

NEXON SAMUEL

nexon.samuel@gmail.com | (813) 465-4216 | Suffern, NY | [LinkedIn](#)

SUMMARY

Results-driven **Data Engineer** with 7+ years of experience in **ETL development, cloud platforms (AWS, Databricks), Python, SQL, and workflow automation**. Skilled in **data modeling, governance, and dashboarding (Power BI/Tableau)**, delivering **scalable pipelines and actionable insights**. Adept at **translating business requirements into data-driven solutions** and driving adoption of analytics platforms.

WORK EXPERIENCE

AMGEN

Feb 2022 - Dec 2025

Platform Data Engineer | Mar 2024 - Dec 2025

Suffern, NY

- Owned **Prophecy platform** roadmap and adoption for 12+ domain teams, enabling **self-service ETL development** through templated components called **Gems**, documentation, and training; shifted 60+ pipelines from centralized to domain-managed ownership, reducing bottlenecks and accelerating delivery cycles.
- Managed end-to-end **AWS** deployment of Prophecy platform (**EC2, Docker, ECR**), automating version upgrades, container builds, and health monitoring; enabled zero-incident deployments across 8 quarterly releases.
- Led **Databricks Unity Catalog** implementation and POCs for Delta Sharing, Workflow Scheduling, and ML model serving; configured **row-level security** and dynamic data masking for 8+ teams, reducing ad-hoc access requests through standardized shared datasets and permission models.
- Built **event-driven workflows** using Python (Flask) and **AWS Lambda**, orchestrated via SQS, SNS, and EventBridge; replaced cron-based scheduling with event triggers for 50+ pipelines, decoupling dependencies and enabling asynchronous processing at scale.
- Designed secure **AWS storage architectures** (S3, EBS, EFS) with **KMS encryption**, automated backups, and S3 Access Points for 10+ TB of regulated data; implemented lifecycle policies and access controls aligned with corporate security standards.
- Developed **Power BI dashboards** and automated alerting for Databricks cluster costs, usage trends, and job performance across 12 teams using **Spark SQL**; delivered email notifications to application leads for cost spikes and SLA breaches, enabling rapid response and informing quarterly capacity planning.

Semantics Data Engineer | Feb 2022 - Feb 2024

Tampa, FL

- Spearheaded biotech oncology **Healthcare Provider semantic domain**, owning end-to-end delivery from sourcing through production support; built 15+ pipelines processing **5M+** provider records, enabling cross-functional teams to query unified provider data via semantic models and reducing duplicate data requests.
- Implemented **MarkLogic** and **Metaphactory** integration enabling federated semantic queries and SPARQL exploration for 30+ analytics and business users; built reusable semantic templates reducing query development time and enabling self-service data discovery across healthcare domains.
- Built shared automation utilities and semantic pipelines across **S3, Databricks, graph ingestion, and Airflow**; partnered with ontology modelers to create **reusable** ingestion frameworks, reducing new pipeline development time from ~3 weeks to ~1 week and improving cross-domain data consistency.
- Migrated semantic datasets from **AWS Glue** to **Databricks Unity Catalog**, improving governance, asset tracking, and lineage visibility; streamlined **Airflow** orchestration workflows to leverage Unity Catalog's native scheduling and metadata management.
- Mentored junior engineers on **Agile/Scrum** practices, standardizing **Git** branching, **CI/CD** pipelines, and **YAML** deployment patterns; established team coding standards for semantic pipeline development.

TERADATA

May 2021 - Dec 2021

Cloud Automation Intern

San Diego, CA

- Developed Teradata **automation scripts** to streamline database control settings and enable abort diagnostics, reducing manual configuration steps for **production database operations** and supporting troubleshooting workflows.
- Enhanced **Groovy** and **Symphony workflows** to support serial, parallel, and hybrid workload execution patterns; maintained **BTEQ** scripts and **SQL procedures** for batch processing and reporting pipelines.
- Implemented **LDAP**-based authentication for **Teradata DSS** via **REST APIs** using Python (**ldap3**, **tdtestpy**, **BeautifulSoup**), enabling secure customer access and supporting enterprise authentication integration.

Data Analyst/ETL System Administrator

Mumbai, MH, India

- Partnered with banking operations and risk teams to **translate business requirements** into analytics solutions, supporting operational performance monitoring and compliance reporting.
- Built **Python**-based analytics workflows (**Pandas**, **NumPy**, **Scikit-learn**) analyzing transaction data and operational metrics to identify performance trends, data quality issues, and anomaly patterns supporting process optimization.
- Designed **PL/SQL data models** and **procedures** for root-cause analysis of operational failures; delivered **Tableau dashboards** visualizing **KPIs** and operational trends for business and analytics teams.
- Transitioned to ETL system administration, **automating job monitoring**, **scheduling**, and **failure recovery** on **Linux**-based infrastructure; maintained stable production workloads for hundreds of daily batch processes.
- Supported migration of legacy ETL applications to cloud environments, contributing to workload assessment, testing, and deployment while ensuring continuity of critical data pipelines.

TECHNICAL SKILLS

Programming & Scripting: Python, PySpark, SQL, Shell/Bash, PL/SQL

Data Platforms & Analytics: Databricks, Prophecy, Metaphactory, Tableau, Power BI, Airflow, Workflows

Database Technologies: Teradata, MySQL, PostgreSQL, SQLite, Snowflake, Unity Catalog, MarkLogic

Cloud Computing & AWS Services: AWS EC2, S3 (Access Points, Event Notifications, Glacier), Lambda, Glue, Athena, EMR, EKS, ECS, RDS, IAM, CloudWatch, AWS Backup

Containerization & Orchestration: Kubernetes (Namespaces, Services, Deployments, ConfigMaps, Secrets, ReplicaSets, DaemonSets, Volumes), Docker, AWS ECR, DevOps, Health Monitoring

Data Quality & Governance: Data Quality Checks, Data Validation, Data Lineage, Data Profiling, Governance Frameworks, Access Controls, Row-Level Security, Dynamic Data Masking, Quality Improvement

Web & API Frameworks: REST APIs, Flask, FastAPI, Streamlit, Postman, HTTP Protocols, Uvicorn

Generative AI & ML Concepts: LLMs, OpenAI GPT, LangChain, FAISS, Embeddings, Vector Databases, Prompt Engineering, Retrieval-Augmented Generation (RAG), AI Agents

EDUCATION

University of South Florida — Master of Science in **Business Analytics & Information Systems** Jan 2020 - Dec 2021

University of Mumbai — Bachelor's in **Information Technology** Jun 2013 - May 2016

University of Mumbai — Associate in **Computer Engineering** Jun 2010 - May 2013

INTERESTS

TopGolf, bowling, K-Speed (mini F1 racing), PS5/arcade gaming, road trips, and exploring new places.