

Adrian Ng, MSc.

Seeking Junior-Level Data Engineering Opportunities

Email: contact@adrian.ng **Website:** 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.

Volatility Estimates

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

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 • built pipeline ingesting data from multiple databases, replacing *Informatica* solution
• achieved speed improvements using efficient DML & DDL (*OPENQUERY*, *MERGE*)

Data Cubes Storing analytical datasets in *Data Cubes* achieved
• up-stream computation of all drill-down/roll-up levels and *GROUP BY* permutations
• reduction in size of dataset, minimising bandwidth across distributed servers
• improved user-experience in *Tableau* front-end

Mentoring Dedicating 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
• aimed towards self-sufficiency in writing database queries/stored procedures

GDPR Pipeline
Technical Lead • integrated new GDPR schema into existing datastores (*SQL*, *Salesforce*)
• provide schema specification to *SQL* developers, advocating for indexable data types
• built efficient *MERGE* 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 *logistic regression*.
• contributed to feature selection via: – data extraction – imputation – normalisation – *R* modelling
• researched alternate models (e.g. *Beta-Geometric/Beta-Bernoulli*), 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 Built a flexible segmentation process able to accommodate the numerous VM mailings and myriad ad-hoc configurations.

adrian.ng/SQL/recursion
adrian.ng/openquery-xml

- 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 *OPENQUERY*, *XML*, dynamic *SQL*
- 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.

- built and tested a new process for segmenting email and SMS from scratch
- provided schema specification to developers for data warehousing

TUI Redesign Collaborated with TUI to integrate a new, responsive design for *Thomson* and *First Choice* large deployment broadcasts (5M+ recipients)

- wrote *TCL* scripts for dynamic *HTML* 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

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.