Bobby Atwal

(408) 466 – 3330 | bobbyatwal85@gmail.com | www.linkedin.com/in/bobbyatwal

Objective

Software Engineer with 5+ years of experience building data-intensive applications, scalable platforms, and AI-driven solutions across enterprise, healthcare, and regulated environments. Skilled in data engineering, backend systems, cloud architecture, and GenAI integrations. Known for bridging complex data ecosystems with intuitive automation to accelerate efficiency, interoperability, and innovation. Recognized for leading cross-functional collaborations and mentoring engineers to deliver scalable, high-impact solutions.

Education

B.S. Computer Science, Graduation: December 2019
San Jose State University, San Jose, CA | GPA 3.6

A.A. Mathematics, Graduation: December 2017
Evergreen Valley College, San Jose, CA | GPA 3.8

Technical Skills

Programming Languages: Python, SQL, Java, JavaScript (ES6+), TypeScript (Node.js), R, HTML, CSS

Data Science: Pandas, NumPy, PySpark, Scikit-learn, MLflow, dbt, Tableau, Power BI, Bash, Git

Cloud: Databricks, AWS (S3, EC2, RDS, Lambda, CloudWatch, Glue, Athena), GCP, Docker, Kubernetes

Databases: Snowflake, PostgreSQL, MySQL, AWS Aurora, AWS DynamoDB, MongoDB, OracleDB

Generative AI: RAG Pipelines (GPT-3.5, LangChain, Azure OpenAI), Vectorization (Qdrant), Prompt Engineering

Frameworks & Web Development: Django, React, Angular, Node.js, FastAPI, Postman, D3, Selenium

Software Systems: Domino Data Lab, Jira, Confluence, TrackWise, E1, Salesforce, VMware, LIMS, EHR

Professional Experience

Senior Data Engineer, Clinical Ink **02/2026 - Present**

- Investigated complex data integrity issues across Clinical Ink's eCOA and EHR-integrated clinical data platform by developing advanced SQL & Python-based workflows against Aurora PostgreSQL, DynamoDB, and Snowflake to detect anomalies, trace root causes, and support large-scale data remediation for active studies
- Automated Python-based data integrity and validation pipelines, eliminating manual workflows to accelerate issue resolution, improve auditability, and streamline GxP-regulated production data operations
- Developed event-driven AWS workflows (Lambda, CloudWatch, RDS, DynamoDB) to automate impact assessments, data monitoring, and clinical data reporting across distributed healthcare systems
- Collaborated with Engineering, Product, and Data Management teams to support clinical data governance, database maintenance, and data reliability initiatives driving high-impact sponsor and stakeholder deliverables

Data Engineer → Senior Data Engineer, Digital Health Automation **01/2023 - 02/2026**
Johnson & Johnson Innovative Medicine - South San Francisco, CA

- Built and maintained scalable data pipelines for wearable sensors (Fitbit, Oura, ActiGraph, Corsano, etc.) transforming multimodal clinical study data into analysis-ready datasets supporting oncology, cardiovascular, neuroscience, immunology, and autoimmune biomarker research and drug development
- Architected cloud-based ETL/ELT workflows and scalable data models to standardize multimodal biosensor, wearable, and clinical research data (HR/HRV, PPG, ECG, accelerometer, gyroscope, telemetry), enabling AI-ready datasets and downstream analytics for scientific and ML teams within a GxP-compliant ecosystem
- Led cross-functional collaborations with scientists, R&D engineers, QA teams, and clinical stakeholders to expand an automated data platform supporting FDA-regulated research workflows, modernize legacy

pipelines, and establish scalable governance, lineage, and validation standards across distributed data systems
- Designed an internal RAG chatbot using Qdrant, Azure OpenAI, LangChain, React, and GPT-3.5 to provide semantic search, metadata exploration, and AI-assisted knowledge retrieval for scientific stakeholders
- Engineered high-volume data processing pipelines leveraging advanced transformation logic for time-series biosensor data to improve scalability, accelerate scientific data delivery, and reduce infrastructure costs
- Developed live visualization dashboards (Tableau & Power BI) to monitor device performance, data quality, and infrastructure costs, driving visibility into platform utilization and performance for R&D leadership
- Mentored junior engineers on ETL/ELT design, testing automation, clinical data validation, code reviews, and debugging best practices within regulated digital health environments
- Technologies: Python, PySpark, NumPy, SQL, AWS (S3, Lambda, EC2, Glue, Athena), Qdrant, Docker, Domino

Software Engineer II, Software Support & Training **11/2021 - 11/2022**

Thermo Fisher Scientific - South San Francisco, CA

- Built a Django-based full stack application that replaced a legacy complaint-tracking system - automating reporting, enabling audit trails, and improving reliability for FDA-regulated software products
- Supported an automated high-throughput medical device platform (8,000+ samples/day across 30+ instruments), ensuring accurate data capture, system uptime, and compliance with FDA-regulated device standards during large-scale Covid-19 diagnostic operations
  - ➢ Worked directly with clinical labs and hospitals worldwide to ensure reliable diagnostic data delivery and device performance during high-volume operations, contributing to $1B+ product revenue in 2022 & $10B+ annual division revenue
- Collaborated with R&D and field support to optimize data reliability across integrated instruments, improving data traceability for SampleManager LIMS & hospital EHR systems (e.g. Epic) in downstream clinical workflows
- Performed QC on qPCR assay data - analyzing amplification curves, Ct values, and controls to validate diagnostic accuracy and communicate findings to leadership for clinical decision making
- Diagnosed global system-level failures across device components with R&D and QA teams, ensuring root-cause resolution and proper documentation within FDA-compliant workflows using Trackwise & Jira
- Identified and prioritized recurring system issues from field support to engineering teams, contributing to design improvements and software scalability across future releases
- Led week-long technical trainings for 30+ global employees across management, field support, and clinical science teams, covering diagnostic software and system design to improve system understanding, troubleshooting ability, and operational efficiency in clinical lab environments

Systems Application Programmer **01/2021 - 04/2021**

City of San Jose, San Jose, CA

- Built a full-stack web application using Angular, TypeScript, Node.js, OracleDB, and Docker to deliver real-time budget analytics and forecasting for the Finance Department and Project Management teams overseeing the $1B+ Capital Project Management System (CPMS) portfolio
- Developed RESTful APIs and dynamic UI components to display project expense data with clean, responsive UX using Bootstrap, HTML, JSON, and CSS/LESS
- Managed OracleDB using Toad and SQL/PLSQL; created views and optimized queries for financial reporting
- Supported Agile development lifecycle using Smartsheet for sprint planning and Atlassian tools (Jira, Confluence, Bitbucket) for issue tracking and documentation

Student Employee, Nuclear Quality Assurance (Internship) **07/2015 – 01/2017**

Electric Power Research Institute, Palo Alto, CA

- Developed Microsoft Access tools to streamline QA process tracking and nuclear supplier oversight workflows
- Delivered trend analysis and process improvement insights to QA leadership and training coordinators

U.S. Citizen