

Adrian Ng, MSc.

Seeking Junior-Level Data Engineering Opportunities

Email

contact@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*.

Futhermore, with my recent technical accomplishments at *Manchester City FC* I now strive to pursue a career in Data Engineering.

Website

Languages

- Java 8
- T-SQL

Software

- IntelliJ IDE
- SQL Server Ma
- Git
- VS Code
- Jira
- Maven

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*

JAVA PROJECTS

Value at Risk Estimating *VaR*, a measure of risk, for an investment portfolio containing stocks, options, and
Dissertation deltas. Implemented these VaR measures and time-series volatility estimates:

adrian.ng/java/var/

- VaR Measures** • Model Building • Historical Simulation • Monte Carlo Simulation.
- Volatility** • *Equal Weighted* • *EWMA* • *GARCH(1,1)*
- Estimates**
- 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 by 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 **MapReduce** Wrote *MapReduce* applications involving:

Large-Scale

Data Storage &
Processing

adrian.ng/java/enron

adrian.ng/scala/enron1

- 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 single-node/distributed clusters

- Self-hosted the single-node cluster
- ETL via *HDFS* utilised **hdfs** **dfs** commands

Spark Utilising the *Apache Spark REPL* to achieve:

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

Option Implemented numerous approaches to pricing options and calculating payoff:

Pricing

Methods of

Computational
Finance

adrian.ng/java/options/

Summarizing

financial data

adrian.ng/java/yahoofinance/

Options • Monte Carlo Simulation • Black Scholes • Binomial Trees

Payoff • American • Asian • European

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

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

NYCFC Data Integration Integrated New York City FC data into our analytical warehouse. Six-month project involving data discovery, analysis, engineering. Liaised with NY stakeholders. • NYCFC • Ticketmaster Project Owner • 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 data set, 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

• 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 Implemented a new segmentation process in *SQL* stored procedures.

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.

• provided schema specification to developers for data warehousing

TUI Redesign Collaborated 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

• 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.