Rajesh Kumar Gupta
rajesh21gupta@gmail.com

PROFESSIONAL SUMMARY

- Distinguished Semantic Data Architect and Knowledge Graph Engineer with 15+ years designing enterprise-scale data platforms, Neo4j knowledge graphs, and AI-driven data architectures across Financial Services, Healthcare, and Manufacturing. Proven expertise in Snowflake, Apache Airflow, and GraphRAG with track record architecting investment intelligence platforms integrating structured, unstructured, and semi-structured data sources. Deep technical knowledge of Cypher, graph traversal algorithms, multi-dimensional Ontology design, and semantic technologies (RDF, OWL, TTL). Demonstrated ability to teach advanced graph concepts, implement adaptive schema evolution using AI-driven feedback loops, and design property-based access control for secure multi-tenant environments. Specialized in real-time data ingestion, temporal versioning, data lineage, and thin-pointer architectures for scalable integration with large datasets.

CORE COMPETENCIES

- Data Platform & Orchestration: Snowflake (2+ years), Apache Airflow (4+ years), real-time ingestion pipelines, ETL/ELT design, Kafka streaming, multi-source integration, data quality frameworks, Great Expectations
- Knowledge Graph & GraphRAG: Neo4j (10+ years), GraphRAG, Cypher, graph traversal algorithms, property graphs, MarkLogic semantic database, graph neural networks (GNN), LangChain integration Semantic & Ontology Technologies: RDF/OWL/TTL formats, SPARQL, Prot´eg´e, healthcare ontolo-
- gies (SNOMED CT, RxNorm, ICD, LOINC, FHIR), pharmaceutical ontologies, multi-dimensional schema
- design, ontology alignment
- AI-Driven Architecture: AI-assisted data architecture, adaptive schema evolution using genetic algorithms, LLM integration (GPT-4, Claude), agentic AI frameworks (LangGraph, LlamaIndex), prompt engineering, RAG pipelines
- Data Governance & Quality: DataHub, Amundsen, metadata management, data catalog design, data lineage tracking, role-based access control (RBAC), property-based access control, compliance (MiFID II, FINRA, FDA, EMA)
- Cloud & DevOps: AWS (S3, Lambda, Glue, EMR, SageMaker, Bedrock, Step Functions), Azure, Docker, Kubernetes, CI/CD pipelines, multi-region deployment, high availability architecture, GCP
- Programming & ML: Python (8+ years), Java (6+ years), SQL, PyTorch, Hugging Face Transformers, NetworkX, DGL, PyTorch Geometric
- Market Data Integration: Bloomberg, Reuters, internal trading feeds, financial data harmonization, investment decision traceability, portfolio analytics

PUBLICATIONS & THOUGHT LEADERSHIP

- IEEE Publication: ”An Instance Based Methodology for Merging Domain Ontology” – Ontology integration algorithms with 50+ citations – ieeexplore.ieee.org/document/5395380
- Transportation Research: ”Road Safety Ontology - SafeOn: Overview & Design” – OWL-based semantic framework – sciencedirect.com/article/pii/S2352146524004630
- Conference Speaker: NextGen Banking Summit 2024 – ”Knowledge Graphs and GenAI in Financial Services” – Keynote on GraphRAG and investment intelligence
- In Progress: ”RAGHeat: A Diffusion-Based GraphRAG Framework for Financial Recommendation” –Novel graph neural network architectures
- Technical Documentation: Authored comprehensive guides on Cypher optimization, graph traversal algorithms, GraphRAG architecture, and adaptive schema design used by 100+ engineers

KEY DATA ARCHITECTURE & KNOWLEDGE GRAPH PROJECTS

- Investment Intelligence Platform (Barclays): Architected enterprise data factory with Neo4j knowledge graph as central data layer processing $2.8T ecosystem. Designed multi-dimensional asset ontology, adaptive schema evolution using AI/genetic algorithms, temporal versioning, thin-pointer architecture, and GraphRAG integration. Built on Snowflake with Airflow orchestration, DataHub catalog, and property-based access control on AWS infrastructure.
- Pharmaceutical Data Platform (Johnson & Johnson): Designed production system integrating Neo4j knowledge graph with MarkLogic semantic database using RDF/OWL healthcare ontologies (SNOMED CT, RxNorm, FHIR). Implemented GraphRAG pipeline, Snowflake data harmonization with Airflow

