

Adrian Ng, MSc.

Seeking Data Engineering Opportunities

Email: adrian.j.ng@gmail.com **Portfolio:** adrian.ng **Location:** London

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.

My professional expertise in *SQL development* focuses on the implementation of segmentation processes for a number of clients including: • Virgin Media • TUI • UPC • MSD • Volkswagen • KwikFit

And with my technical and data-oriented accomplishments at *Manchester City FC*, I now pursue a career in the field of Data Engineering.

EDUCATION

Royal Holloway 2017
Data Science & Analytics

- Master of Science
- Pass with Distinction
- Department of Computer Science

King's College London 2010
Mechanical Engineering

- Bachelor of Engineering
- Upper Second Class with Honours
- School of Engineering

LANGUAGES

- Java 8
- T-SQL

SOFTWARE

- IntelliJ IDEA
- MS SQL Server
- VS Code
- Git
- Jira
- Maven

JAVA PROJECTS

Value at Risk

Dissertation

adrian.ng/java/var/

Estimating *Var*, a measure of risk, for an investment portfolio containing stocks, options, deltas.

Var Measures

- Model Building
- Historical Simulation
- Monte Carlo Simulation.

- implemented *Levenberg-Marquardt* algorithm for optimisation of *GARCH(1,1)* parameters

- made use of object-oriented techniques and patterns

- gained efficiencies using Java's **concurrency** APIs to parallelize the 100,000+ random walks generated by *Monte Carlo* when simulating stock price movements

Volatility Estimates

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

- utilised *Google Finance/Yahoo Finance* APIs to source time-series financial data

Data Mining

Large-Scale Data
Storage & Processing

adrian.ng/java/enron

adrian.ng/scala/enron1

MapReduce Wrote *MapReduce* applications involving:

- aggregation of *Twitter* data, utilising **twitter4j** API
- scraping a large collection of emails in the *Enron Corpus*
- extraction of communications graph consisting of nodes/edges

Hadoop

- Applications ran on both single-node (self-hosted)/ distributed-node clusters
- Interfaced with *HDFS* via terminal command-line

Spark

Utilised an *Apache Spark REPL* to achieve:

- translation of *MapReduce* applications to **Scala**
- reduced code verbosity
- ETL via *HDFS* using **sparkcontext** API

Option Pricing

Methods of
Computational Finance

adrian.ng/java/options/

Implemented numerous approaches to pricing options and calculating payoff:

- 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/

An exercise in using 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.

Webpage Scrapping

Wrote a program for scraping data from webpages.

- Utilised **HtmlUnit** and **Selenium** APIs
- Traversed DOM and parsed child elements via **xPath**

MANCHESTER CITY FOOTBALL CLUB

Data Analyst

Fan Relationship Management

Jan. - July 2018

NYCFC Data Integration Project Owner	Integrated New York City FC data into our analytical warehouse. Six-month project involving data discovery, analysis, engineering. Multiple data sources were involved: • NYCFC • Ticketmaster • Salesforce • Major League Soccer
Data Pipeline	<ul style="list-style-type: none">• built pipeline ingesting data from multiple databases, replacing <i>Informatica</i> solution• achieved speed improvements using efficient DML & DDL (<i>OPENQUERY</i>, <i>MERGE</i>)
Data Cubes	Storing analytical datasets in <i>Data Cubes</i> achieved <ul style="list-style-type: none">• up-stream computation of all drill-down/roll-up levels and <i>GROUP BY</i> permutations• reduction in size of dataset, minimising bandwidth across distributed servers• improved user-experience in <i>Tableau</i> front-end
Mentoring	Dedicating time to mentoring junior colleagues remotely in Manchester/New York <ul style="list-style-type: none">• organised weekly workshops teaching basic DML and advanced DDL• developed additional material on my website to supplement these workshops• aimed towards self-sufficiency in writing database queries/stored procedures
GDPR Pipeline Technical Lead	<ul style="list-style-type: none">• integrated new GDPR schema into existing datastores (<i>SQL</i>, <i>Salesforce</i>)• provide schema specification to <i>SQL</i> developers, advocating for indexable data types• built efficient <i>MERGE</i> process featuring relational database design• implemented a process to wipe personalised data belonging to any non-consenting individual stored in our data-warehouse
Customer Churn Model	Modelling MCFC/NYCFC customers' future propensity to churn via <i>logistic regression</i> . <ul style="list-style-type: none">• contributed to feature selection via: – data extraction – imputation – normalisation – R modelling• researched alternate models (e.g. <i>Beta-Geometric/Beta-Bernoulli</i>), academic papers, R APIs

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 <small>adrian.ng/SQL/recursion adrian.ng/openquery-xml</small>	Built a flexible segmentation process able to accommodate the numerous VM mailings and myriad ad-hoc configurations. <ul style="list-style-type: none">• wrote a flexible import process to efficiently ingest millions of tuples distributed across multiple flat-files, gaining time-savings over the built-in import wizard• achieved efficient joining of local and remote tables via use of <i>OPENQUERY</i>, <i>XML</i>, dynamic <i>SQL</i>• implemented efficient regex parsing via recursion, producing a one-to-many tuple mapping
Volkswagen Onboarding	Worked with .NET developers and project managers to on-board a new client. <ul style="list-style-type: none">• built and tested a new process for segmenting email <i>and</i> SMS from scratch• provided schema specification to developers for data warehousing
TUI Redesign	Collaborated with TUI to integrate a new, responsive design for <i>Thomson</i> and <i>First Choice</i> large deployment broadcasts (5M+ recipients) <ul style="list-style-type: none">• wrote <i>TCL</i> scripts for dynamic <i>HTML</i> merges and gained efficiencies by moving expensive operations upstream• provided testing; gave feedback; managed expectations on technical feasibilities• gained recognition with client and was awarded at the end of this three-month project