

GOVINDA NEUPANE

Senior Data Engineer | Snowflake & Snowpark | Big Data (Hadoop, Spark) | Big Data and Distributed Systems | Healthcare Analytics | HIPAA

Kathmandu / Lalitpur, Nepal • +977 980 9213141 • govindaneupane81@gmail.com

LinkedIn: linkedin.com/in/govind-neupane • GitHub: github.com/govindan66

PROFESSIONAL SUMMARY

Senior Data Engineer with expertise in building and migrating large-scale healthcare data platforms. Specialized in Snowflake, Snowpark Python, Hadoop, Spark, Java, and AWS EMR, with deep experience in distributed systems and HIPAA-compliant healthcare analytics. Currently leading a migration team modernizing legacy Hadoop pipelines to Snowflake through Augmented Engineering, driving code conversion, parity checking, and remediation of missing logic while ensuring data accuracy, performance, scalability, and reliability.

ENGINEERING PHILOSOPHY & CAREER GOALS

Philosophy. I believe in building data systems that are not just functional, but elegant — systems that scale gracefully, fail safely, and evolve with the business. Every pipeline I build is designed with observability, testability, and maintainability at its core.

Goals. Seeking senior-level opportunities in data engineering where I can architect and build next-generation data platforms, mentor engineering teams, and drive the adoption of modern cloud-native data technologies.

TECHNICAL SKILLS

Data Engineering: Snowflake, Snowpark, Apache Spark, Hadoop, ETL Pipelines, Data Modeling, Data Validation, Data Reconciliation, Data Quality, Cascading

Programming: Python, Java, SQL, Bash/Shell, TypeScript

Cloud & Infrastructure: AWS EMR, AWS S3, Elasticsearch, Docker, Git

Web Technologies: Django, REST APIs, PostgreSQL, FastAPI

Soft Skills: Collaboration, Documentation, Problem Solving, Code Reviews, Training & Mentorship, Agile/Scrum

PROFESSIONAL EXPERIENCE

Senior Data Engineer

IQVIA | Lalitpur District, Nepal · Hybrid | Full-time | Oct 2026 – Present

Leading the migration team modernizing large-scale healthcare data pipelines from legacy Hadoop systems to Snowflake, driving code conversion through Augmented Engineering while ensuring functional parity and data accuracy.

- Leading the migration team converting legacy big data pipelines to Snowflake using Augmented Engineering (AI-assisted) workflows
- Driving code conversion from Hadoop (Java, Cascading, Spark) to Snowpark Python with accelerated, AI-augmented engineering practices
- Performing rigorous parity checking between legacy and converted pipelines to guarantee functional equivalence and data accuracy
- Identifying missing or incomplete business logic in AI-converted code and remediating gaps to ensure correctness
- Reviewing converted Snowpark implementations and establishing engineering standards across the migration team
- Coordinating cross-functional efforts to resolve data quality issues and keep the migration on track
- Mentoring engineers and leading knowledge transfer on Snowflake, Snowpark, and Augmented Engineering workflows

Data Engineer

IQVIA | Lalitpur, Nepal | Full-time | Nov 2025 – Oct 2026

Engineered modern cloud-native data pipelines on Snowflake, migrating legacy healthcare big data workloads to Snowpark Python with a focus on data accuracy and reliability.

- Built and migrated big data pipelines from Hadoop to Snowflake using Snowpark Python
- Developed scalable Snowpark Python transformations for large-scale healthcare datasets
- Conducted data validation and reconciliation to ensure parity between legacy and modern platforms
- Performed code reviews to validate business logic, transformations, and best practices

- Collaborated with cross-functional teams to resolve data quality issues and clarify requirements

Data Engineer

Cedar Gate Technologies | Lalitpur District, Nepal · Hybrid | Full-time | Apr 2025 – Nov 2025

Supported end-to-end migration of big data pipelines from Hadoop to Snowflake using Snowpark Python, ensuring functional parity and data accuracy across legacy and modern platforms.