orchestration, temporal versioning, and comprehensive data lineage serving 500+ researchers.
- SwarmCare Healthcare Platform: Architected intelligent health data platform integrating clinical ontologies (SNOMED CT, RxNorm, ICD, LOINC, FHIR) with Neo4j/MarkLogic. Implemented GraphRAG using BioBERT, FHIR-compliant data integration, multi-dimensional patient ontology, and role-based access control across 50+ healthcare providers.
- Manufacturing Intelligence System (S&P Global): Built Neo4j-based knowledge graph with GraphRAG processing 1M+ documents. Integrated Snowflake data warehouse with Airflow orchestration, Amundsen catalog, and real-time Kafka ingestion achieving 85% accuracy for supply chain optimization on AWS.
- RAGHeat Innovation: Invented GraphRAG framework combining GATs with LangChain for financial intelligence. Designed adaptive schema evolution using genetic algorithms, investment decision lineage graph, thin-pointer architecture, and comprehensive teaching materials on GraphRAG concepts.

INDEPENDENT RESEARCH & INNOVATION

RAGHeat: Advanced GraphRAG Framework for Financial Intelligence 2024 – Present

- Invented breakthrough GraphRAG framework combining Graph Attention Networks (GAT), heat diffusion modeling, and retrieval-augmented generation using LangChain/Amazon Bedrock achieving 13.2% improvement in financial recommendation accuracy
- Architected real-time Neo4j-FAISS vector search integration with property graph reasoning achieving sub-1.5s latency for complex financial queries
- Developed adaptive graph schema evolution using genetic algorithm principles enabling AI-driven ontology refinement and automated taxonomy learning
- Implemented investment decision lineage graph with comprehensive traceability from research signals to portfolio execution supporting regulatory compliance
- Designed thin-pointer architecture enabling efficient integration of large-scale market data with graph analytics layer
- Created comprehensive documentation and teaching materials on GraphRAG architecture, Cypher optimization, and graph traversal algorithms

SwarmCare: Intelligent Healthcare Data Platform

- Architected comprehensive healthcare knowledge graph using Neo4j and MarkLogic integrating clinical ontologies (SNOMED CT, RxNorm, ICD-10/11, LOINC, FHIR) for intelligent patient care coordination
- Designed semantic health information exchange platform using RDF/OWL/TTL formats with SPARQL federated queries enabling interoperability across 50+ healthcare providers
- Implemented GraphRAG system for medical literature synthesis combining BioBERT with property graphs for evidence-based treatment recommendations
- Built FHIR-compliant data integration layer with ontology-driven transformation supporting HL7 standards and comprehensive data lineage tracking
- Developed multi-dimensional patient ontology enabling flexible categorization by condition, treatment, provider, and outcome
- Architected role-based access control with property-level permissions ensuring HIPAA compliance and secure patient data visibility
- Established data quality framework with automated validation, anomaly detection, and comprehensive metadata management

EDUCATION

Master of Technology, Software Engineering – National Institute of Technology, Allahabad, India (2009)
Bachelor of Engineering, Computer Science – Rajiv Gandhi Technological University, Bhopal (2007)

Research Focus:

Graph Learning,
Data Architecture, Ontology Engineering,
Semantic Web Technologies

CERTIFICATIONS:

Neo4j Graph Dta Science Professional — Neo4j Certified Professional
Snowflake Data Engineer — Snowflake SnowPro Core
Apache Airflow Developer — AWS Data Analytics Specialty — AWS Solutions Architect DataHub Metadata

Management — LangChain AI Development
Awards & Recognition: National Mathematics Olympiad Winner — IEEE Best Paper Award — NextGen Banking Summit Keynote Speaker

TECHNICAL PROFICIENCY

Data Platforms: Snowflake (2+ years), Apache Airflow (4+ years), Kafka, Apache Spark, dbt, Databricks
Knowledge Graph: Neo4j (10+ years), MarkLogic, GraphRAG, Cypher, Gremlin, SPARQL
Data Governance: DataHub, Amundsen, Great Expectations, Collibra, Alation, data catalog design, metadata management
Ontology & Semantic: RDF/OWL/TTL, Prot´eg´e, Apache Jena, RDFLib, SNOMED CT, RxNorm, ICD, LOINC, FHIR
Programming: Python (10+ years), Java (5+ yrs), SQL, Cypher, SPARQL
AI/ML: LangChain, LlamaIndex, LangGraph, PyTorch, TensorFlow, Hugging Face Transformers, OpenAI API
Cloud: AWS (S3, Lambda, Glue, EMR, SageMaker, Bedrock, Step Functions, CloudWatch), Azure, Docker, Kubernetes
Graph Technologies: NetworkX, DGL, PyTorch Geometric, GraphFrames, TigerGraph, Amazon Neptune, Gephi, Nebula Graph, GraphDB(ontotext)

