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

- **Implementation of Value at Risk (VaR) measure**   (https://adrian.ng/java/var/)   (https://github.com/Adrian-Ng/VaR)
  I implemented a number of approaches to estimating *VaR*, a measure of risk, for a hypothetical investment portfolio containing stocks, options, and corresponding deltas. I utilised *Google Fiance/Yahoo Finance* APIs to read time-series data, for which I implemented a number of *moving average processes* for estimating variance.

VaR Measures	**Moving Average Processes**
○ Model Building	○ *Equal Weighted*
○ Historical Simulation	○ *Exponentially Weighted Moving Average (EWMA)*
○ Monte Carlo Simulation.	○ *GARCH(1,1)*

  In addition, an implementation of the *Levenberg-Marquardt* algorithm was used for optimisation of *GARCH(1,1)* parameters via maximum likelihood estimation. I made use of object-oriented techniques and patterns to accommodate these numerous approaches. I used `Java's` concurrency APIs to parallelize the 100,000+ random walks generated by *Monte Carlo* when simulating stock price movements, which resulted in a highly efficient solution.

- **Large-Scale data processing with Apache Hadoop**   (https://github.com/Adrian-Ng/HadoopEnron)
  During my postgraduate module *Large Scale Data Storage and Processing*, I wrote a number of *MapReduce* applications. These projects included: ○ the aggregation of *Twitter* data ○ scraping a large collection of emails in the *Enron Corpus* ○ the extraction of nodes/edges from this communications network. I ran my applications on a self-hosted, single-node cluster as well as on the university's distributed cluster. To load/extract data in *HDFS*, I used `hdfs dfs` commands.

  - ○ **Apache Spark**   (https://adrian.ng/scala/spark/enron1)
    In a self-learning exercise, I translated some of these *MapReduce* applications to `Scala`. This code was less verbose and ran in an *Apache Spark REPL*, which could still interface with *HDFS* via `sparkcontext` APIs.

- **Option Pricing**   (https://adrian.ng/java/options/)   (https://github.com/Adrian-Ng/OptionPricer)
  As part of the postgraduate module, *Methods of Computational Fianance*, I implemented three approaches to estimating option prices: ○ Monte Carlo Simulation ○ Black Scholes ○ Binomial Trees. And where applicable, I computed the *payoff* for American, Asian, and European options. These approaches make probabilistic assumptions, so *Apache Commons Math* API was used.

- **Java 8 Streams with financial data**   (https://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 estimators 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:** I implemented a data pipeline to ingress data from multiple databases. This process was encapsulated in *stored procedures* in which I wrote appropriate DML & DDL (`OPENQUERY`, `MERGE`) for efficient ETL. This pipeline replaced the slower front-end *Informatica* solution.

    - **Data Cubes:** I used an aggregated dataset to compare the distribution of `NULL` values. These analyses were transformed to *Data Cubes* to pre-compute every possible roll-up/drill-down. As such, bandwidth was minimised across our distributed servers and need for real-time computation in *Tableau* front-end was eliminated, resulting in an improved user-experience.

    - **Mentoring:** As part of this project, I dedicated time to mentoring a junior colleague remotely in New York. I organised weekly workshops to teach basic DML and more advanced DDL with a goal toward self-sufficiency in writing database queries and stored procedure implementation/scheduling. Additional material on my website helped supplement these workshops.

- **GDPR Preference Pipeline:** I worked on the integration of a GDPR preference pipeline into our data stores (*SQL*, *Salesforce*) and the subsequent refactoring of numerous processes downstream. I worked with SQL developers to provide specification and UAT testing. I built an efficient, automated `MERGE` process using relational database design (primary key constraints, clustered indexes, triggers).

- **Customer Churn Model:** In an intra-team project to estimate MCFC/NYCFC customers' propensity to churn in following seasons, I assisted with feature selection thereby contributing to data extraction, imputation, and normalisation. Model selection landed on *logistic regression* but I also researched alternate models (involving more probabilistic assumptions) such as *Beta-Geometric/Beta-Bernoulli*; for which I looked at the corresponding academic papers and `R` `Studio` API.

Creator (now Inspired Thinking Group)

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

The majority of my work in this role involved working with `SQL` processes which were used to transform customer data into CRM segmentations. Having been promoted to the senior position, I developed a number of these processes. On occasion, I took responsibility for resourcing and managing the team's workload using *Jira*.

- **Virgin Media Segmentation**   (https://adrian.ng/SQL/cte/Recursion/)   (https://adrian.ng/SQL/misc/openquery-xml) I built an end-to-end segmentation process in `SQL`. This included building a fast, flexible, and bespoke import tool around `BULK INSERT` to efficiently ingest and union millions of tuples distributed across multiple flat-files. Remote server queries (`OPENQUERY`) made use of `XML` to effectively `INNER JOIN` local and remote tables resulting in speed and minimial resource use on a busy live server. I used recursive queries to implement efficient regex parsing similar to `flatMap` in `Java 8`/`Scala`.

- **Volkswagen Onboarding:** I worked with `.NET` developers and project managers to bring Volkswagen on-board as a new client. This required building and testing a new segmentation process for broadcasting email *and* SMS. In addition, I provided specification to developers for their data warehousing/archiving ingress schema.

- **TUI Redesign:** I collaborated with the TUI client integrate a new design of their large deployment broadcasts (5M+ recipients) for *Thomson* and *First Choice* brands. `TCL` scripts were developed to dynamically merge fields into the `HTML` body and, where possible, efficienices were gained by moving expensive operations upstream. Over the course of this three-month project, I gained recognition with the client and was awarded for my efforts. – they sent me an e-card! how best to say this??

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.