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 Data Engineering and I seek opportunities that meet my growing experience in *Java* – a language I have used in numerous academic projects ranging from the implementation of models pulling real-world financial data to large-scale data processing with *Apache Hadoop MapReduce.*

Prior to postgraduate study, my expertise was in *SQL development* focusing on the implementation of segmentation processes for a number of clients including: *Virgin Media*, *TUI*, *UPC*, *MSD*, *Volkswagen*, *KwikFit*. After graduation, my most recent accomplishments as a Data Analyst at *Manchester City FC* were in the technical parts (e.g. pipelines, architecture) of the projects I worked on, which leads me now to pursue a career in Data Engineering.

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)
  Assuming a number of hypothetical investment portfolios, my dissertation project implemented a number of approaches to estimating *VaR*, a measure of risk, and variance/volatility (for model parameterization).

    ○ **VaR:** a) *Model Building* b) *Historical Simulation* c) *Monte Carlo Simulation.*
    ○ **Variance/Volatility:** a) *Equal Weighted* b) *Exponentially Weighted Moving Average* c) *GARCH(1,1).*

  Because of these numerous approaches, object-oriented techniques and patterns were implemented. In addition, I used Java's concurrency APIs to parallelize the 100,000+ random walks generated by *Monte Carlo* for simulating stock prices. Real world financial data was obtained via the Google Finance/Yahoo Finance APIs.

- **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.

Professional Experience

- **Manchester City Football Club**   Euston, London
  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 a number of remote *SQL* databases. This process was encapsulated in *stored procedures* which used 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 improved user-experience.
    - **Mentoring:** As part of this project, I was dedicated 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 working with stored procedures. Additional material on my website helped supplement these workshops.
  - **GDPR Stream Integration:** I worked on the integration of a GDPR preference stream into our data stores (*SQL, Salesforce*). I implemented a new pipeline and refactored numerous processes downstream . I worked with the development team to provided specification and UAT testing. I built an efficient, automated `MERGE` process using primary key constraints, clustered indexes, triggers.
  - **Customer Churn Model:** I contributed datasets and collaborated on feature/model selection. In particular, looking at *logistic regression* and *Beta-Geometric/Beta-Bernoulli* models in R Studio.

- **ITG Creator (Digital Marketing Agency)**   Westminster, London
  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. As senior team member, I developed a number of these processes. On occasion, I held responsibility for resourcing and managing the team's workload using *Jira*.
  - **Virgin Media Segmentation Process**   (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`. 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. Recursive queries were used to implement a solution (similar to `flatMap` in *Java 8*) for efficient *regex*.
  - **Volkswagen Client Onboarding:** I worked with `.NET` developers and project managers to bring Volkswagen on-board as a new client. This required implementing a new segmentation process for broadcasting email *and* SMS. In addition, I provided specification to developers for their data warehousing/archiving ingress schema.
  - **TUI Mailing Redesign:** I collaborated closely with the TUI client during a three-month project to redesign the existing *Thomson* and *First Choice* mailings. `TCL` scripting was used to merge dynamic content into the `HTML` body. My efforts on this project were recognised by the client.

- **Seatwave (now Ticketmaster)**   Moorgate, London
  Marketing Analyst Intern – Commercial Team   *May 2013 - Dec. 2013*
  In this position I gained my first experience writing database queries in *SQL Server Management Studio*. I wrote *DML* and *DDL* capable of querying the transactional/customer databases to return data for warehousing, reporting, and segmentation.

Technologies

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