

CV

Greg Nwosu

Summary

I'm known for having a can-do attitude, I find solutions where most others can't. I often pioneer new approaches for teams to get around issues and find approaches to issues that most see as unsurmountable. I'm friendly and fit in, I bring up the quality of the less experienced and love to learn from anyone that I can. I have good knowledge of Java , Python , Scala, and Haskell, SQL, bash shell scripting. I have experience with different cloud platforms GCP, Azure and AWS. I love learning new things and have a passion for trading. I have deep knowledge of Agile/Scrum, and I've been on many teams in the past and draw on this experience of what works to figure out a way forward when things get difficult.

I am [certified](#) in machine learning from Stanford University's [course](#). I'm a [graduate](#) of [\[\[https://classroom.udacity.com/nanodegrees/nd101\]](https://classroom.udacity.com/nanodegrees/nd101) Udacity's Deep Learning Foundation Course]] and now I'm halfway through the [Udacity's Deep Reinforcement Learning NanoDegree](#). My goal is to continue in the position of being adept in cutting-edge machine learning techniques and data science together with my current skills in big data engineering and ETL.

Current Interests

- Algo-Trading, wrote a home project that analyses candles and recommends trades based on technical analysis.
- Arbitrage: wrote a system that uses, bellman ford algorithm to find negative weight cycles in currencies so I can make money by transferring money between currencies at my bank.
- Intermediate Advanced [Haskell programming](#)
- deep reinforcement learning.
- [category theory](#)
- Bayesian Methods for Data Analytics.
- Machine Learning, started playing with Large Language Models

Other hobbies include:

- Cycling
- Callisthenics
- studying civil litigation and human rights law
- Zen Meditation.
- I have also trained to teach adults and am a mentor to children learning to code at [CoderDojo](#).

Commercial & Technical Experience

October 2021 - Present : **Shell**

Responsibilities

- Front Office , Author and Designer of **Power Purchase Agreements Reconciliation Tool** Technical

Lead Engineer, Designer, Developer

- Front Office , **Power Purchase Agreement Inter-book Transfer Tool** Lead Engineer, Designer, developer

Technologies

Azure Data Factory, Azure DevOps, Databricks, Pyspark, Databricks DeltaLake, Databricks DeltaLive, Databricks Auto-loader

Achievements:

- Wrote a spark UDF that decompresses zip files, zip decompression is not supported in spark, but I managed to support it, without having to write to temporary storage.
- (re)Wrote a **reconciliation tool to reconcile Power Purchase Contracts (PPAs)**

between Shell and a third-party broker within the first 6 months of joining. Currently, the system is used directly by traders on its newly formed power trading desk. The system provides near real-time tracking of PPAs from broker order books through shell middleware and into the Shell order books. The system achieved 4 x9's up-time and became so reliable it was moved to a "run and support" model. The system includes data-quality analysis, slicers and dice controls, summaries, drill-down and graphical visualizations.

- The existing system for transferring PPAs from the broker to Shell was slow and unreliable. Delivered [\[\[https://adb-8189235268511358.18.azure-databricks.net/?o=8189235268511358#notebook/4039834138178088/dash-board/1665125115797297/present\]\]](https://adb-8189235268511358.18.azure-databricks.net/?o=8189235268511358#notebook/4039834138178088/dash-board/1665125115797297/present) a real-time streaming support system Inter-book Transfer Tool

for transferring trades]] between an external PPA broker and Shell. The system generates a VWAP curve from energy price forecast data and executes orders to cover predicted liabilities of third-party renewable energy contracts. The system uses Databricks DeltaLake format for full audit coverage over time. It is possible to view and query the state of the system at any time in the past. The new Power Purchase Inter-Book transfer tool leverages DeltaLive and Databricks auto loader for real-time streaming updates, so that traders can ensure timely transfer of PPAs to the Shell order book.

Qualifications

- Mennta **Introduction To Trading**
- Mennta **Introduction To Commodity Derivatives**

March 2021 - September 2021 : **Gousto**

Responsibilities

Senior Data Engineer.

