Pushkar Pramod Wani

Willing to Relocate | (607) 297 9920 | pwani@binghamton.edu | Linkedin | Github | Portfolio

SUMMARY

Analytics and Data Engineer with 4 plus years building data pipelines, star schemas, and analytics serving layers using SQL, Python, Spark, and Snowflake. Delivered governed self service analytics through Semantic Modeling and validated Text to SQL workflows. Applied ML experience shipping forecasting, anomaly detection, and smart narratives that enable faster decision making for business stakeholders.

SKILLS

- **Languages**: Python, SQL, TypeScript, Java
- **Data**: Snowflake, Snowpark, dbt, Airflow, PySpark, Databricks
- **Datastores**: PostgreSQL, Redshift, DynamoDB, MongoDB, Redis
- **Practices**: Data Modeling, SCD Type 2, Schema Design, Data Contracts, Unit Tests, Validation, Deduplication, Lineage Logging, Monitoring
- **Analytics**: PowerBI, Streamlit, Semantic Modeling, Reporting Views, KPI Definitions, Tableau
- **Cloud and DevOps**: AWS, Azure, GCP, Docker, Kubernetes, Linux, Git
- **Certifications**: Microsoft Azure AI Fundamental, Microsoft Azure Data Fundamentals, Building and Evaluating Data Agents in Snowflake

EXPERIENCE

Regeneron Pharmaceuticals
May 2025 - Dec 2025

Data & ML Engineering Intern *Sleepy Hollow, NY*

- Built Snowflake ingestion pipelines from S3 (Parquet), Redshift, and SFTP, enforcing schema alignment, source to target mappings, and reject handling to publish analytics ready tables for commercial data marts
- Implemented automated Data Quality and Observability with Snowpark plus Streamlit, including schema drift detection, null and distribution checks, and KPI anomaly flags via scheduled runs, cutting downstream query latency by about 70%
- Designed a governed Semantic Layer using YAML based Semantic Modeling with entities, relationships, metrics, and join rules, and integrated Cortex Analyst with Cortex Search in Snowflake Intelligence to deliver traceable Text to SQL, cohort analysis, and KPI diagnostics
- Implemented Snowpark stored procedures for statistical profiling, Holt Winters forecasting, anomaly detection, and decision tree based top insights, integrating outputs into Snowflake Intelligence to support alerts and planning workflows

Binghamton University, Watson Career Services & School of Computing
Aug 2024 - Dec 2025

Graduate Software Developer *Binghamton, NY*

- Developed a micro-internship platform using Next.js, React, and PostgreSQL that streamlined job posting, student application, and employer review workflows, resulting in faster matching of candidates to opportunities
- Designed a relational schema in PostgreSQL and built SQL reporting views for applicant tracking system matching, including indexing and query tuning to improve admin analytics performance
- Built Python REST APIs with authentication, RBAC, and request validation, and set up automated deployments via GitHub Actions and Docker, which reduced deployment time and ensured consistent environment provisioning

KrtrimaIQ Cognitive Solutions
Dec 2021 - Dec 2023

Data Engineer *Bangalore, India*

- Designed Azure Data Factory ELT pipelines with SCD Type 2, validation, deduplication, and lineage logging, publishing dimension and fact tables for dashboards used by 5+ teams
- Built Databricks PySpark batch jobs for ingestion and transformations, adding schema and null checks plus alerting to prevent bad data from reaching reporting tables
- Standardized dbt SQL models and partnered with stakeholders to define KPIs and star schemas, publishing curated datasets for PowerBI dashboards
- Delivered BI focused assistants using Rasa rule based NLP to generate smart narratives from tabular KPIs, and published curated datasets for PowerBI dashboards. Reduced reliance on manual SQL and improved reporting adoption across business teams.

PROJECTS

Clinical Trial ETL Pipeline | Clinical Trial ETL Pipeline

- Built an asynchronous ingestion pipeline with Redis queue orchestration, worker services, status polling APIs, PySpark transforms, and loading into curated PostgreSQL tables
- Designed a star schema and measurement view serving layer, adding schema validation, reject handling, idempotent reruns, and structured error reporting for reliable processing

Airflow Orchestrated Micro Batch Pipeline | Airflow Orchestrated Micro Batch Pipeline

- Orchestrated a scheduled micro batch pipeline with Airflow, triggering AWS Glue PySpark transforms from S3 raw events to curated Parquet datasets on S3
- Aggregated KPIs and upserted into DynamoDB for near real time dashboards and Lambda APIs, with local end to end testing using LocalStack and Docker

Hugging Face Seq2Seq NLP Pipeline | Hugging Face Transformer NLP Pipeline

- Fine tuned and evaluated seq2seq transformer workflows for translation and summarization, packaging preprocessing, inference, and task specific quality reporting into a reproducible pipeline.

EDUCATION

Binghamton University, SUNY
Jan 2024 - Dec 2025

MS, Computer Science *Binghamton, New York*

- **GPA:** 3.8
- **Coursework:** Distributed Systems, Cloud Computing, Data Mining, Machine Learning, Artificial Intelligence, Data Structures

PES Institute of Technology
Aug 2017 - Aug 2021

BE, Electronics and Communication *Bangalore, India*

- **GPA:** 3.9