VIGNESH NATARAJAN

705-978-5294 | vikky33.nv@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#) | [Credly](#)

TECHNICAL SKILLS

Programming Languages: Python, C#, Java
Libraries & Frameworks: Pandas, NumPy, FastAPI, Flask, PySpark, Streamlit
Databases: PostgreSQL, MS SQL Server, MongoDB
ETL tools: Airflow, dbt
Cloud & DevOps: AWS(S3, Lambda, Glue, Athena), Microsoft Azure DevOps, Docker, Git, GitHub
Others: DBeaver, SSMS, Jupyter Notebooks, Selenium, JIRA, ServiceNow

PROFESSIONAL EXPERIENCE

Data Engineer Intern   May 2025 – Aug 2025
University of Windsor, ON

- Developed and automated ETL pipelines to extract data from external APIs, processing 1,000+ records per batch for analytics
- Transformed raw datasets into structured DataFrames, loading 800+ cleaned records for real-time dashboard consumption
- Created interactive dashboards to visualize operational metrics, enabling 20+ team members to monitor trends
- Optimized data pipeline performance by refactoring transformation logic, reducing processing time by 30%

Software Engineer (Data)   Jul 2020 – Apr 2023
Cognizant - Chennai, India

- Implemented AWS Glue jobs to automate ETL processes, transforming and cataloging 500+ datasets from S3
- Designed and optimized SQL stored procedures to enhance data retrieval time for ETL pipelines upto 40%
- Designed data models using dbt, creating 20+ transformation models that improved data lineage tracking by 25%
- Cleaned and standardized master data across 15+ tables, improving reporting accuracy by 30%
- Automated parsing of 100MB+ XML data dumps, enabling efficient ingestion of 1,000+ files into PostgreSQL for pipeline integration

Intern   Feb 2020 – Jun 2020
Cognizant - Chennai, India

- Utilized PySpark to process 300MB+ datasets, assisting in initial data transformation for ETL pipelines
- Authored SQL queries to extract and validate 1,000+ records to support data quality checks for reporting
- Assisted in configuring a basic data ingestion workflow using Python scripts, handling 500+ CSV files
- Collaborated on monitoring a batch processing job, reducing data load errors by 10% through log analysis

EDUCATION

Master of Applied Computing   May 2024 – Aug 2025
University of Windsor — Windsor, ON

Bachelor of Technology in Computer Science and Engineering   Jul 2016 – Jun 2020
SRM Institute of Science and Technology — Chennai, India

PERSONAL PROJECTS

[CryptoXtract](#) | *Real-time ETL pipeline powered by modern data stack for cryptocurrency market intelligence*

- Built production-ready ETL pipeline using Python, PostgreSQL, and Docker processing cryptocurrency market data with automated extraction from external API every 6 hours
- Implemented comprehensive data transformation pipeline using dbt for creating analytics-ready datasets from raw API data
- Designed containerized data warehouse with proper schema optimization, handling complex financial data types & error handling

[FilterIt](#) | *Python based web app powered by AI and ML to prevent data leakage*

- Utilized SpaCy to detect and redact sensitive information from patient records, preventing data leakage before sending it to LLM for summarization
- Integrated OpenAI API to generate detailed patient record summaries from various formats, enhancing data accessibility
- Integrated two-factor authentication (2FA) using a Google Authenticator-style approach, enhancing login security and eliminating the need for traditional passwords

CERTIFICATIONS

- [AWS Certified Data Engineer – Associate](#)
- [Microsoft Certified: Fabric Data Engineer Associate](#)