**Gabriel Olokunwolu**

**Phone: 516-580-5050|Email:** [**golokunwolu@gmail.com**](mailto:golokunwolu@gmail.com)

---------------------------------------------------------------------------------------------------------------------------------------------------------------

Certified Big Data/ Machine Learning Engineer with 7+ years of IT Experience comprising of Statistical Modelling & Machine learning algorithm developments, AWS Big Data Engineering, Application development; having acquired extensive experience in designing, deploying and operating highly available, scalable and fault tolerant systems using big data technologies

---------------------------------------------------------------------------------------------------------------------------------------------------------------

**EDUCATION AND PROFESSIONAL CERTIFICATIONS**

* AWS Certified Data Analytics Specialty
* AWS Certified Machine Learning Specialty
* AWS Certified Developer Associate (AWSCDA)
* AWS Certified Solutions Architect Associate (AWSCSAA).
* Certified Scrum Master

---------------------------------------------------------------------------------------------------------------------------------------------------------------

**EXPERTISE**

* Demonstrable knowledge & experience in various Big Data or distributed systems. Such as Hadoop, NoSQL, AI/ML, ETL and Streaming
* Developed Cloud Formation scripts to build on demand EC2 instance formation.
* Strong experience in Hadoop and Big Data eco system related files such CSV/JSON & Parquet
* Capable of extracting data from an existing database, Web sources or APIs
* Experience in statistical modeling and data visualization with large data sets of structured and unstructured data
* Experienced with machine learning algorithm such as logistic regression, random forest, XGboost, KNN, SVM, neural network, linear regression, lasso regression and k – means
* Extensively worked on Python 3.5/2.7 (Numpy, Pandas, Matplotlib, NLTK and Scikit-learn)
* Implemented Bagging and Boosting to enhance the model performance
* Knowledge of Deep Learning (CNN, RNN, DNN) with an understanding of RNN, LSTM
* Experience in implementing data analysis with various analytic tools, such as Anaconda 4.0Jupiter Notebook 4.X,R 3.0(ggplot2, Caret, dplyr) and Excel 2010/2013
* Strong experience in Big Data technologies like Spark 1.6, Sparksql, pySpark, Hive 1.X, AWS(S3, Redshift, EC2)
* Experience designing and implementing fast and efficient data acquisition using Big Data processing techniques and tools.
* Experience in real time processing Framework (Apache Spark)
* Experience with running queries on Athena from s3 buckets
* Experience of database technology such SQL, PLSQL, and MySQL including NoSQL databases.
* Experience in Hadoop, MapReduce and Apache Spark
* Strong Experience using AWS Glue for ETL process
* Excellent understanding of source-code management principles and systems, particularly Git
* Strong hands-on experience in scripting languages such as Python, Shell, Bash scripting.
* Experience in Data Engineering (RedShift, Kinesis Data Stream)
* Strong exposure to creating Docker images and docker containers.
* Knowledge of how to automate the creation, deployment and maintenance of security rules. (WAF)
* Strong exposure to AWS and Google cloud platforms.
* Strong knowledge in Tomcat, WebLogic servers on different OS’s like Windows, Linux.
* Participated in the release cycle of the product which involved environments like Development, QA UAT and Production.
* Strong experience in scripting languages SQL, T-SQL, JSON, HTML, CSS, PHP, XML, Java(basic), R, Python, VB
* Worked with project documentation and documented other application related issues, bugs on internal wiki website.

---------------------------------------------------------------------------------------------------------------------------------------------------------------

**TECHNICAL SKILLS**

* **Cloud Platforms**: AWS, Azure, GCP
* **AWS SERVCIES**: VPC, Identity and Access Manager (IAM), EC2 Container Service, Elastic Beanstalk, Lambda, S3, CloudFront, Glacier, RDS, DynamoDB, ElastiCache, Redshift, Direct Connect, Route 53, CloudWatch, CloudFormation, CloudTrail, Athena, Amazon Elastic MapReduce (EMR), AWS Glue, Redshift, AWS IoT, SNS, API Gateway, SES
* **Versioning Tools**: Git, GitHub, Bitbucket, GitLab, Perforce.
* **IAC Tools:** Terraform CloudFormation.
* **Containers**: Docker.
* **Application/Web Servers**: Apache Tomcat, WebLogic 9.X/10.X/12C, Apache 2.X/1.3.X, JBoss 7.1.
* **Operating Systems**: Red Hat Linux, Ubuntu, Windows, MAC OS X, Windows Server.
* **Database**: Oracle, SQL Server, MySQL, Postgres, Redshift, Athena
* **Framework/APIs**: Django, Flask, Fastapi
* **Programming Language:** Logistics Regression, Naive Bayes, Decision Tree, Random Forest, KNN, Linear Regression, Lasso, Ridge, SVM, Regression Tree, XGboost, K - means Python 2.7/3.5 (Numpy, Scipy, Pandas, Seaborn, scikit learn, NLTK), R 3.0, SAS 9.1, SQL
* **Analytic Tools:** **Hadoop Ecosystem**Anaconda 4.0 (Jupyter NoteBook 4.X, Spyder), Rstudio Hadoop 2.X, Spark 1.6+ (Pyspark, Sparksql, MLlib**)**, MapReduce, Hive 1.X,
* **Database: Data Visualization**SQL Server 2008/2014, MongoDB 3.2 Tableau 9.4, Python - matplotlib

