

Adrian Ng, MSc.

Junior Metadata Engineer at IAG Cargo

Email: contact@adrian.ng **Portfolio:** adrian.ng **Location:** Harrow

PROFILE

*NB. I am **not** job-hunting. May this document simply serve to give you an idea of what I'm about professionally.*

I am a *Junior Metadata Engineer* at IAG Cargo, an airline based at Heathrow. I have just started my job here but I will be building data pipelines and working with big-data technologies.

Content to keep this section rather sparse for now...

- Adrian

EDUCATION

Royal Holloway 2017
Data Science & Analytics

- Master of Science
- Pass with Distinction
- Department of Computer Science

King's College London 2010
Mechanical Engineering

- Bachelor of Engineering
- Upper Second Class with Honours
- School of Engineering

CODING

- T-SQL, PLSQL
- Java 8
- Powershell

TOOLS

- VS Code
- Git
- Markdown
- Jira

IAG CARGO

Junior Metadata Engineer

Data Platform Team

June 2019 - Present

Oracle ETL Automation

Using *Powershell* scripts and modules (*Sharepoint PnP*, *Credential Manager*), this process pulls files from *Sharepoint*, invoking *SQL*Loader* for bulk import into *Oracle* and *sqlcl* for calling stored procedures and executing DML.

Sales Reference db Created a relational schema in *Oracle* with *ETL* processes and auditing.

BA Internal Mentoring Programme

Mentored a colleague based in Spain over a 9 month period.

moving-ahead.org

MANCHESTER CITY FOOTBALL CLUB

Data Analyst

Fan Relationship Management

Jan. - July 2018

NYCFC ETL Project Owner

Integrated New York City FC data into our analytical warehouse. Six-month project covering data discovery, analysis, engineering. Multiple data sources were involved: • NYCFC • Ticketmaster • 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 dataset, minimising bandwidth across distributed servers
- improved user-experience in *Tableau* front-end

training

Dedicating time to training 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 Technical Lead

- integrated new GDPR schema into existing datastores (*SQL*, *Salesforce*)
- provide schema specification to *SQL* developers, advocating for indexable data types
- built efficient *MERGE* process featuring relational database design
- implemented a process to wipe personalised data belonging to any non-consenting individual stored in our data-warehouse

Customer Churn Model	Modelling MCFC/NYCFC customers' future propensity to churn via <i>logistic regression</i> . <ul style="list-style-type: none"> contributed to feature selection via: – data extraction – imputation – normalisation – R modelling researched alternate models (e.g. <i>Beta-Geometric/Beta-Bernoulli</i>), academic papers, R APIs
-----------------------------	--

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	Built a flexible segmentation process able to accommodate the numerous VM mailings and myriad ad-hoc configurations.
----------------------------------	--

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.
------------------------------	--

- implemented a new process for segmenting email *and* SMS from scratch
- provided schema specification to developers for data warehousing

TUI Redesign	Collaborated with TUI to integrate a new, responsive design for <i>Thomson</i> and <i>First Choice</i> large deployment broadcasts (5M+ recipients)
---------------------	---

- wrote TCL optimised scripts for merging HTML documents in SQL tables.
- tested, gave feedback, managed expectations on technical feasibilites
- gained recognition with client and was awarded at the end of this three-month project

JAVA PROJECTS

Value at Risk Dissertation	Estimating <i>Var</i> , a measure of risk, for an investment portfolio containing stocks, options, deltas.
-----------------------------------	--

adrian.ng/java/var/

Var Measures

- Model Building
- Historical Simulation
- Monte Carlo Simulation.

Volatility Estimates

- Equal Weighted*
- Equal Weighted Moving Average (EWMA)*
- GARCH(1,1)*

- implemented *Levenberg-Marquardt* algorithm for optimisation of *GARCH(1,1)* parameters
- made use of object-oriented techniques and patterns
- gained efficiencies 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 Large-Scale Data Storage & Processing
--

adrian.ng/java/enron
adrian.ng/scala/enron1

MapReduce	Wrote <i>MapReduce</i> applications involving:
------------------	--

- aggregation of *Twitter* data, utilising **twitter4j** API
- scraping a large collection of emails in the *Enron Corpus*
- extraction of communications graph consisting of nodes/edges

Hadoop

- Applications ran on both single-node (self-hosted)/ distributed-node clusters
- Interfaced with *HDFS* via terminal command-line

Spark

Utilised an *Apache Spark REPL* to achieve:

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

Option Pricing Methods of Computational Finance
--

adrian.ng/java/options/

Implemented numerous approaches to pricing options and calculating payoff:

Options

- Monte Carlo Simulation • Black Scholes • Binomial Trees

Payoff • American • Asian • European

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

Summarizing financial data

adrian.ng/java/yahoofinance/

An exercise in using 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.