Big Data Engineer

Anton Todorov

# Contact

|  |  |
| --- | --- |
| Address: | 25 Wills Street, Melbourne, 3000 |
| Phone: | 04 2374 2568 |
| Email: | anton@openfab.com.au |

# Experience

## ERM Power

|  |  |
| --- | --- |
| Job Description: | Senior Data Engineer |
| Dates: | 2018 Jul - On going |

Worked on transition to AWS and Linux from local Microsoft SQL server setup. Leveraged Redshift and Snowflake for data lake, EMR for distributed compute using PySpark/Hadoop.

Deigned an implemented local Vagrant/Packer/Ansible based development environment using infrastructure as code and mirroring production environment build scripts. Build CI/CD pipelines using AWS code\* for Python based projects.

Developed forward projection models for usage of AWS resources and tied in with automated orchestration engine for cost reduction using Python/SciKitLearn.

Developed serverless AWS Lambda based application for ingesting batch power consummation and pricing data based on power network tariffs.

## Telstra

|  |  |
| --- | --- |
| Job Description: | Big Data Engineer |
| Dates: | 2017 Jan - 2018 Jul |

Worked with PySpark/Spark on big data project to extract real time information about health of the Telstra network and provide information used by other teams in minimising disruptions.

Developed data ingestion pipelines for real time data from multiple data brokers using Flume and Kafka.

Fed data from pipelines into HDFS for temporary storage before sending the information as batch jobs to Spark/Scala using Hive.

Adapted Java based work flow for Python in Jenkins in new CD/CI environment adapted to AWS deployment.

Worked on porting legacy Scala code into performant PySpark where possible, optimised PySpark code for execution in Spark/Scala. Wrote thin python wrappers (using PyPy) for memory heavy operations (mostly to do with date conversions).

## ITG

|  |  |
| --- | --- |
| Job Description: | Data Scientist |
| Dates: | 2015 May - 2016 Dec |

Developed machine learning based trade risk analyser, used as static classifier to decide if a trade would be a business or regulatory risk and flag it for further high touch handling. Used new data lake solution with Apache Hive and Hadoop, connected 15+ >1TB older databases, to train model in batch mode (up to once a day).

Worked on integration of historic disparate data sources (20+ databases) into big data lake using using Hive and HDFS. Used Spark for real time trading orders alternative execution routes. Fed data into python application using PySpark.

Developed in house crisis impact analysis/resolution tool using hypergraph/simplex algorithms/data structures (using Haskell), integrated with in house provisioning system to minimise trade disruptions (using Python).

Used Django/Bootstrap/d3.js/Pandoc to deliver live documentation to company stakeholders. Automated reports of business recovery gaps and worked with various business units to improve key recovery times.

## OpenFab

|  |  |
| --- | --- |
| Job Description: | Co-founder |
| Dates: | 2013 Nov - 2015 May |

Worked on using Linux real time kernels with hard real time requirements.

Developed firmware for open hardware laser cutter in C/assembly.

Developed numerical simulation suite to enable machine learning gradient based descent methods for firmware motor control algorithms using repeated simulation.

## National Oilwell Varco

|  |  |
| --- | --- |
| Job Description: | Numerical Programmer/Developer |
| Dates: | Jul 2011 - Nov 2013 |

Numerical programmer on a computational fluid dynamics project used to model turbulent drilling fluid flow in oil well. Worked on numerical integration for stiff differential equations implemented in Fortran.

Developed a web app for National Oilwell Varco/Noble Denton rig projects to automate documentation generation. Web app used Python framework (Django) and custom back end (Tesseract, Django templating engine, TeX and html) with a simple front end to upload raw text and images and produce reports in pdf and html.

# Skills

## Languages

* Python
* Bash
* C
* SQL
* Lisp
* Scala
* Haskell
* R

## General Skills

* Machine learning
* Numerical programming
* Could deployment (AWS)
* Sysadmin
* Developer
* Mathematics
* Floating point experience

# Education and qualifications

## Degrees

### BSc Physics

|  |  |
| --- | --- |
| Institution: | Melbourne University |
| Dates: | Feb 2006 - Oct 2010 |

### International Baccalaureate

|  |  |
| --- | --- |
| Institution: | Chatsworth International School |
| Dates: | Sep 2004 - Jun 2006 |