

OVERVIEW

I chose to study Biology to satisfy my interest in the natural world and my desire to use data to understand it better. Throughout my course I found my passion for data and statistics when reading scientific papers and learning about new ways to analyse and present information. Subsequently this led me to the decision that a career in Big Data was what I wanted to pursue. As a Big Data Engineer I get the opportunity to work with raw data and model it into useful and game changing information which allows businesses to make more informed decisions and distinguish themselves from others.

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.

WEBSITE DESIGN AND MAINTENANCE – CANTERBURY CAR CENTRE LTD

June 2015 – September 2015

Ensuring that new stock was placed on company and associated websites and using coral draw to create professional stock photos. I was also responsible for designing new layouts for the website itself.

EDUCATION

UNIVERSITY OF NOTTINGHAM, BSC BIOLOGY, 2:1

The largest project I undertook at university was a human genetics project looking at gene inheritance. In order to do this, I had to be very detail-oriented when setting up the DNA mixtures, as well as when analysing the output through a software called GeneViewer. I then used R to test the significance of the results gained. Overall, I feel that this project cultivated my skills of self-discipline and focus as well as the more technological skill.

ADDITIONAL TECHNOLOGY SKILLS

R was the main language utilised throughout my undergraduate degree. One project undertaken for my Experimental Design and Analysis module used R to evaluate p values and to produce GLMs based on this data.

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

ABIGAIL DUSKE

Big Data Engineer

ABOUT ME

Coming from a STEM background it is clear to me that data is at the heart of everything, and Big Data is at the heart of company decision making. My curiosity and confidence ensure that I always want to get to the root of the matter and subsequently, Big Data is the path for me.

I am highly eloquent and an effective communicator. I am keen to learn new skills, am a confident networker and assertive in decision making in my personal life as well as in a work environment.

Currently my personal development involves taking Brazilian Jiu-Jitsu classes and playing competitive quidditch.

CONTACT US

Kubrick Group

T: 020 3866 4620

E: consultants@kubrickgroup.com

W: kubrickgroup.com