PROFESSIONAL EXPERIENCE

Bessemer Trust, Somerset, NJ **OCT 2025– Present**
Data Architect (Ontology & Knowledge Graph)

Responsibilities:

- Insurance Advisory Document Productionization: Architected end-to-end semantic processing pipeline for high-net-worth insurance advisory documents from PURE, Chubb, and Berkley One carriers, extracting policy structures, coverage clauses, endorsements, and underwriting attributes into a unified RDF/OWL knowledge graph enabling cross-carrier comparison, gap analysis, and advisor-facing recommendation workflows for ultra-high-net-worth client portfolios.
- Multi-Domain Enterprise Ontology Development: Designed and curated production-grade ontologies spanning insurance (P&C, life, umbrella, excess liability), wealth planning (trusts, estates, gifting, tax strategies), and equity/investment domains using OWL 2 DL, RDF, TTL, and SHACL, achieving semantic interoperability across 15+ business capability areas and aligning with industry standards (FIBO, ACORD).
- Enterprise-Level Bessemer Trust Ontology: Led architecture and delivery of the firm-wide enterprise ontology framework establishing a single source of semantic truth across client, advisor, portfolio, product, and regulatory entities, enabling 360-degree client views and reducing data silos across investment, trust, tax, and family office service lines.
- Ontology Release Management & Governance: Engineered a Git-based ontology release management pipeline with semantic versioning (SemVer), automated SHACL validation, CI/CD-driven deployment to Amazon Neptune, and a formal ontology governance framework covering change request workflows, stewardship councils, naming conventions, deprecation policies, and backward compatibility guarantees.
- Enterprise Data Lineage Strategy: Architected end-to-end data lineage strategy tracing data flow across Bessemer Trust's investment, operations, compliance, and client-reporting departments, mapping source-to-consumption paths through 50+ pipelines and enabling regulatory traceability, impact analysis, and root-cause investigation for data quality issues.
- Ontology & Knowledge Graph Visualization Platform: Built an in-house interactive visualization tool for browsing, querying, and exploring the enterprise knowledge graph with class hierarchies, property paths, SPARQL/Cypher query interfaces, force-directed graph layouts, and subgraph drill-downs, accelerating ontology adoption among non-technical business stakeholders and SMEs.
- AWS Neptune Integration & Analytics Workflows: Established secure AWS Neptune connectivity with IAM-authenticated SPARQL/Gremlin endpoints, automated bulk-load pipelines from S3, and local extraction workflows for offline ontology analysis, conformance testing, reasoning validation, and graph metric tracking (node degree, centrality, clustering coefficients).
- GraphRAG & Multi-Agent Query Systems: Engineered a LangGraph-orchestrated multi-agent framework including a GraphRAG agent, Text2Cypher agent (Neo4j), and Text2SPARQL agent (Neptune RDF) leveraging Claude/GPT-4 LLMs with few-shot prompting, schema-aware grounding, and self-correction loops, enabling natural language querying of the enterprise knowledge graph for advisors and analysts.

Barclays – Whippany, NJ **May 2023 – Oct 2025**
AVP, Data Architect & Knowledge Graph Engineering Lead

Responsibilities:

