

About Me

Delivery oriented Engineer specialising in data engineering, Dev[Sec]Ops, and AWS solutions. Skilled in managing and working within large-scale multi-workstream projects. Passionate about Big Data and ML/AI integration.



Core Skills

While I have a track record of adapting and learning quickly “on-the-job” my core skills include but are not limited to:

- Team leadership and motivation through Scrum and Agile Practices.
- Ability to communicate with all levels of the hierarchy both technical and non-technical.
- Programming Languages: Python, JavaScript, TypeScript, Ruby
- Data Engineering and ETL: AWS Glue, Athena, Lambda, ECS; Python Bonobo, JSONata, JSON Schema, Pydantic, Camunda Zeebe...
- Data Analysis: Jupyter, Dask, AWS SDK for pandas (formerly AWS Data Wrangler)...
- Knowledge Graph: JSON-LD, dgraph, Neo4j
- Infrastructure as Code (IaC): Terraform, OpenTofu, Cloudformation, AWS CDK
- Container Technologies: Kubernetes, Helm, Docker...
- Git and CI/CD: Bitbucket, Github, Gitlab, Github Actions, Jenkins...

Career Summary

Sky UK Ltd - Contract (June 2023 - August 2023)

Joining an Equal Experts team deployed to shadow a squad within Sky Identity I provided consultancy on DevOps practice with a particular focus on security; on a stack consisting of Springboot Java applications built using a highly custom Jenkins pipeline and deployed to Kubernetes with manifests being generated via a custom Gradle plugin.

Key Responsibilities

- Engaging with DevOps and SRE colleagues to assess the “State of DevOps” in Sky Identity.
- Assist in accelerating the “path to production” for new microservices.
- Improve the level of security in the Continuous Integration pipeline.

Key Achievements

- Added SonarQube and Veracode security scanning to the CI pipeline as a quality gate resulting in the number of security and code quality issues reaching production reducing to near zero.
- Introduced cookiecutter Java project which slashed new project first iteration delivery to production from 3 months to a matter of days.
- Simplified the CI pipeline by removing 3 unnecessary deployed environments meaning:
 - deployment to production reduced from 2 weeks to 1 day and;
 - cost of running additional Kubernetes namespaces was removed.
- Consulted on the practice of Canary deployments resulting in fewer false negatives and subsequent rollbacks.

AstraZeneca UK Ltd - Contract (October 2019 - March 2023)

Spending 3.5 years with AstraZeneca I provided my services as a certified AWS Solution Architect and Python Engineer to build a number of Platforms from their initial Extract, Transform, Load (ETL) phase right up to onboarding and delivering value to other parts of the business. During that time I engaged with Cloud and Security teams within AstraZeneca to establish new ways of working in the cloud. Additionally, I had the privilege of coaching a number of other engineers in Agile Practice, Python and; Cloud Architecture and Engineering.

UNIFIED MEMORY PROJECT (January 2022 - March 2023)

As Tech Lead on this project, I led in the technical analysis, redesign and build of an ETL solution for ingesting unstructured data using NLP to extract key data attributes.

Key Responsibilities

- Lead in the design and development of a new service collaborating with Solution Architecture, Enterprise Search, Reference Data Management and AI Engineering colleagues.
- Advise on resourcing requirements.
- Communicate technical requirements to both technical and non-technical colleagues.
- Champion and demonstrate good practices in development, DevOps, Solution Architecture and Cloud Engineering.

Key Achievements

- Ensured good development practice by ensuring Unit Test coverage remained over 90% throughout
- Applied automation to the development lifecycle where possible using GitHub Actions
- Ensured standard development practice across the team by sharing IDE configurations. In this case Visual Studio Code
- Adapted to resource constraints within teams the project was dependant on by actively filling knowledge gaps
- Managed the flow of technical deliverables across 5 work streams to deliver milestones on time.
- Prevented over-spending on head count by reducing requirement and saving the equivalent of 3 FTE over a period of over 6 months.

FEATURE TEAM ENABLEMENT (July 2021 - December 2021)

The purpose of this team was to provide support to the 4 work streams across the Data Platform in the design, development and deployment of AWS based solutions.

