

## OVERVIEW

I graduated last year with a mathematics degree and was looking for an industry that my skills and background would transfer well to. I discovered Big Data when looking at programming languages and found that my preference for pure mathematics related well as the defining characteristics of Big Data require more of an abstract approach when creating robust solutions. I am now extremely excited about the opportunity to work as a big data engineer.

## WORK

### BIG DATA ENGINEER – KUBRICK GROUP, LONDON

At Kubrick Group I gained experience and qualified in: Professional and Communication Skills, IOC (Investment Operations Certificate) Level 1, Advanced Excel, Web Application development (UI/UX, JavaScript, CSS and HTML), Agile and SQL.

I then specialised in the Big Data Engineering course leading to accreditation as a Cloudera CCA Developer.

#### August 2016 - present

I worked on Big Data projects which exercised my skills in data ingestion and processing using Hadoop EcoSystem tools such as Sqoop, Flume and OpenSource Python Libraries including Pandas and Twython. We cleansed and munged data after exploratory data analysis through Python libraries including Plotly and Pandas then moving onto Spark to process, transform, analyse and stage the data into HDFS.

Projects included:

MetOffice Weather Data: involving Python API and Twitter processing into HDFS followed by further processing in the Hadoop Ecosystem.

Police Crime Data Analysis involving SSIS processing, SQL Analysis, and presentation through custom web front end.

### TECHNOLOGY TRAINING – GETUM LTD, SHREWSBURY

#### September 2015 – January 2016

I gained experience and accelerated my skills by learning about and seeing how the company used different technologies, including: Scala, Play, Github, Agile, Cassandra and Spark.

## EDUCATION

### UNIVERSITY OF EAST ANGLIA – MATHEMATICS BSC, 2:1 HONS

#### October 2012 – June 2015

I was part of a group project in second year which involved solving a stream function of point vortices as part of my fluid dynamics module. The solution required using a Runge-Kutta numerical method and I presented the project to my year group and several senior lecturers.

### ADDITIONAL TECHNOLOGY SKILLS

I've used Scala to create a simple login system which is based on user information stored in a Cassandra database. I've also been working through the Functional Programming Principles in Scala course as a way to advance my knowledge of Scala.

## KEY SKILLS

#### Python Libraries (Anaconda):

PyMongo, Pandas, Numpy, scikit learn, Plotly and Twython

#### Cloudera Hadoop version 5.4:

HDFS, Yarn, Sqoop, Flume, Avro, Parquet, Impala, Hive and Hue

#### NoSQL:

MongoDB. API with PyMongo

#### SQL:

SSIS, Advanced T-SQL and Stored Procedures

#### Apache Spark:

RDD, DataFrames, SparkSQL, PySpark with some exposure to and understanding of Scala

#### Source Control and IDE:

Git, SourceTree and Git Extensions, PyCharm

#### Other Skills:

Linux CentOS 6, SSH, Putty, Agile, JSON and BSON, Python Pickle and Shelve



**GEORGE HUGHES**

Big Data Engineer

## ABOUT ME

I am highly logical, hard-working, ambitious and intelligent with a can-do attitude and a drive to solve problems. I have a natural attention to detail and thrive when presented with new obstacles to overcome. I am looking to work for a company that can challenge and engage me as well as allow me to put my skills to the test.

I discovered Big Data while working with Getum last year and soon realised it was a rapidly growing industry and is quickly becoming vital to companies in several different industries. It definitely is an area I wanted to learn more about and I am determined to become an expert in this field.

## CONTACT US

Kubrick Group

T: 020 3866 4620

E: [consultants@kubrickgroup.com](mailto:consultants@kubrickgroup.com)

W: [kubrickgroup.com](http://kubrickgroup.com)