

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* – a language 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*, and *KwikFit*.

In my role as a Data Analyst as *Manchester City FC* I implemented data pipelines, data cubes, and other technical miscellany note - write something better here. Furthermore, my abilities here proved to be more technical than analytical – which now leads me 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 implements 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
Jan. - July 2018
 - *Data Analyst – Fan Relationship Management*
 - **New York City FC Project**
Built a data pipeline to ingest data from *Ticketmaster* and *NYCFC*.
Liaised with stakeholders (New York City FC) and partners (Major League Soccer).
 - **Lifetime Customer Value**
Churn, SQL data mining. Modeling of LCV via *beta-gaussian*, *beta-binomial*
 - **GDPR Customer Preferences**
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.
 - **Tableau Dashboard Automation/Optimization**
Implemented *Data Cubes* to pre-aggregate data along all combinations of categorical fields. That is, every possible drill-down and roll-up was computed in advance. As a result, front-end dashboards retained their exploratory flexibility but removed real-time computational burden. Thus improving user-experience.
 - **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
Dec. 2013 - Sept. 2016
 - *Senior CRM Campaign Executive – SQL Development*
 - **Segmentation Processes**
Built a number of automated segmentation process using SQL stored procedures for team members to use.
Recipient data were imported via **BULK INSERT**, stored in database tables and indexed (clustered). Segmentation data was output and linked to HTML content to be broadcast to recipients.
 - **Recursion** (<https://adrian.ng/SQL/cte/Recursion/>)
Used recursive queries (CTEs) to clean data e.g. removing *n*-number of leading zeros from mobile phone numbers in order to prefix with dialling codes; or splitting strings and mapping into relational format.
 - **Cross-Server Query Optimisation** (<https://adrian.ng/SQL/misc/openquery-xml>)
Improved cross-server query execution speeds by using **OPENQUERY**, which transmits a string of SQL for execution on the remote (a live database under constant heavy load). Further, *Dynamic SQL* was utilized to include XML data in the string. Mapping via a *CTE*, this XML could be transformed into a relational object capable of joining to remote objects. As a result, filtering via join occurs remotely and only a small data set is returned via the **OPENQUERY**.
 - **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
May 2013 - Dec. 2013
 - *Marketing Analyst Intern – Commercial Team*
 - **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