- Supported end-to-end migration of big data pipelines from Hadoop (Java, Cascading, Spark) to Snowflake using Snowpark Python, ensuring functional parity and data accuracy
- Performed code reviews on Snowpark implementations to validate business logic, transformations, and adherence to data engineering best practices
- Conducted data validation, reconciliation, and root cause analysis to identify discrepancies between legacy and modern data platforms
- Collaborated with cross-functional and third-party teams to resolve data quality issues, clarify requirements, and improve pipeline reliability
- Developed and executed unit testing strategies for Snowpark pipelines to ensure data integrity, correctness, and regression prevention
- Designed and maintained technical documentation for Snowflake architecture, Snowpark workflows, and data pipeline migration patterns
- Delivered training sessions and knowledge transfer (KT) on Snowflake and Snowpark, improving team adoption of modern data engineering tools

Software Engineer

Cedar Gate Technologies | Lalitpur District, Nepal · Hybrid | Full-time | Dec 2023 – Apr 2025

Designed and developed scalable big data pipelines using Java, Hadoop, and Cascading to process large-scale healthcare datasets, including Eligibility and Medical claims data.

- Designed and developed scalable big data pipelines using Java, Hadoop, and Cascading to process large-scale healthcare datasets, including Eligibility and Medical claims data
- Analyzed and enhanced existing Java and Python-based data processing systems, improving HCC (Hierarchical Condition Categories), MARA (Milliman Advanced Risk Adjusters), and Utilization Metrics logic
- Worked extensively with Hadoop, Apache Spark, and AWS (EMR), and explored Elasticsearch for efficient data indexing, querying, and performance optimization
- Translated complex healthcare domain requirements into technical implementations, ensuring accurate processing of Eligibility, Medical, and clinical datasets
- Improved Utilization Metrics (clinical measures) by refining business logic and aligning with evolving healthcare analytics requirements
- Implemented unit testing, code reviews, and defect resolution practices to ensure data accuracy, system reliability, and high-quality software delivery

Data Engineer

Cedar Gate Technologies | Lalitpur District, Nepal · Hybrid | Full-time | Apr 2023 – Dec 2023

Developed and supported ETL data pipelines using Java, Hadoop, and Cascading on AWS EMR clusters, applying data modeling and cleansing to ensure high-quality datasets for downstream analytics.

- Developed and supported ETL data pipelines using Java, Hadoop, and Cascading on AWS EMR clusters
- Assisted in processing large-scale datasets by implementing data transformations and ensuring efficient pipeline execution
- Applied data modeling techniques to structure and organize data for downstream analytics and reporting
- Performed data cleansing and validation to ensure high data quality and consistency across datasets
- Collaborated with senior engineers to understand pipeline architecture, troubleshoot issues, and optimize data processing workflows

Technical Support Executive

Machnet Technologies Inc | Kathmandu, Nepal · Hybrid | Full-time | Mar 2022 – Apr 2023

Provided technical support and API testing services, resolving tickets and service requests while maintaining compliance and strengthening technical documentation.

- Responsively addressed customer queries and promptly resolved tickets and service requests, ensuring high customer satisfaction

- Participated in the creation and review of technical documents, verifying accuracy and improving the overall documentation process
- Initiated timely site maintenance and recommended corrective measures, improving system reliability and customer service
- Conducted API testing to ensure functionality and compatibility of APIs within the technical support framework
- Maintained control mechanisms for monitoring transactions and their release, guaranteeing support operation integrity
- Adhered to procedures, regulatory guidelines, and compliance requirements, and reported compliance risks to the supervisor

Client Services Executive (Recruiter)

BnB Softed | Zorba Softed | Kathmandu, Nepal · Remote | Full-time | Aug 2021 – Mar 2022

Managed end-to-end processing of client applications for USA-based clientele, acting as a liaison between clients and external organizations while maintaining high data accuracy.

- Facilitated end-to-end processing of client applications, managing the journey from submission to settlement for USA-based clients
- Acted as a liaison between clients and external organizations, coordinating and communicating on their behalf
- Delivered exceptional customer service, addressing inquiries and concerns promptly and professionally
- Maintained a high level of data accuracy with meticulous attention to detail during daily data entry
- Established open lines of communication with supervisors to ensure alignment with team goals
- Embraced additional responsibilities as needed, demonstrating flexibility and commitment to team success

Python (Django) Developer

Techrida | Kathmandu, Nepal · Hybrid | Full-time | Mar 2021 – Sep 2021

Developed web applications and RESTful APIs using the Python Django framework in an Agile environment, following the MVT architectural pattern.