**PROFESSIONAL EXPERIENCE**

---------------------------------------------------------------------------------------------------------------------------------------------------------------

**Big Data/Machine Learning Engineer** (Sept 2020 – till date)

Nike –Remote

* Worked as a Big Data Engineer in the Market Planning Team where the company wanted to incorporate business and product drivers that predict market demand to be incorporated in business planning process to drive sales and boost demands
* Created DAGs in Apache Airflow for scheduled jobs execution
* Created DDL and DML to aide weekly data migration from Snowflake to Postgres
* Used F-Score, Precision, Recall, and A/B testing to evaluate Model performance
* Performed univariate and multivariate analysis on the data to identify any underlying pattern in the data and associations between the variables
* Worked on outlier identification with boxplot, studentized, K-means clustering using Pandas, NumPy
* Tested the models for problems like goodness of fit, over-fitting, multicollinearity, residual normality, etc.
* Worked on Machine Learning Algorithms Development for analyzing click stream data using Spark and Scala.
* Designed Distributed algorithms for identifying trends in data and processing them effectively.
* Used Spark and Scala for developing machine learning algorithms which analyses click stream data.
* Participated in features engineering such as feature creating, feature scaling and One-Hot encoding
* Tackled highly imbalanced Fraud dataset using under sampling with ensemble methods, oversampling with SMOTE and cost sensitive algorithms with Python Scikit-learn
* Developed machine learning code using spark MLLIB Used Spark SQL for data pre-processing, cleaning and joining very large data sets.
* Created data lake using spark which is used for downstream applications Designed and Developed Scala workflows for data pull from cloud-based systems and applying transformations on it.
* Created ETL jobs to load data and server data into Buckets and transported S3 data into the Data Warehouse.
* Created reports and dashboards using structured and unstructured data.
* Deployed Lambda and other dependencies into AWS to automate EMR Spin for Data Lake jobs
* Migrated Databases from Traditional Data Warehouses to Spark Clusters.
* Collaborated with various team & management to understand the requirement & design the complete system
* Implemented the complete Big data pipeline with real-time processing
* Created deployment pipelines for infrastructure provisioning using GitHub & Jenkins
* Used Spark API to stream data from various sources in real-time
* Built distributed high-performance systems using Spark and Scala Experience developing Scala applications for loading/streaming data into NoSQL databases (MongoDB) and HDFS.
* Developed Spark code in python using Spark SQL & Data Frames for aggregation
* Worked with Sqoop to ingest & retrieve data from various RDBMS like Oracle DB & MySQL
* Built the machine learning model include: SVM, random forest, XGboost to score and identify the potential new business case with Python Scikit-learn
* Wrote Hive queries to transform data for further downstream processing

**AWS Big Data/Machine Learning Engineer** (Mar 2020–Sept 2020)

PSEG – Bethpage, NY

* Deployed Lambda and other dependencies into AWS to automate EMR Spin for Data Lake jobs
* Set up continuous integration/deployment of spark jobs to EMR clusters
* Scheduled spark applications/Steps in AWS EMR cluster.
* Installed and configured Apache Hadoop, Hive and Pig environment on the prototype server.
* Configured database to store Hive metadata.
* Used big data tools (Sparksql, Mllib) to conduct the real time analysis of dataset
* Built the machine learning model include: SVM, random forest, XGboost to score and identify the potential new business case with Python Scikit-learn
* Involved in variable selection, model optimization and model selection
* Loaded unstructured data into Hadoop File System (HDFS).
* Created ETL jobs to load data and server data into Buckets and transported S3 data into the Data Warehouse.
* Created reports and dashboards using structured and unstructured data.
* Joined various tables using spark and Scala and ran analytics on top of them in EMR
* Applied spark streaming for real time data transforming.
* Created multiple dashboards in tableau for multiple business needs.
* Installed and configured Hive and written Hive UDFs and used piggy bank a repository of UDF’s for Pig Latin.
* Implemented Partitioning, Dynamic Partitions and Buckets in HIVE for efficient data access.
* Implemented test scripts to support test-driven development and continuous integration
* Developed and executed a migration strategy to move Data Warehouse from SAP to AWS Redshift.
* Designed and built multi-terabyte, full end-to-end Data Warehouse infrastructure from the ground up on Confidential Redshift for large scale data handling Millions of records every day.
* Implemented and Managed ETL solutions and automating operational processes.
* Optimized and tuned the Redshift environment, enabled queries to perform up to 100x faster for Analytic software
* Designed and Developed ETL jobs to extract data from Salesforce replica and load it in data mart in Redshift.
* Published interactive data visualizations dashboards, reports /workbooks on Tableau and SAS Visual Analytics.
* Designed Logical and Physical data modelling for various data sources on Confidential Redshift

