

Data -> KG -> Serving - End-to-End Flow

No-NAT VPC | Cognito JWT | Neptune (SPARQL) | OpenSearch

YOUR LOGO

Data -> KG -> Serving - End-to-End Flow (No-NAT VPC, Cognito JWT, Neptune, OpenSearch)

YOUR LOGO

DATA

[0] CQs

- Define must-answer questions (e.g., trims with ACC & sales by region/date)
- Drive ontology and mappings

[1] Sources -> Bronze (S3/DDFS)

- 8 inputs: Sales DB, R&D PDFs, OCR PDFs, CAD/Images, XML/JSON, Reviews, Model/Parts, Patents
- Tools: Airbyte/NiFi/Glue, Texttract/Tesseract, Tika, Scrapers

[2] Ingest & Clean (Silver)

- Parse CSV/JSON/XML; extract text from PDFs
- QA (nulls/types/dates) with Great Expectations
- Entity cues: model codes, trims, parts

[3] Conform (Gold)

- Dims/Facts + bridges (models, trims, features, parts, sales, regions, docs/ings/patents)
- Tools: Spark/SQL and dbt models & tests

[4] Ontology & Vocabulary

- OWL/RDFS classes and properties
- SKOS labels and synonyms (ACC <-> Adaptive Cruise Control)
- Map CQs to required relations

[5] RDF Mapping

- R2RML/RML or Python to mint URIs
- Serialize N-Triples or Turtle

[6] Validation & Reasoning

- SHACL shapes (Trim belongsTo 1..1; modelYear >Year)
- Optional OWL reasoner (Hermit/ELK)

[7] Load Graph DB (Neptune)

- Bulk loader from S3 (N-Triples)
- Expose SPARQL endpoint; set prefixes

[8A] OpenSearch (BM25)

- Index doc text (R&D/OCR/manual)
- Store docId to entity links

[8B] Vector Index (optional)

- Embeddings to KNN (OpenSearch/qdrant)
- Chunk to entity metadata for RAG

[9] Query Orchestrator (API GW + Lambda)

- JWT auth (Cognito)
- Router: CQs -> SPARQL; open-ended -> vector/BM25
- Federation (SPARQL SERVICE) for external brands/regions

Neptune (SPARQL)

- Facts: trims, features, parts, sales aggregates
- SigV4 if IAM auth enabled; VPC-only :8182

OpenSearch (Snippets)

- BM25 snippet for evidence
- SigV4; VPC domain

[10] Final Answer (UI)

- Structured facts and evidence citations
- optional external links (Wikidata)

Runtime ping-pong (JWT-protected):