- Collaborated productively with software development team members and Agile stakeholders
- Developed web applications within the Django framework, adhering to the Model-View-Template (MVT) pattern
- Performed CRUD operations using Object Relational Mapping (ORM), streamlining data interactions
- Contributed to the implementation of RESTful APIs using Python and Django, enhancing data accessibility and integration
- Migrated the Django database from SQLite to PostgreSQL while ensuring data integrity throughout the process
- Prepared comprehensive reports and documentation to support fellow development team members

KEY PROJECTS

Hadoop to Snowflake Migration Platform (Data Engineering · In Progress)

A comprehensive migration platform that automates the process of moving petabyte-scale healthcare data from on-premise Hadoop clusters to Snowflake. Features include automated schema mapping, data type conversion, parallel migration execution, and real-time validation dashboards.

- Migrated 50+ Hadoop MapReduce jobs to Snowflake
- Achieved 99.9% data accuracy through automated reconciliation
- Reduced pipeline execution time by 60%
- Zero data loss during migration

Technologies: Snowflake, Snowpark, Python, Apache Spark, Hadoop, AWS S3, SQL

Snowpark ETL Validation Framework (Data Quality · Completed)

Built to ensure data integrity across complex ETL transformations, this framework provides a declarative way to define validation rules, execute them against Snowflake tables, and generate comprehensive quality reports.

- 100+ automated validation rules
- Reduced data quality issues by 80%
- Integrated with CI/CD pipelines
- Supports custom validation logic

Technologies: Python, Snowpark, Snowflake, pytest, Great Expectations, SQL

Healthcare Claims Analytics Engine (Healthcare Analytics · Completed)

A distributed analytics engine built on Apache Spark and Hadoop that processes millions of healthcare claims daily. Provides real-time risk scoring, cost analysis, and quality metrics for healthcare payers and providers.

- Processes 10M+ claims records daily
- Sub-second query response with Elasticsearch
- Supports HCC and MARA risk models
- Deployed on AWS EMR for auto-scaling

Technologies: Java, Apache Spark, Hadoop, AWS EMR, Elasticsearch, Python, SQL

Spark-based Data Processing System (Data Engineering · Completed)

Designed for enterprise-scale data processing, this system handles both batch and near-real-time data workloads using Apache Spark. Includes a job orchestration layer, monitoring dashboard, and intelligent resource management.

- Handles 1TB+ daily data volumes
- 99.5% pipeline reliability
- Automated failure recovery
- Cost-optimized EMR configurations

Technologies: Apache Spark, Python, Scala, AWS EMR, Apache Airflow, Docker

CERTIFICATIONS

Snowflake Data Engineer — Snowflake (2024)

Advanced certification validating expertise in designing and building data engineering solutions on Snowflake.

Snowflake Data Engineer II — Snowflake (2024)

Expert-level certification for advanced Snowflake data engineering patterns and architectures.

Snowflake Fundamentals — Snowflake (2023)

Foundational certification covering core Snowflake concepts, architecture, and features.

Snowpark for Data Engineers — Snowflake (2024)

Specialized certification for building data pipelines using Snowpark Python and Java APIs.

Advanced Prompt Engineering — LinkedIn Learning (2024)

Advanced techniques for crafting effective prompts for large language models and AI systems.

Learning Python — LinkedIn Learning (2022)

Comprehensive Python programming certification covering core language features and best practices.

Programming Foundations — LinkedIn Learning (2021)

Foundational programming concepts including algorithms, data structures, and software design principles.

EDUCATION

BSc (Hons) in Information Technology

Asia Pacific University of Technology & Innovation (APU), Malaysia — studied at Lord Buddha Education Foundation, Kathmandu (APU-affiliated center) | Kathmandu, Nepal | 2018 - 2022 | GPA: 3.01/4.0