**AWS Data Engineer** (April 2019 – Mar 2020)

World Wrestling Entertainment– Stamford, CT

* Used EMR to process and analyze vast amounts of 3rd party data using frameworks and related open-source projects
* Processed data for analytics purposes and business intelligence workloads with EMR Clusters
* Continuously captured gigabytes of data per second from various sources using Kinesis firehose
* Implemented data availability of data for collection in millisecond to enable real-time analytics use cases such as real-time dashboards, real-time anomaly detection and dynamic pricing.
* Selected models implemented include logistic regression, decision trees, random forest, SVM with Python 3.3 Scikit-learn
* use Amazon EMR to transform and move large amounts of data into and out of other AWS data stores and databases, such as Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB.
* Defined bootstrap actions on EMR clusters and specify apps to be installed on instances
* Ran complex Data Queries from s3 buckets in Athena
* Automated cloud metrics statistics and large dataset collection and general Analysis on how to reduce cost
* Provisioned AWS resources using management console as well as Command Line Interface.
* Created scripts in Python which integrated with Amazon API to control instance operations.
* Implemented the Extraction, Transformation and Loading of Large Datasets with AWS Glue
* Wrote Ansible Playbooks with Python SSH as the Wrapper to Manage Configurations of AWS nodes and Tested Playbooks on AWS instances using Python.
* Used Shell scripts and bash to make API calls in order to generate data reports in AWS environment
* Automated the ingestion of Server metrics into redshift table in python
* Built external Dashboard in Tableau to visualize resources performance for executive view
* General Cloud and infrastructure operation and maintenance
* Weekly tag compliance reporting in AWS config (Scripted)
* Used Shell scripts and bash to make API calls in order to generate data reports in AWS environment
* Set up and built AWS infrastructure with various services available by writing cloud formation templates in Json.
* Planned, built and configured network infrastructure within VPC with public and private subnets, configure routing tables, security group and internet gateway.
* Migrated the production MYSQL DB to new AWS RDS Aurora instance
* Configured NACL, Security Groups and WAF to secure the application and VPC network depending on access parameters provided and conditions.
* Automated provisioning of cloud infrastructure using CloudFormation, Terraform; and application

**AWS Big Data Engineer** (Sep 2018 – Mar 2019)

Vexcel, Inc – New York, NY

* Ingested Performance Metrics (Large Dataset) into Redshift tables via Scripting (API Calls)
* Automated Athena Query with lambda functions
* Implemented data availability of data for collection in millisecond to enable real-time analytics use cases such as real-time dashboards, real-time anomaly detection and dynamic pricing.
* Implemented the Extraction, Transformation and Loading of Large Datasets with AWS Glue
* Provisioned AWS resources using management console as well as Command Line Interface.
* Used Amazon EMR to transform and move large amounts of data into and out of other AWS data stores and databases, such as Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB.
* Ran direct from s3 using Athena environment
* Automated the Transformation of large dataset in python for better Analysis
* Migrated the production MYSQL DB to new AWS RDS Aurora instance
* Configured NACL, Security Groups and WAF to secure the application and VPC network depending on access parameters provided and conditions.
* Created Python scripts to automate AWS services which include web servers, ELB, Cloud front distribution, databases, security groups and application configuration
* Created SNS topic that triggers based on Datapoint threshold
* Setup and configured VPC PrivateLink (Endpoints), Direct Connect, Transit VPC, VPC Peering, Reverse Proxy Caching, CDN, IPsec and VPN for secure data transfer, and secured connection to other AWS services and other AWS accounts.
* Automated provisioning of cloud infrastructure using CloudFormation, terraform; and application configuration and deployment using Terraform and Ansible.
* CI/CD pipeline management through Jenkins. Version control through GIT.
* Tested and deployed automation with Docker containerization and Jenkins. Automation of manual tasks using Shell scripting.
* Used Docker in Environment variables, Configuration files, Option types and Strings & integers.
* Leveraged AWS cloud services such as EC2, auto-scaling and VPC to build secure, highly scalable and flexible systems that handled expected and unexpected load bursts.
* Created and managed IAM user accounts and role-based policies for access to AWS services.
* Implemented and maintained monitors, alarms, and notifications for EC2 instances using CloudWatch and SNS.
* Performed S3 buckets creation, access policies, and archive outdated data to Glacier through LifeCycle Policy configuration.
* Launched databases via RDS to support web application functionality.
* Created low latency website using S3, CloudFront, and Route53.
* Installed Apache and Nginx Web Server on EC2, (proxy server as needed).
* Created and manage users, permissions, directories and files in Linux environment.
* Set up Ansible environment for configuration management, Dynamo DB, Elastic cache and Lambda.
* Implementation of compliance with CIS Benchmark on AWS environment.
* Enforced company’s policies and procedures during design and implementation.
* Defined and deployed monitoring metrics, and logging systems on AWS environment. Application and system log analysis using CloudWatch, CloudTrail, Data Dog, Logic Monitor.
* Designed, developed and implemented high availability, backup and recovery strategy for applications and database.
* Hands-on experience and good knowledge of AWS services: EC2, S3, VPC Endpoints, VPC PrivateLink, VPC Peering, Transit VPC, Network Protocol, Subnetting (Private and Public), IPsec, VPN, Route53, ELB, Auto scaling, SQS, SNS, IAM, DynamoDB, Lambda, Amazon Kinesis, Cloud Front, Cloud Formation, CloudWatch, CloudTrail, Storage Gateway, internet Gateway.

