Babajide Ogunjobi

Phone No: (850)339-9983
Email Address: jogunjobi@gmail.com

Summary:
Meticulous Data Engineer and Machine Learning Engineer with a Business/Functional background with immense experience building large-scale distributed data platforms, and data warehouses and deploying top-level production ML models

Skills:
- Python
- SQL
- JavaScript / NodeJS
- Scala
- Amazon Web Services (AWS)
- Google Cloud (GCP)
- Docker
- Apache Spark
- Snowflake
- PyTorch
- Sagemaker
- Apache Airflow
- Neo4j
- BigQuery
- DBT
- Microsoft Azure

Certifications & Courses:
- AWS Certified Machine Learning – Specialty
- Neo4j Certified Professional
- **Instructor:** Data Warehouse Optimization in Snowflake (Link)

Education:
- M.S. Information Systems (A.I. Specialization), Northwestern University (Part Time)   **08/2020 – Present**
- M.S. Business Management, University of Florida   **01/2008 – 04/2010**
- B.S. Management, Florida State University   **08/2004 – 12/2007**

Professional Experience:

Tracc Systems, Chicago, IL   **06/2022 – Present**
Chief Technology Officer
- Leading the design, architecture, and development of the full Tracc platform for last mile, freight management, and freight visibility including the front-end backend and infrastructure development
- Lead Data Architect and Data Engineer designing and building out the Tracc data infrastructure which included a data warehouse in DuckDB, graph database in Neo4j, data lake in S3, transactional database in Postgres as well as API layers
- Development of the real-time delivery and order prioritization module using a graph-based traveling salesman algorithm
- Building out a scalable pub-sub data platform to receive real-time data from multiple data providers and process them to multiple downstream systems
- Developing Large Language Models (LLM) using OpenAI, Langchain, and Neo4j to read our internal data and preemptively answer questions on the status of logistics and supply chain processes that we're managing

AlphaSense Inc., New York, NY   **06/2021 – 06/2022**
Director of Data Engineering & Architecture
- Developing and executing an extensive migration plan which featured:
  - Totally redesigning the architecture and data flow processes to allow for parallel and distributed source data migration, data catalog and data lineage tracking, interdependent DAGs, and external APIs
  - Replacing subscription-based data integration tools (FiveTran) with custom-built data pipelines using Apache Airflow and custom Python & SQL scripts
  - Reconfiguring our data storage layer and data warehouse from an append and insert-update logic to a fully versioned data lake strategy

- Mentoring a distributed team of 8 analytics engineers, data engineers, and analysts on data processing practices and agile development methodologies.
- Deployed multiple natural language models for document search and retrieval along with a pipeline using AWS Sagemaker, S3, and Apache Airflow

MOOVE, Remote   **02/2022 – 04/2022**
Chief Technology Officer
- Built out end-to-end technology infrastructure (internal and external) on the way to a $105 million Series A raise. Major accomplishments include:
    - Leading architecture, roadmap, and infrastructure planning, and development of multiple applications to support operations
    - Building out the technology infrastructure from scratch (including data platforms, pipelines, and algorithms) while designing and deploying Information Security policies
    - Hiring, building out, and managing internal development teams as well as working with external vendors to accelerate the development of applications and infrastructure
- Designing and developing scalable fault-tolerant enterprise big data platforms for data ingestion, processing, and storage
- Worked on real-time machine learning and optimization proof of concepts that powered driver onboarding, recruitment, and revenue forecasts.

Publicis Sapient, Chicago, IL   **01/2019 – 06/2020**
Manager Data Engineering
- Working with clients to create architecture documents for current and future platforms, strategic framework documents, and RFPs.
- Leading project teams to execute high-level data engineering projects that have been won by the organization:
    - Developing a graph database solution in Azure for a major tax company to unify customer data from multiple internal databases.
    - Developed a distributed enterprise graph solution in order to build out an automated "knowledge as a solution" platform within a major bank.
    - Worked within a major vehicle manufacturing organization to create an end-to-end data pipeline to analyze and visualize vehicle sales and promotions data.

Quartz, New York, NY   **04/2018 – 01/2019**
Lead Data Engineer Consultant
- Lead Engineer in charge of designing and developing the new Quartz Data Infrastructure from scratch. This involved working with multiple development teams and the Product Management teams to build a solution, which encompassed each team's needs.
- Setting up a new data lake and data stores to make all of our article and advertising-related data queryable (using AWS Redshift & Athena) and searchable (using Elasticsearch).
- Automating multiple Python jobs and daily reports for the Analytics team using AWS Lambda and by creating a dedicated Apache Airflow instance.
- Worked closely with our data scientist to build a dedicated ephemeral cluster to facilitate training, testing, and deploying production machine learning (ML) models

Grubhub, Chicago, IL   **03/2016 – 03/2018**
Senior Data Engineer
- Lead engineer coordinating the migration of multiple MySQL and Redshift databases into our new Hadoop, Hive, and AWS S3-based data platform.
- Developing data integration and transformation PoCs, using Python and Pyspark scripts needed to build our downstream reporting data store.
- Building coordinated and scheduled pipelines using Apache Airflow to manage multiple ETL jobs.
- Working with data analysts and data scientists to build Python helper modules for custom data analysis, transformations, and integrations.
- Worked alongside team data scientists to build APIs that would read data from our database for their models and leading deployment of the models
- Designing and building our cluster management schedule and script to reduce our AWS EMR footprint and costs while still taking advantage of available memory to process all required jobs.

J.P. Morgan Chase & Co.   **07/2014 – 02/2016**
Data Solutions Architect
- Reserve engineering and recreating logical and physical models to adhere models to Chase Consumer Bank standards.
- Forward engineering of new and remediated physical data models for instantiation.
- Representing the application development team in all Data Review Board meetings to present specifications, changes, and updates to our data models
- Creating outbound metadata documents and processes for exporting data to downstream data warehouses.
- Maintaining metadata repository, data catalogs, and keeping track of data lineage.

Citigroup   **10/2011 – 07/2014**
Solutions Lead Analyst – ETL & Metadata (Vice President)
- Lead Data Processing and Data Management Subject Matter Expert for Citigroup's Global Finance Backbone Initiative
- Defining Business & Functional requirements, and user interface wireframes for the Finance Backbone Master Data Management (MDM) solution.
- Designing solutions, test scenarios, and creating prototype applications to test the data integrity of Fact data, Reference data and Extract Transform Load (ETL) processes
- Profiling and analyzing financial data from all source systems for system compliance, data integrity, and target record conformity.
- Designing security and data entitlements model for all users across the enterprise system (System permissions, User Profiles, Page(s) security, etc.)
- Running mass reconciliations on all Master Data Management (MDM) setups as well as supporting databases after deployments to capture breaks and missing setups.