- Neo4j Knowledge Graph Architecture: Architected enterprise-scale Neo4j-powered financial knowledge graph representing $2.8T trading ecosystem, serving as central data layer for investment intelligence with 500+ entity types and 2M+ nodes
- Investment Decision Traceability: Designed comprehensive investment decision lineage graph linking research reports, investment strategies, portfolio actions, and regulatory reporting into unified traceability system ensuring full audit trails
- Multi-Dimensional Asset Ontology: Developed sophisticated multi-dimensional asset ontology using RDF, OWL, TTL formats enabling flexible categorization, cross-asset analysis, and semantic reasoning across equities, fixed income, derivatives, and alternatives
- Adaptive Schema Evolution: Architected AI-driven schema evolution system using genetic algorithm principles and reinforcement learning, enabling automated ontology refinement based on user queries and market dynamics achieving 35% improvement in schema adaptability
- Temporal Versioning: Implemented comprehensive temporal versioning for time-aware queries enabling historical context analysis, point-in-time valuations, and regulatory compliance lookback with
- microsecond precision
- Thin-Pointer Architecture: Designed thin-pointer architecture enabling scalable integration of petabyte-scale Snowflake datasets with Neo4j graph layer, optimizing storage efficiency and query performance by 60%
- GraphRAG Integration: Pioneered GraphRAG implementation transforming 10TB+ daily unstructured regulatory documents, research reports, and market commentary into queryable graph entities using LangChain and GPT-4 with 92% entity extraction accuracy
- Property-Based Access Control: Implemented sophisticated property-based access control framework ensuring secure multi-tenant data visibility with fine-grained role-based permissions supporting 500+ portfolio managers and analysts
- Snowflake Data Platform: Architected Snowflake data warehouse with 50+ integrated market data sources (Bloomberg, Reuters, internal feeds) processing 5TB+ daily with automated data quality validation and lineage tracking
- Airflow Orchestration: Designed and maintained 200+ Apache Airflow DAGs orchestrating real-time data ingestion pipelines with sub-100ms latency, complex ETL/ELT workflows, and automated data catalog updates achieving 99.99% uptime
- Data Catalog & Metadata Management: Established comprehensive DataHub-powered data catalog with automated metadata registration, business glossary, data lineage visualization, and quality metrics tracking 1M+ financial data assets
- Data Quality Framework: Implemented end-to-end data quality approach using Great Expectations with 500+ validation rules, automated anomaly detection, and continuous monitoring ensuring 99.9% data accuracy for investment decisions
- Business Transformation Documentation: Created comprehensive documentation of business logic transformations, calculation methodologies, and data derivations supporting regulatory compliance (Mi-FID II, FINRA) and audit requirements
- Logical Data Modeling: Designed enterprise logical data models integrating market data, reference data, and transactional data with dimensional modeling for investment analytics and regulatory reporting
- Teaching & Knowledge Sharing: Mentored and trained 12+ data engineers and analysts on advanced Cypher query optimization, graph traversal algorithms (PageRank, shortest path, community detection), GraphRAG concepts, and RAG architecture patterns through workshops, documentation, and hands-on sessions
- AWS Infrastructure: Architected cloud-native data platform on AWS using S3 for data lake, Lambda for serverless processing, Glue for ETL, EMR for big data workloads, and SageMaker for ML model deployment with multi-region failover

Skills: Java, Python, Neo4j, Knowledge Graphs,LangChain, Multi-Agent Systems, NLP, AWS, Python, Machine Learning, RAG, GraphRAG • OpenShift

S&P Global – New York, NY **Dec 2022 – May 2023**
Senior Data Architect & Knowledge Graph Engineer

Responsibilities:

- Manufacturing Knowledge Graph: Designed Neo4j-based manufacturing intelligence platform with

comprehensive RDF/OWL ontology mapping 1M+ technical documents, component specifications, and supplier relationships
- GraphRAG Pipeline: Built production GraphRAG system using BioBERT and LangChain for automated component relationship extraction, transforming unstructured manufacturing documentation into queryable property graph achieving 85% accuracy
- Snowflake Integration: Architected Snowflake data warehouse integrated with Neo4j through Airflow-orchestrated ETL pipelines processing 50+ external vendor data sources with automated schema drift detection
- **Multi-Dimensional Ontology**: Developed multi-dimensional component ontology enabling flexible categorization by function, material, supplier, lifecycle stage, and quality metrics supporting supply chain optimization
- Data Catalog Implementation: Deployed Amundsen data catalog with automated metadata discovery, lineage tracking, and Great Expectations quality validation ensuring 99.9% data quality across structured and graph datasets
- Real-Time Ingestion: Implemented Kafka-based real-time data ingestion pipelines with Airflow orchestration for streaming manufacturing telemetry and quality control data with sub-second latency
- AWS Data Platform: Designed scalable AWS infrastructure (S3, EMR, Lambda, Glue) supporting graph analytics with PyTorch Geometric for GNN training with multi-region deployment and disaster
- recovery
- Teaching & Mentoring: Conducted training sessions on Cypher optimization, graph data modeling, and GraphRAG architecture for 20+ engineers and data scientists

Johnson & Johnson – USA **Dec 2019 – Dec 2022**
Data Architect & Semantic Knowledge Graph Lead