**AWS Developer/Engineer**. (May 2015– Sep 2018)

New York City HRA - New York, NY

* Planned, designed and developed solutions on Amazon Web Services.
* Worked closely with customers to get to know their business and technical goals and challenges.
* Managed the relationship between the organization's technical demands and the capacity to be pulled from the AWS cloud.
* Ensured compliance with the company’s policies and procedures during design, development and implementation.
* Automated of cloud infrastructure configuration and deploymsent using cloud Formation and Ansible.
* Designed and deployment of public and hybrid cloud solutions for customers using appropriate services based on requirements.
* Provisioned secure environment for AWS VPC network, and implement infrastructure as code using configuration management tools. Implementing CI/CD pipelines using Jenkins and monitor infrastructure.
* Migrated customers' on-premise workloads(applications) to the AWS cloud platform.
* Designed, developed and implemented high availability, backup and recovery strategy for applications and database.
* Evaluated and improved existing AWS deployments.
* Defined and deployed monitoring metrics, and logging systems on AWS.
* Identified appropriate use of AWS and deployment, and operational best practices.
* Hands-on experience and good knowledge of AWS services: EC2, EBS, S3, RDS, VPC, VPN, Route53, ELB, Auto scaling, SQS, SNS, IAM, DynamoDB, Cloud Front, Cloud Formation, ECS, CloudWatch, CloudTrail, Storage Gateway, internet Gateway, NAT Gateway.
* Scripted using with Shell, SQL, JSON, YAML, XML, Linux script.

**ETL Data/Database Analyst**  (Jan 2013 – May 2015)

Argo Groups, New York, NY

* Created SSIS packages to extract data from OLTP to OLAP systems, and scheduled jobs to call the packages.
* Applied various data transformations like Slowly Changing Dimension (SCD), Aggregate, Sort, Multicasting, Conditional split, Derived column etc.
* Created complex T-SQL, Stored Procedures, functions, triggers and DB objects as and when needed to convert data logically as per the business requirements.
* Designed and prepared standard and ad hoc reports and summaries for statistical analysis.
* Extracted actionable insight from data to provide clear recommendations for the users.
* Data modelling, database design, development, testing and implementation with RDBMS.
* Designed, developed and maintained customized reports and dashboard using excel and tableau
* Created ODBC connection, ADO DAO objects to connect Oracle SQL Server databases

• Wrote Python Scripts to extract data from SQL Server database for further Analysis

• Automated weekly reports with Python Scripting

* Data migration and Data validation/Testing through VBA modules between MS Access and MS Excel
* Wrote VBA codes for automated transferring of data between spreadsheets and workbooks
* Used Pandas, NumPy, Seaborn, SciPy, Matplotlib, Scikit-learn, and NLTK in Python for developing various machine-learning algorithms.
* Automated Importing of data to Access and SQL Server from flat files and Excel spread sheets
* Developed Python scripts to automatically run vba macros that generated daily reports

**Other competencies**

Skype For Business/Lync || Documentation of processes and procedures || Writing Business Requirement||Project Management|| Tools: ServiceNow, Remedy, Clear-Vision. Customer Focus Service. Analytical. Troubleshooting Crisis Management/Escalation. Client Relationship Management. Cross-Functional Team Management. Remote Desktop Support. Program Management. System Multitasking. Remarkable analytical and problem-solving abilities. Ability to work flexible schedules.

**EDUCATION AND PROFESSIONAL CERTIFICATIONS**

* Master’s in Economics (City College of New York)
* Bachelor’s in Economics (Stony Brook University).

**PERSONAL MINUTIAE**

References: Furnished upon request