Key Responsibilities

- Establish an AWS estate of 14 accounts to a level at which over 200 engineers are able to deploy solutions.
- Establish CI/CD pipelines and standards to facilitate the deployment of infrastructure and services.
- Champion and coach teams in good practices when developing infrastructure and service code.
- Advise Platform Leadership.

Key Achievements

- Successfully introduced the standard practice of using Terraform for defining infrastructure.
- Introduced a Proof of Concept for cataloguing projects in GitHub using Backstage (<https://backstage.io>).
- Designed and built a solution for automating Installation Qualification (IQ) to expedite GxP compliance verification.
- Developed `cookiecutter` templates to provision projects using Backstage and reduce project spin up time
- Mentored apprentice engineers in good practice DevOps and Cloud Engineering.

PATIENT SAFETY DATA INGESTION (January 2021 - June 2021)

I joined this team to take a proof-of-concept Veeva ETL solution to production. In doing so I established good unit testing practices and a full CI/CD deployment pipeline. In addition, I provided on-the-job coaching to 2 Kubrick colleagues on python best practices (OOP, Unit testing and coding standards such as style and secure coding) and AWS.

Key Achievements

- Built a fully unit tested ingestion pipeline
- Provided sufficient coaching for both Kubrick colleagues to be offered permanent Data Engineer positions
- Instigated a wholesale shift to Data Platforms own AWS estate saving time and cost of delivery
- Mentored apprentice engineers in the good practice Data and Python Engineering.

CPSS IMAGING DATABASE (50% July 2020 - December 2020)

The CPSS Imaging Database is an image data and metadata management system for digital pathology, in-vivo, and mass spectrometry imaging designed around a BPMN engine (Camunda Zeebe). Its purpose is to gather metadata from services internally - such as eSlide Manager, bioELN - and externally, such as BSI

- before combining these data with metadata extracted from image files (in various formats: SVS, NDPI, bif, DICOM), transforming it using JSONata, and storing in Dgraph.

Key Achievements

- Lead on the design and building of infrastructure in AWS CDK for deploying more than 15 services to AWS Fargate. Services included: Camunda Zeebe; Camunda Operate; Elasticsearch & Kibana; Dgraph
- Implemented persistent storage for containers running in AWS Fargate using AWS EFS
- Designed and implemented inter-service connectivity by applying Service Discovery using Cloud Map and on-prem connectivity using a combination of internal Application and Network Load balancers
- Designed and implemented hub-and-spoke CI/CD pipeline using Jenkins Pipeline, samver
- Assumed Scrum Master role. Facilitated daily Stand-up events and regular Backlog Refinement events. Key Skills: Team Leadership, Mentoring, AWS Solution Architecture, AWS CDK, CI/CD, Jenkins, TypeScript, Python

SCIENCE DATA FOUNDATION IMAGING PLATFORM (50% October 2019 - December 2020)

The SDF Imaging Platform serves as the “one stop shop” data lake for biomedical images and related metadata at AstraZeneca. It provides the ability to discover and access images relevant to a particular avenue of research.

Key Achievements

- Lead on design and delivery of the SDF Imaging Platform release 1 using AWS cloud services. Taking an existing PoC and producing a cloud native solution.
- Develop nano-services to perform extraction of medical images in DICOM format and Whole Slide Images in a range of formats including SVS and NDPI
- Designed Step Function state machine to orchestrate activities on images and their metadata based on source and format and eventual destination
- Designed an API for providing access to the complex and diverse imaging metadata using AppSync (GraphQL) and DynamoDB in addition to supplying pub/sub capabilities using SNS
- Implemented dashboard in CloudWatch to monitor image processing activity and throughput
- Implemented custom alerting based on application log output using custom metrics, CloudWatch Alarms and SNS More recently in this role:
- Developed a cross-platform CLI utility written in Python 3 for obtaining AWS STS keys when authenticated via Ping Federate SAML AWS integration and assuming a Federated Role to replace the distribution of IAM Keys
- Instrumental in the Solution Architecture of an API that enables:
 - Multiple imaging workstreams to contribute to a universal database of imaging metadata
- -Free text searching of image metadata
 - More complex querying of metadata using GraphQL
- Lead on the design and delivery of a solution for an internal HTTP API to enable Validation, Transformation and Governance of imaging metadata using API Gateway (Private), VPC Interface Endpoint
- Liaised with MuleSoft architecture and development to establish integration with the internal API
- Lead on the design and delivery of a solution for deploying distributed graph database, Dgraph using AWS Fargate

HALO UPLIFT PROJECT (50% October 2019 - June 2020)

The Halo Uplift Project was a Proof of Concept to answer the questions of “Can we migrate on-prem Halo to the Cloud?” and “How much will it cost to deploy and run?”