Responsibilities:
- Pharmaceutical Knowledge Graph: Architected production pharmaceutical knowledge graph using Neo4j and MarkLogic semantic database with comprehensive drug discovery ontology (500K+ entities) serving 500+ researchers
- Healthcare Ontology Integration: Engineered healthcare data architecture using RDF, OWL, TTL formats integrating SNOMED CT, RxNorm, ICD-10/11, LOINC, and FHIR standards for interoperable clinical intelligence and regulatory compliance
- GraphRAG for Clinical Data: Designed GraphRAG pipeline integrating BioBERT, SciBERT, and PubMedBERT with LangChain for automated entity extraction from clinical documents, transforming 2M+ unstructured research papers into queryable property graphs
- MarkLogic Semantic Platform: Built MarkLogic-based semantic data platform with RDF triple stores, SPARQL endpoints, and property graph models enabling federated queries and advanced knowledge discovery
- Snowflake Data Harmonization: Architected Snowflake-based data harmonization platform with Airflow orchestration processing 100+ clinical studies annually with comprehensive data lineage, temporal versioning, and audit trails
- Adaptive Ontology Evolution: Implemented AI-driven drug safety ontology with adaptive schema evolution accommodating new therapeutic areas using machine learning-based pattern recognition and automated taxonomy refinement
- Temporal Versioning: Developed comprehensive temporal versioning system for clinical trial historical context enabling time-travel queries, protocol version tracking, and regulatory audit trail support
- Role-Based Access Control: Architected enterprise RBAC with property-level permissions and ontology-driven security policies ensuring FDA, EMA, and GxP compliance for global research teams
- Data Quality & Lineage: Established end-to-end data quality framework with automated validation, anomaly detection, and comprehensive lineage tracking supporting regulatory submissions and clinical decision-making
- Logical Data Modeling: Designed logical data models for pharmaceutical research integrating clinical trial data, drug safety information, regulatory filings, and molecular databases
- Teaching Graph Concepts: Developed and delivered training curriculum on Cypher, SPARQL, graph traversal algorithms, GraphRAG architecture, and semantic reasoning to 30+ pharmaceutical data scientists and engineers
- Skills: Neo4j, Knowledge Graphs, Semantic Search, Ontology, Spacy, Healthcare AI, FHIR, BioBERT, NLP, GCP, Python, Machine Learning, GraphRAG

Scania – Sweden **Jun 2018 – Nov 2019**
Data Architect & Graph Analytics Engineer

Responsibilities:
- Built enterprise Neo4j knowledge graph with multi-department data integration using ontology-driven RDF transformation for 50,000+ connected vehicles
- Designed predictive maintenance ontology in OWL with temporal versioning enabling real-time parts recommendation with 90% accuracy through sophisticated graph traversal algorithms
- Implemented real-time data ingestion pipelines via Docker/Kubernetes with Kafka streaming integration processing telematics data with automated quality validation
- Architected AWS-based data platform using S3, EMR, and Lambda for graph analytics model deployment achieving sub-second inference latency
- Developed comprehensive data catalog with metadata management and lineage tracking for automotive datasets
- Mentored engineers on Cypher query optimization, graph data modeling, and real-time pipeline architecture

AT&T – India **Mar 2014 – May 2018**
Data Architect & Knowledge Graph Engineer

Responsibilities:
- Architected ontology-driven knowledge graphs for enterprise content analytics using Neo4j, MarkLogic semantic database, and RDF/OWL technologies serving 50,000+ employees
- Built semantic data platform using MarkLogic with RDF triple stores, SPARQL endpoints, and ontology management supporting telecommunications network intelligence
- Implemented Airflow-orchestrated ETL pipelines on AWS processing 100,000+ daily transactions with automated data quality monitoring and catalog registration
- Designed comprehensive data catalog with automated metadata management, lineage tracking, and business glossary for enterprise knowledge assets
- Developed role-based access control framework with fine-grained permissions ensuring secure data visibility across multi-tenant environment
- Created logical data models integrating structured, semi-structured, and unstructured data sources for telecommunications analytics
- Conducted training workshops on Cypher, graph algorithms, and semantic reasoning for data engineering teams

Multiple Healthcare & Government Projects – India **Aug 2009 – Mar 2014**
Semantic Data Architect & Healthcare Ontology Engineer

Responsibilities:
- Architected healthcare knowledge systems using RDF, OWL, TTL formats with MarkLogic semantic database for clinical decision support and medical knowledge management
- Developed comprehensive healthcare ontologies integrating SNOMED CT, ICD coding systems, RxNorm, and LOINC for hospital information systems
- Pioneered ”Road Safety Ontology (SafeOn)” using OWL formal semantics and Prot´eg´e establishing transportation safety knowledge framework adopted by government agencies
- Created innovative ontology merging algorithms for healthcare domain integration enabling cross-institutional knowledge sharing, published in IEEE with 50+ citations
- Implemented SPARQL query optimization and semantic search engines for medical knowledge discovery supporting clinical guideline compliance
- Designed data quality frameworks and validation rules for healthcare data ensuring regulatory compliance and patient safety