Technologies

AWS DMS, AWS Glue, AWS EC2, AWS EMR, AWS Cloudformation, Python, Databricks Spark, Databrick Delta Lake,

Airflow, Kubernetes

Achievements:

- Wrote a proof of concept to show how to attach hooks to spark streams in python, to enable automatic alerting on ingestion failure of a streaming process. – Proposed two ways, one of which used the sparks internal python java bridge.
- introduced BDD to help BAs verify business logic.
- Used AWS DMS Service to safely migrate data from databases into AWS S3 storage in real-time.

March 2019 - March 2021 : **Shell**

Responsibilities

1. Part of the SWAT (SoftWare / Analytics Team) team in The Agile Hub tasked with improving the architectural approach and development of extracting business value (monetizing) of data from under-achieving shell projects. Projects usually exist for 5
 - 7 weeks. Projects typically deliver technical advisory reports and a proof of concept.
2. Cargo Tracking wrote software for front desk trading analysts that provide consensus on the location of crude and distillates from various providers.

Technologies

Helm Charts, Pyspark, Azure Data Factory, Azure Cosmos, AKS, Airflow, Kubernetes

Achievements:

- Led a SWAT team to increase the performance of pricing reports, we managed to render a report in 15 seconds that was previously taking 15 minutes. Deployed proof of concept airflow pipeline in Kubernetes that also has jobs that run in AKS
- mentored junior members in development practice, GitHub, unit testing, code-quality.
- trained remote developers in Github, Git
- wrote a data pipeline that ingests from google sheets which enabled business analysts to collaborate on cargo tracking reference
- introduced the team to event-driven architecture, which resulted in 10x speed of data processing.
- used graph algorithms (using networks in python) to enable

visualization and therefore aid simplify complex decision paths in business logic introduced BDD to help BAs verify business logic.

September 2018 – March 2019: [Quantexa](#)

Senior Data Engineer

Responsibilities

- Mentoring architecture and quality refactoring of Quantexa's award-winning analysis and fraud detection platform.
- A team member responsible for accelerating the code quality of Quantexa implementations as they are rolled out onto banks; my duties

include code reviews of client implementations and code quality and improvement of the core product.

Technologies

- GCP
- Spark
- Frameless
- Scala
- Spark testing framework
- Cats
- ScalaCheck
- Git-lab
- typesafe config

Achievements

- Redesign and refactoring the core scoring framework.
- Introduced property-based testing using ScalaCheck to bring test coverage over 90%
- Implementation of typesafe config for reducing release errors due to misconfiguration

June 2017 - January 2019: [MyDrive](#)

Senior Data Engineer

Responsibilities

Design and implementation of MyDrive Data Lake ETL process Ad-Hoc Data Cleansing for DataScientists

Technologies

- [AWS](#)

- [EC2](#)
- [EMR](#)
 - automation
- [S3](#)
- [Jupyter-Notebooks](#)
- [Lambda](#)
- [Redshift](#)
- [RDS](#)
 - [postgres](#)
- Data visualization
 - [Matplotlib](#)
- [AnswerRocket](#)
- [hive](#)
 - Hive performance tuning
- [Scala 2.11.6 \(for EMR compatibility\)](#)
- [Haskell \(GHC 8.0.1\)](#)
- [clojure](#)
- [Spark 2.1.0 \(for EMR compatibility\)](#)
- [Glove \(python implementation of word2vec\)](#)
- [PyMc \(for Bayesian analysis\)](#)
- [Parquet](#)
- [Nginx](#)
- [Gitlab](#)

Achievements

- Co-Design and implementation of a Data Lake for organizing the ingestion and processing locations of OLTP /OLAP data streams in scala/spark, airflow
- Audit of all of MyDrive data on S3 for GDPR, currently around 30TB
- Introduced the best industry standards for python development; via type-based Python (mypy) better virtualization (pipenv) and a stronger project framework
- began to learn some production-level terraform for cloud-agnostic infrastructure

Skills Gained

- AWS Lambda
- AWS DMS
- AWS Athena
- AWS Glue
- Apache Airflow

- AWS ECS
- AWS ECR
- AWS Kinesis
- Mypy
- Terraform
- Boto3

July 2016 - April 2017: [Aimia](#)

Machine Learning Engineer

Responsibilities

Working within a new team for monetizing Aimia's vast data repository. My responsibilities initially were helping migrate Aimia's data processing services to a more robust platform (EMR). I then moved into co-designing and developing a platform to capture all the data needed for the machine-learning techniques we wished to use. This completed, I then developed real-time ETL spark streams for data out of the legacy hive data warehouse so data could easily be used in a variety of machine learning algorithms. Choosing parquet because of its schema evolution and performance properties. I completed a brief prototype migration to the [answer rocket](#) platform so that some less technical analysts could evaluate natural language analytics.

