

# 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.

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

- made use of object-oriented techniques and patterns

#### Volatility Estimates

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

- 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.

### 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

<i>Data Analyst</i>	<i>Fan Relationship Management</i>	<i>Jan. - July 2018</i>
<b>NYCFC Data Integration</b> 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	
	<b>Data Pipeline</b>	<ul style="list-style-type: none"><li>• built pipeline ingesting data from multiple databases, replacing <i>Informatica</i> solution</li><li>• achieved speed improvements using efficient DML &amp; DDL (<i>OPENQUERY</i>, <i>MERGE</i>)</li></ul>
	<b>Data Cubes</b>	Storing analytical datasets in <i>Data Cubes</i> achieved <ul style="list-style-type: none"><li>• up-stream computation of all drill-down/roll-up levels and <i>GROUP BY</i> permutations</li><li>• reduction in size of dataset, minimising bandwidth across distributed servers</li><li>• improved user-experience in <i>Tableau</i> front-end</li></ul>
	<b>Mentoring</b>	Dedicating time to mentoring junior colleagues remotely in Manchester/New York <ul style="list-style-type: none"><li>• organised weekly workshops teaching basic DML and advanced DDL</li><li>• developed additional material on my website to supplement these workshops</li><li>• aimed towards self-sufficiency in writing database queries/stored procedures</li></ul>
<b>GDPR Pipeline</b> Technical Lead	<ul style="list-style-type: none"><li>• integrated new GDPR schema into existing datastores (<i>SQL</i>, <i>Salesforce</i>)</li><li>• provide schema specification to <i>SQL</i> developers, advocating for indexable data types</li><li>• built efficient <i>MERGE</i> process featuring relational database design</li><li>• implemented a process to wipe personalised data belonging to any non-consenting individual stored in our data-warehouse</li></ul>	
<b>Customer Churn Model</b>	Modelling MCFC/NYCFC customers' future propensity to churn via <i>logistic regression</i> . <ul style="list-style-type: none"><li>• contributed to feature selection via: – data extraction – imputation – normalisation – <i>R</i> modelling</li><li>• researched alternate models (e.g. <i>Beta-Geometric/Beta-Bernoulli</i>), academic papers, <i>R</i> APIs</li></ul>	

## CREATOR (NOW INSPIRED THINKING GROUP)

<i>Senior CRM Campaign Executive</i>	<i>SQL Development</i>	<i>Dec. 2013 - Sept. 2016</i>
I developed a number of <i>SQL</i> processes to transform customer data into CRM segments. On occasion, I took responsibility for resourcing and managing the team's workload in <i>Jira</i> .		
<b>Virgin Media Segmentation</b> <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"><li>• 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</li><li>• achieved efficient joining of local and remote tables via use of <i>OPENQUERY</i>, <i>XML</i>, dynamic <i>SQL</i></li><li>• implemented efficient regex parsing via recursion, producing a one-to-many tuple mapping</li></ul>	
<b>Volkswagen Onboarding</b>	Worked with .NET developers and project managers to on-board a new client. <ul style="list-style-type: none"><li>• built and tested a new process for segmenting email <i>and</i> SMS from scratch</li><li>• provided schema specification to developers for data warehousing</li></ul>	
<b>TUI Redesign</b>	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"><li>• wrote <i>TCL</i> scripts for dynamic <i>HTML</i> merges and gained efficiencies by moving expensive operations upstream</li><li>• provided testing; gave feedback; managed expectations on technical feasibilities</li><li>• gained recognition with client and was awarded at the end of this three-month project</li></ul>	

## SEATWAVE (NOW TICKETMASTER)

<i>Marketing Analyst Intern</i>	<i>Commercial Team</i>	<i>May 2013 - Dec. 2013</i>
---------------------------------	------------------------	-----------------------------

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.