

Adrian Ng MSc.

Seeking Junior-Level Data Engineering Opportunities

Email: contact@adrian.ng

Website: adrian.ng

SUMMARY

I am a Computer Science graduate passionate about Data Engineering. I seek opportunities that further my growing experience in *Java* – which I have used in numerous academic projects ranging from the implementation of financial models to large-scale data processing with *Apache Hadoop MapReduce*.

Prior to postgraduate study, my expertise was in *SQL development* focusing on the implementation of CRM segmentation processes for a number of clients including: *Virgin Media*, *TUI*, *UPC*, *MSD*, *Volkswagen*, and *KwikFit*.

In my role as a Data Analyst at *Manchester City FC*, I implemented end-to-end data pipelines (ingress, ETL, data cubes) for use in reporting dashboards. My strengths in this role were more technical than analytical – which leads me now to pursue a career in programming.

EDUCATION

- **Master of Science in Data Science and Analytics** with Distinction
Department of Computer Science, Royal Holloway, University of London Sept. 2016 – Dec. 2017
- **Bachelor of Engineering in Mechanical Engineering** Upper Second Class with Honours
School of Engineering, King's College London, University of London Sept. 2007 – July 2010

JAVA PROJECTS

- **Implementation of Value at Risk (VaR) measures in Java** (<https://adrian.ng/java/var/>) (<https://github.com/Adrian-Ng/VaR>)
This dissertation project implemented various approaches to estimating *VaR*, a measure of risk. These are: *Model Building*, *Historical Simulation*, and *Monte Carlo Simulation*. In addition, the following approaches to estimating market variance/volatility were implemented: *Equal Weighted*, *Exponentially Weighted Moving Average*, and *GARCH(1,1)*.
 - **Object Oriented Design**
As we have a number of approaches to estimating both *VaR*, *variance* and *volatility*, object oriented techniques and patterns were implemented.
 - **Concurrency**
The *Monte Carlo* approach generates a large number of random walks, which can take a long time to fully execute in series. I used Java's concurrency API's to write a highly efficient solution.
 - **Data Ingress**
Real-world market data was sourced using the *Yahoo Finance API*. These daily closing prices were transformed into daily price changes.
Other inputs included the parameters of our hypothetical investment portfolio (market assets, deltas).
- **Option Pricing** (<https://adrian.ng/java/options/>) (<https://github.com/Adrian-Ng/OptionPricer>)
This project implements three approaches to estimating option prices in Java: *Monte Carlo simulation*, *Black-Scholes equations*, and *Binomial Trees*.
Apache Commons Math API was used to deal with some probabilistic assumptions.
- **Data Mining with Hadoop MapReduce** (<https://github.com/Adrian-Ng/HadoopEnron>)
A number of *MapReduce* applications were written in Java with a variety of purposes including extracting the communications network from the *Enron Corpus*, a large dataset of emails, or aggregation of Twitter data. Applications were exported and executed on *Hadoop* clusters (both single node and distributed). Input/Output datasets were stored in HDFS and accessed via `hadoop fs` commands.
A subsequent exercise was undertaken to minimise the verbosity of these *Hadoop MapReduce* applications by translating them to *Scala* for use in a *Spark REPL*.
- **Java 8 Streams with financial data** (<https://adrian.ng/java/yahoofinance/#stream>)
A small exercise involving the use of *Java 8 Streams*. Processing real-world financial data to return *mean* and *equal-weighted variance* of some market asset.
- **Google PageRank**
This is the implementation of Google's *PageRank* algorithm. I simulate the behaviour of someone browsing a series of webpages by computing a transition matrix from an input graph and mixing a Markov Chain.

PROFESSIONAL EXPERIENCE

- **Manchester City Football Club** Euston, London
Data Analyst – Fan Relationship Management *Jan. - July 2018*
 - **New York City FC Project**
A long-term project involving many phases including: data-discovery, pipeline building, data aggregation, and dashboarding.
As part of this project, I spent 8 hours per week mentoring a junior colleague remotely in New York.
 - * **Data Pipeline**
Implemented a data pipeline to ingress data from a number of distributed *SQL* databases. This process was encapsulated in *stored procedures* which used appropriate DML & DDL (*OPENQUERY*, *MERGE*) for efficient ETL. This process replaced the existing *Informatica* solution.
 - * **Data QA**
Data came from multiple external sources (*NYCFC*, *Ticketmaster Salesforce*, *Major League Soccer*). For QA I built an aggregated view to compare the distribution of *NULL* values. I made use of *CUBE* to store these analyses in *Data Cubes* which minimised bandwidth use across distributed databases and improved user experience when visualising these datasets in *Tableau* as every combination of *roll-up* and *drill-down* was pre-computed.
 - **Customer Lifetime Value**
Implementation of a discrete-time periodic sales model (blah blah) Churn, *SQL* data mining. Modeling of LCV via *beta-gaussian*, *beta-binomial*
 - **GDPR Overhaul**
This project involved the creation of a number of automated processes to merge GDPR preference data with the analytical database. To accomplish this, *SQL* DML such as *merge* was utilized.
I worked closely with the development team to test and produce an automated process to merge these new customer preferences seamlessly into the existing database.
 - **Guiding and Mentoring**
Instructing junior colleagues on *SQL* best practices and fundamentals. E.g. understanding DDL & DML for writing *SQL* queries and creating database objects; when to return a *product join* vs *semi-join*; making use of *information_schema*; utilizing *SQL Agent* to schedule jobs.
- **ITG Creator (Digital Marketing Agency)** Westminster, London
Senior CRM Campaign Executive – SQL Development *Dec. 2013 - Sept. 2016*
 - **Virgin Media Segmentation Process**
Built an end-to-end segmentation process (data ingress + config file = segmentation data). Close coordination with account executives Innovative use of *XML* + dynamic *SQL* + *OPENQUERY* resulted in efficient data fetching from remote server.
 - **Volkswagen New Client Onboarding**
Work assessment group Testing Mobile and Email CRM Efficient approach using recursive query for regex
 - **TUI Content Redesign**
Three-month project. Close liaising with client. Integration of new *HTML* from client into existing system. Ensuring design is reactive. Integrating *SQL* table with *HTML* content dynamically using robust *TCL* scripts to handle multiple design configurations.
Efforts on this project were recognised by client.
 - **Soft Skills**
 - Attended inter-departmental work assessment groups and advised on work specifications.
 - As senior team member, served as point of contact for clients and colleagues looking to resource our team.
 - On occasion I held responsibility for resourcing and managing the team's workload using *Jira*.
- **Seatwave (now Ticketmaster)** Moorgate, London
Marketing Analyst Intern – Commercial Team *May 2013 - Dec. 2013*
 - **Basic SQL**
In this position I gained my first experience writing database queries in *SQL Server Management Studio*. With basic understanding of *DML* and *DDL*, I was able to query the ticketing and customer databases to extract data for warehousing, analysis, and CRM segmentation.

TECHNOLOGIES

- **Languages:** Java 8, T-SQL
- **Software:** IntelliJ IDEA, SQL Server Management Studio,, Git, Jira, Maven