Technologies

- [AWS](#)
 - [EC2](#)
 - [EMR](#)
 - automation
 - [S3](#)
 - [Jupyter-Notebooks](#)
 - [Lambda](#)
 - [Redshift](#)
 - [RDS](#)
 - [postgres](#)
 - Data visualization
 - [Matplotlib](#)
- [AnswerRocket](#)
- [hive](#)
 - Hive performance tuning
- [Scala 2.11.6](#) (for EMR compatibility)
- [Haskell](#) (GHC 8.0.1)
- [clojure](#)

- Spark 2.1.0 (for EMR compatibility)
- Glove (python implementation of word2vec)
- PyMc (for Bayesian analysis)
- Parquet
- Nginx
- Gitlab

Achievements

- Set up Continuous Integration and Unit Testing Infrastructure
- Helped complete migration from EC2 to EMR for greater resilience to failure
- Implemented an HQL (Hive SQL) Parser in Haskell to auto-generate Spark streaming schema from the abstract syntax tree
- Engineered, Designed and developed real-time streaming for the majority of data warehouse into big-data platform in AWS
- Helped set up the continuous integration environment
- Implemented word2vec for cluster classification of websites
- Made a prototype answer rocket database for an evaluation of natural language analytics

Skills Gained

AWS Clojure Nginx Docker Kafka Bayesian Analysis (PyMC)

April 2015 - June 2016, Self Employed: Barclays Capital

Big Data ETL Engineer

Responsibilities

Ingesting Risk Data into Barclays BigData System Design meetings and code quality

Technologies

- Hadoop
- Apache Spark
- Apache Flume
- Kafka
- Protobuf/Parquet/Avro
- Berkley DB

Achievements

- Set up Continuous Integration and Unit Testing Infrastructure
- developed systems to ingest terabytes of risk profile data into hdfs
- helped set up a continuous integration environment

- helped mentor graduate intern
- developed comprehensive testing using ScalaCheck test generation
- integrated apache flume with Barclays in-house data warehouse format
- re-engineered Barclays interface to Solace Messaging in Scala

Skills Gained

Apache Flume Apache Spark ScalaCheck Solace Messaging Kafka

September 2014 - February 2015, Blinkbox Books

Senior Scala Engineer

Responsibilities

- Design of and implementation of REST APIs, in swagger
- Automated verification of APIs against swagger in Tests
- Wrote property-based testing code for storage service
- Interfacing with Microsoft Azure Storage Framework with Scala
- Implementation of Scala code
- Writing functional tests in Property Based BDD style
 - ScalaCheck Property
 - FlatSpec for BDD
- Review and Merging of Pull Requests in Git hub
- Diagnosis of issues with Continuous Integration and Deployment preparation
- AMQP configuration

Technologies

- Scala
- ScalaCheck
- Spray.io
- FlatSpec
- Akka
- Github, Git
- Swagger
- REST
- HTTP
- Azure
- RabbitMQ AMQP

Achievements

- Designed, Developed and Deployed the first version of REST endpoint for storage agnostic cloud-based big data service, with redundancy across storage providers

- Improved Scala, Git, Github, REST knowledge, AMQP/RabbitMQ knowledge

Skills Gained

- AMQP/ RabbitMQ
- REST
- Spray.io/ Akka

August 2013 - August 2014, RBS

Infrastructure Developer

Working with the maintenance and monitoring of RBS's big-data risk aggregation platform. I used a combination of

- java 6
- oracle coherence
- Unix bash shell scripts
- Haskell
- Scala
- Python

I am responsible for

- capacity planning
- monitoring bandwidth throughput and latency to ensure the smooth running of the platform.
- Bidding for budget and rationalizing legacy infrastructure.

