing, and implementation, while demonstrating exceptional troubleshooting capabilities to resolve technical obstacles encountered during development and release phases.

EDUCATION

MS, Computer Science (specialization in Data Science) **Seattle University**, Seattle, WA Sept 2022 – Jun 2024
Recipient of Seattle University's Dean's Honor Roll
Courses: Distributed Systems, Machine Learning, Big Data Analytics, AWS Cloud Computing, Artificial Intelligence

MS, Computer Applications **Symbiosis International University**, India July 2015 - Apr 2018
Courses: Python, Linux scripting, Data Structure Algorithms, Relational Database management, Data Mining and Warehousing

BS, Information Technology **University of Mumbai**, Mumbai, India Jun 2011 – Jan 2015
Courses: Database management, SQL, Linux, Data Warehousing

PUBLICATIONS & CERTIFICATIONS (*Research Papers:* github.com/jay-singhvi/publications)

- **Incremental SMOTE with Control Coefficient for Classifiers in Data Starved Medical Applications**, published in the 26th International Conference on Big Data Analytics and Knowledge Discovery (DAWAK 2024).
- **A Retrieval-Augmented Framework for Meeting Insight Extraction**, accepted to be published in SAC_2025 (The 40th ACM/SIGAPP Symposium on Applied Computing, Track: Intelligent Systems for Digital Era)
- **Hybrid Deep Learning Framework using Transfer Learning as Feature Extractor in Env. Health Risk Prediction**, in peer-review IEEE JBHI 2025
- CITI Program - Responsible Conduct of Research – Engineers | Human Subjects Research for IRB (Faculty, Staff, and Student) *(Other Certificates:* **linkedin.com/in/jay-singhvi/details/certifications/***)*

PROJECTS *(GitHub Portfolio*: github.com/jay-singhvi/*)*

Resonate AI Chatbot *(Tech Stack: Python, Transformers, LangChain, Pinecone, Hugging Face, LLM, RAG, AWS S3 & AWS Transcribe,*

Infra as code, NLP, QLoRA)

- Architected and deployed a production-grade retrieval-augmented generation (RAG) system with semantic graph clustering techniques, attaining 90% BERT similarity scores and 89% precision/recall metrics while implementing sophisticated document chunking algorithms with overlapping segments to preserve contextual integrity across document boundaries.
- Engineered a high-performance vector embedding persistence layer utilizing Pinecone, maintaining 85% cosine similarity retention while optimizing dimensional reduction techniques for balancing query performance and semantic accuracy in high-volume retrieval.
- Designed and executed a comprehensive LLM evaluation framework facilitating systematic A/B testing across OpenAI (GPT-3.5), Anthropic (Claude 3.5), and Google (Gemini 1.5)models, implementing controlled prompting strategies with standardized performance metrics and developing quantitative evaluation pipelines for measuring hallucination rates, factual accuracy, and response relevance.
- Created instrumentation for tracking token utilization, latency profiles, and cost metrics across different model configurations, enabling data-driven decisions while designing a semantic routing architecture that analyzes query intent and complexity to dynamically select appropriate models, optimizing for both performance and computational efficiency.
- Fine-tuned Llama 2 (7B parameter model) utilizing QLoRA (Quantized Low-Rank Adaptation) techniques, achieving parameter-efficient domain adaptation while reducing computational requirements by 70% compared to full-parameter tuning and implementing gradient checkpointing and mixed precision training methodologies to optimize VRAM utilization.
- Constructed specialized training datasets through careful curation of domain-specific examples with robust validation processes, resulting in enhanced response quality for targeted enterprise use cases while contributing to the open-source AI community.
- Developed a distributed inference system with intelligent caching mechanisms that reduced average response latency by 65% while maintaining response quality, significantly enhancing user experience metrics and establishing a continuous deployment pipeline for iterative improvements.

AI-Agentic Synthetic Data Generation: *(Tech Stack: Python, Docker, Anthropic API, Claude AI, CSV manipulation, Environment management, CLI)*

- Architected and implemented a sophisticated containerized AI system with specialized agent architecture, engineering distinct analyzer and generator components that operate in concert to produce synthetic datasets with statistical fidelity to source data distributions while developing a modular, extensible framework that decouples data analysis from generation processes.
- Engineered robust error propagation mechanisms between system components with comprehensive logging and monitoring, implementing an advanced batch processing framework with dynamic batch sizing algorithms that automatically adjust processing parameters based on memory availability and CPU utilization patterns.
- Optimized data throughput with parallelized processing pipelines while maintaining strict data consistency guarantees, engineering memory-efficient data handling routines that minimize footprint during large dataset transformations through intelligent streaming and chunking methodologies.
- Leveraged Anthropic API and Claude 3.5 Sonnet through sophisticated prompt engineering techniques, developing specialized context-aware prompts that preserve statistical properties while implementing adaptive prompting strategies that dynamically adjust instruction specificity based on data complexity.
- Developed comprehensive validation pipelines to verify synthetic data distributions against source datasets, implementing statistical comparison methodologies to ensure generated data maintains essential characteristics of the original information.
- Engineered robust parameter validation, contextual help systems, and intelligent defaults while designing and implementing a secure API key management system with environment-based configuration and just-in-time authentication.
- Published the containerized solution to Docker Hub with comprehensive documentation, continuous integration testing, and versioned releases, facilitating widespread adoption while maintaining quality control over distributed components.

