Adrian Ng, MSc.

Seeking Junior-Level Data Engineering Opportunities **Email:** contact@adrian.ng **Website:** adrian.ng

Profile

I am a Computer Science graduate passionate about programming and a career in Data Engineering. I seek opportunities that meet my growing experience in *Java* – a language I have used in numerous academic projects ranging from the implementation of financial models to large-scale data processing with *Apache Hadoop* and more.
Prior to postgraduate study, my expertise in *SQL development* focused on the implementation of segmentation processes for a number of clients including: *Virgin Media*, *TUI*, *UPC*, *MSD*, *Volkswagen*, and *KwikFit*.
After graduation, my accomplishments as a Data Analyst at *Manchester City FC* leaned more towards Data Engineering, which leads me now to pursue a career in this field.

Education

- **Royal Holloway – Department of Computer Science** Sept. 2016 – Dec. 2017
  Master of Science in Data Science and Analytics *with Distinction*
- **King's College London – School of Engineering** Sept. 2007 – July 2010
  Bachelor of Engineering in Mechanical Engineering *Upper Second Class with Honours*

Technologies

Languages: • Java 8 • SQL
Software: • IntelliJ IDEA • SQL Server Management Studio • Git • VS Code • Jira • Maven

Java Projects

Value at Risk
adrian.ng/java/var/

Dissertation project: estimating *VaR*, a measure of risk, for an investment portfolio containing stocks, options, and deltas. A number of VaR measures and time-series variance estimates were implemented:

VaR Measures

- Model Building
- Historical Simulation
- Monte Carlo Simulation.

Moving Average Processes

- *Equal Weighted*
- *Exponentially Weighted Moving Average (EWMA)*
- *GARCH(1,1)*

- implemented *Levenberg-Marquardt* algorithm for optimisation of *GARCH(1,1)* parameters
- made use of object-oriented techniques and patterns to accommodate these numerous approaches
- gained efficiencies in using Java's `concurrency` APIs to parallelize the 100,000+ random walks generated by *Monte Carlo* when simulating stock price movements
- utilised *Google Fiance/ Yahoo Finance* APIs to source time-series financial data

Data Mining
adrian.ng/java/enron

Building *MapReduce* applications for large scale data mining and processing.

MapReduce
- aggregation of *Twitter* data
- scraping a large collection of emails in the *Enron Corpus*
- extraction of nodes/edges in communications network

Hadoop
- applications ran on self-hosted, single-node cluster and distributed clusters
- `hdfs dfs` commands used to ETL via *HDFS*

Spark
adrian.ng/scala/spark/enron1

A subsequent self-learning exercise achieving:
- translation of *MapReduce* applications to `Scala`, running in a *Apache Spark REPL*
- reduced code verbosity, maintained ETL via *HDFS*

Option Pricing
adrian.ng/java/options/

Implementing numerous approaches to pricing options and their various payoffs

Options • Monte Carlo Simulation • Black Scholes • Binomial Trees
Payoff • American • Asian • European

These approaches made probabilistic assumptions, so *Apache Commons Math* API was used.

Summarizing financial data
adrian.ng/java/yahoofinance/#stream

A self-taught exercise to gain familiarity with Java 8's `Stream` API. I was able to implement approaches to computing mean and variance estimates from an immutable collection of time-series financial data.

Manchester City Football Club

Data Analyst | *Fan Relationship Management* | *Jan. - July 2018*

New York City FC Project

I took ownership of this project to integrate *NYCFC's* transactional and demographic data with *City Football Group's* data-warehouse. This six-month project involved many phases including: discovery, engineering, and analysis. Data came from multiple external sources each with differing schema: *NYCFC*, *Ticketmaster Salesforce*, *Major League Soccer*.

Data Pipeline

- built pipeline ingesting data from multiple databases, replacing *Informatica* solution
- achived efficient ETL process via effective DML & DDL (`OPENQUERY`, `MERGE`)

Data Cubes

- implemented up-stream computation of drill-down/roll-up for all permutations, minimising bandwidth across distributed servers
- eliminated real-time computation in *Tableau* front-end, improving UX

Mentoring

- dedicated time to mentoring junior colleagues remotely in Manchester/New York
- organised weekly workshops teaching basic DML and advanced DDL
- developed additional material on my website to supplement these workshops

GDPR Pipeline

- integrated new GDPR schema into existing datastores (*SQL*, *Salesforce*)
- worked with SQL developers to provide schema specification and UAT testing on new processes
- built efficient `MERGE` process featuring relational database design

Customer Churn Model

Modelling MCFC/NYCFC customers' future propensity to churn via *logistic regression*.

- contributed to feature selection involving: data extraction, imputation, and normalisation
- researched other models (e.g. *Beta-Geometric/Beta-Bernoulli*), academic papers, `R Studio` API

Creator (now Inspired Thinking Group)

Senior CRM Campaign Executive | *SQL Development* | *Dec. 2013 - Sept. 2016*

I developed a number of `SQL` processes to transform customer data into CRM segments. On occasion, I took responsibility for resourcing and managing the team's workload in *Jira*.

Virgin Media Segmentation

(https://adrian.ng/SQL/cte/Recursion/ (https://adrian.ng/SQL/misc/openquery-xml)

- built a bespoke import tool around `BULK INSERT` to ingest millions of tuples distributed across multiple flat-files
- achieved efficient joining of local and remote data by combining `OPENQUERY`, `XML`, and dynamic `SQL`
- implemented efficient regex parsing via recursion (operating akin to `flatMap` in `Java 8`/`Scala`)

Volkswagen Onboarding

- built and tested a new segmentation process for broadcasting email *and* SMS.
- provided schema specification to developers for data warehousing

TUI Redesign

- worked closely with TUI to integrate a new, responsive design of their *Thomson* and *First Choice* large deployment broadcasts (5M+ recipients)
- wrote `TCL` scripts for dynamic `HTML` merges and gained efficiencies by moving expensive operations upstream
- gained recognition with client and was awarded at the end of this three-month project

Seatwave (now Ticketmaster)

Marketing Analyst Intern | *Commercial Team* | *May 2013 - Dec. 2013*

Using *SQL Server Management Studio*, I wrote DML capable of querying the transactional/customer databases to return data for warehousing, reporting, and segmentation. I also worked on pricing and spatial analyses, using *QGIS* as a visualisation tool.