Responsibilities

- Dev Ops
- Capacity management
- Infrastructure Bidding.
- Technologies
 - Java 6
 - Python
 - Scala
 - Scalaz
 - Continuous Integration (TeamCity)
 - Dev-ops
 - Coherence
 - capacity planning
 - performance profiling
 - Scala-sbt

- ScalaCheck
- Scala-Specs

Skills Gained

- Bidding
- Budgeting
- Coherence
 - performance
 - capacity analysis
- FX
- Git
- Scala
- Scalaz
- Scala Check
- Scala Specs
- Python
- Haskell
- DevOps
- Scrum

Achievements

- learned scrum/agile in depth here, gained in-depth knowledge of scrum.
- Recently developed a £500k proposal for new infrastructure as a result of a profiling and capacity plan I put in place.
- Presented plan to the RBS board and won approval for the spend for updating the nodes in a coherence cluster based on profiling, coherence cluster shock and datagram analysis measurements.
- Dev-ops scripts written in Haskell
- 6 months of commercial advanced
 - Scala
 - Scalaz
 - ScalaCheck

Jun 2010 – September 2013, IG Group

Direct Market Access & Smart Order Routing Java Developer

Responsibilities

- General FIX Connectivity
- Instrument Downloads and Trading

- Designed coded and accredited IG trading Gateways to be compliant with external exchange trading protocols.
- Daily instrument downloads from exchanges
- API client connectivity and accreditation
- Smart Order Routing (SOR)
 - tweaking SOR trading strategies
 - Fault Diagnosis and SOR Order Resolution
- certification with external companies
- Last line of support for trading gateways and connectivity issues

Technologies

- Java 6
- Java 7
- LMAX disruptor
- Multi-threading
- Linux
- Oracle SQL
- SQL Developer
- Clover
- Sonar
- Maven2
- Maven 3
- Bamboo
- Python 2.6
- Python-Requests
- BDD
- JBehave
- Domain-Driven Design
- Concurrent Programming Functional Programming
- Low Latency Algorithms
- Disruptor Pattern
- Bash Shell Scripting

Achievements

- Introduced BDD/TDD to the team and increased productivity by 20%
- Designed and implemented the initial framework for IG's Gateways
- CHIX, Bats, Bloomberg, CommerzBank, UBS
- LSE, (Including its winning LSE Millennium Gateway, IG had no downtime on LSE launch compared to 80% of finance houses)

- Designed and implemented Connectivity for Algorithmic Exposure Hedging System
- Standardised a way to debug running processes across multiple firewalled SSL zones
- Introduced BDD and Domain Driven Design to the DMA Connectivity team

Skills Gained

- Trading
- FX
- Securities
- EasyMock Mockito
- JBehave
- SOR
- Order Routing
- Trading
- FIX 4.2
- FIX5SP2
- Cameron
- git-svn

Apr 2008 – June 2009, Stan James

Working with a top gambling company; Developing a trading platform and desktop application for traders in sports betting. I played key roles in technical decision making, agile estimating, planning and retrospectives, as well as implementation, testing, refactoring and maintenance. Initially responsible for the inception of the quants module for event pricing and later contributed all other modules.

Skills Gained

- Agile Methodology
- Scrum
- Agile Estimating and Planning
- Sports Betting
- GWT
- Java Swing
- Selenium
- Fitness
- Oracle Coherence
- Hibernate
- Spring
- core Java
- JUnit

- Weblogic
- Oracle

Education

2002-2003 University College London

M.Sc. Intelligent Systems (Incomplete)

1. Course Content
 - Neural Networks
 - SVMs
 - Decision Trees
 - Learning theory
 - Maximum Likelihood Estimation
 - Bayesian Decision Theory
 - Hidden Markov Models
 - EM Algorithm
 - ICA
 - Clustering
 - Factor Analysis
 - Mixture Models
 - Monte Carlo Sampling Methods
 - Graphs
 - Bayesian Networks
2. Software Research paper:
Detecting Faces in Images a Survey of different approaches

1994-1997 University of Birmingham

2.1 B.Sc. Computer Science & Artificial Intelligence

1. Course Content:
 - Concurrent and Object Orientated Programming in C++
 - TCP-IP
 - UNIX real-time shared Memory and Semaphores
 - Computer Graphics
 - Advanced Interface Design
 - Human-Computer Interaction
 - Relational Database Theory
 - HTML Design / CGI Programming

- Expert Systems
 - Neural Networks
2. Software Research paper:
Melody Composition using Web-based Genetic Algorithms.

1992-1994 St Francis Xavier College

3 A-levels including A in Computer Science

1987-1992 John Paul Secondary School

9 GCSE's Grade A-C