Serverless Employee Management System: *(Tech Stack: Python, AWS S3 & AWS DynamoDB, Docker, AWS EKS, RESTful APIs, OAuth 2.0, Microservices)*

- Architected and implemented a sophisticated cloud-native SaaS platform for enterprise workforce management, leveraging AWS services and microservices architecture while designing specialized data partitioning strategies within DynamoDB that optimized for common access patterns while minimizing read/write capacity consumption, resulting in 40% lower operational costs.
- Engineered multi-tiered storage architecture that intelligently routes data between hot and cold storage tiers based on access frequency analysis, implementing automated lifecycle policies that reduced storage costs while maintaining sub-second query performance without compromising data availability.
- Developed a comprehensive RESTful API ecosystem with granular endpoint permissions and rate-limiting mechanisms, implementing OAuth 2.0 authentication with custom scope definitions while engineering a robust API versioning system with backward compatibility guarantees to protect client implementations.
- Implemented containerized deployment architecture using Docker with optimized multi-stage builds, reducing image sizes by 65% while configuring advanced Kubernetes deployments on AWS EKS with automated horizontal pod scaling based on custom metrics and sophisticated readiness and liveness probes.
- Designed event-driven processing workflows leveraging AWS Lambda functions with carefully calibrated memory allocations and execution timeouts, implementing concurrency management strategies that delivered 5x throughput improvements for batch operations.

- Engineered a fault-tolerant dead-letter queue system with automatic retry mechanisms and intelligent backoff strategies, ensuring reliable data processing even during intermittent service disruptions while enabling secure third-party integrations without service degradation.

Personalized Marketing Campaign Optimizer: *(Tech Stack: Python, Scikit-learn, Pandas, Matplotlib, Seaborn, SMOTE, GridSearchCV, Machine Learning)*

- Architected a sophisticated marketing campaign optimization system using ensemble machine learning techniques (Decision Tree, KNN, Random Forest), achieving 86% prediction accuracy in identifying high-conversion customer segments while developing custom evaluation metrics aligned with business objectives.
- Engineered comprehensive model interpretation methodologies that translated complex algorithmic patterns into actionable marketing insights and strategic recommendations for targeted campaign optimization.
- Implemented advanced class imbalance mitigation strategies combining SMOTE and Random Under Sampling techniques, significantly improving minority class prediction without compromising overall model accuracy.
- Designed robust feature engineering pipelines that transformed raw customer interaction data into predictive indicators, creating compound features that captured complex behavioral patterns while systematically eliminating noise variables.
- Developed sophisticated exploratory data analysis workflows with custom visualization frameworks using Matplotlib and Seaborn, implementing multi-dimensional analysis methodologies that revealed previously undetected patterns across diverse market segments.
- Engineered automated data quality systems incorporating statistical outlier identification algorithms and anomaly detection, substantially improving pipeline integrity prior to model training phases.
- Implemented comprehensive model optimization protocols combining stratified cross-validation with hyperparameter tuning systems using GridSearchCV and custom scoring functions, establishing automated workflows that systematically identified optimal configurations for production deployment.

SQL Query Assistant using Snowflake Cortex Analyst: (Tech Stack: AWS S3, Python, Snowflake, Streamlit, SQL, LLM, Snowflake Cortex LLM)

- Architected and developed an advanced SQL query generation system leveraging Snowflake Cortex Analyst, creating a sophisticated conversational interface that transforms natural language questions into optimized SQL queries for non-technical users.
- Engineered a comprehensive semantic model framework with YAML configurations defining logical tables, dimensions, measures, and custom expressions, enabling high-precision query generation through detailed metadata mapping of complex database.
- Designed and deployed a production-grade Streamlit chatbot interface with robust error handling, request management, and sophisticated session state capabilities, creating an intuitive experience for natural language query composition while maintaining secure connections throughout user sessions.
- Implemented advanced data ingestion workflows for multiple revenue datasets with optimized ETL processes, precise data type handling, and error management protocols to ensure data integrity across the analytical ecosystem.
- Engineered sophisticated API integration with Snowflake Cortex Analyst REST endpoints, implementing secure token-based authentication and efficient request payload structuring while developing intelligent caching mechanisms that significantly reduced API call frequency.
- Designed a verified query repository system for capturing validated SQL patterns, enhancing query accuracy and performance by leveraging proven structures for similar natural language inputs while providing comprehensive error handling with detailed logging.