Key Achievements

- Lead on the design and delivery of configurable and reusable (dev, test etc..) infrastructure using AWS CDK to deploy scalable clustered EC2 instances of varying size and a MySQL Database cluster using AWS RDS
- Created Ansible playbooks to provision the EC2 instances and install: Halo Clients; Halo AI; Halo Link; Halo API; Halo Analysis
- Lead a Proof of Concept to assess the possibility of using AWS AppStream for end-user access management and governance
- Liaised with Data Scientists on system testing
- Liaised with Product Owner, Business Analyst and Indica Labs on solution costing and viability

Other Relevant Roles

yboo Ltd - Contract (October 2018 - August 2019) Skills: AWS Services and Tools, in particular: AWS CDK, API Gateway, IAM, Lambda, S3, Kinesis Data Streams and Firehose, Redshift, Ruby, NodeJS, TDD (RSpec, Jest), Docker, Bitbucket Cloud (including Bitbucket Pipelines), DevOps
The Dining Club Group - Contract (May 2018 - August 2018) Skills: AWS Services and Tools, in particular: Terraform, API Gateway, IAM, Lambda, S3, DynamoDB, DeepLens, Amplify, PHP, Go, TDD (PHPUnit, testing.T), MySQL, Docker, Jenkins, DevOps
William Hill Plc - Contract (March 2016 - February 2018) Skills: Team Leadership, Mentoring, Python, Go, TDD (unittest.py, testing.T), Splunk, Docker, API Integration, Ruby, JavaScript, TDD (RSpec, Jest), Ruby on Rails, Postgresql, Docker, Marathon/Mesos, Jenkins, DevOps, Scrum
IntechnologyWIFI - Contract (October 2015 - January 2016) Skills: Python, PHP, MySQL, API integration
Vodafone UK - Contract (May 2015 - August 2015) Skills: Python, Django, Scrum
LED Hut Ltd (June 2014 - November 2014) Skills: PHP, Magento, Laravel, SysAdmin, MySQL
Wilson and Cooke (November 2013 - June 2014) Skills: PHP, WordPress, SysAdmin, MySQL
Docnet Ltd (January 2013 - November 2013) Skills: PHP, Zend Framework, SysAdmin, MySQL
Begbies Traynor Group Plc (February 2012 - January 2013) Skills: PHP, Python, Django, Symphony, Doctrine, MySQL
Zimo Communications Ltd (June 2011 - February 2012) Skills: PHP, Zend Framework, MySQL
Manchester Metropolitan University (April 2007 - June 2011) Skills: PHP, CakePHP, MySQL, SysAdmin

Past/Present Certifications

	Expiry
AWS Certified Solutions Architect - Associate	Oct 2018 - Oct 2021
Certified ScrumMaster (CSM)	Feb 2018 - Feb 2022

Hobbies/Interests

Sports

Having grown up in Moston, Manchester I have followed in my fathers footsteps and am a Manchester City fan. Since University I've enjoyed snowboarding and try to go every Winter; best place I've boarded so far... Whistler, BC in Canada!

Computing

I'm keen to get more involved in IoT and ML. I enjoy tinkering with a Arduino and a Jetson Nano whenever I have spare time. Currently reading "Deep Learning with PyTorch". I also enjoy playing the odd game; football obviously but also taking my children through the games I played at their age such as Day of the Tentacle and Monkey Island.

Volunteering

With any spare time I have I happily support local micro-businesses with my technical knowledge with services including but not limited to advise and/or management of DNS, static websites, email and Shopify. I find it rewarding to see the stress drain away when I offer to help and the relief when I say "for free!".