Callum Squire

West Hampstead
London
England
✉ *ctsquire@hotmail.co.uk*

Employment History

Data Engineer - Mercuria Energy Group Ltd, *January 2023 -> Current role*, London.
Building highly secure, greenfield projects alongside the Head of Data and BI. I set up the modern data solutions requested. Responsibilities include:

- Developed a Python poetry template with GitLab CI/CD, streamlining the linting, typing, testing, and uploading of Python modules from GitLab to Artifactory using tools like Black, Mypy, and Pytest. Enhanced automation, contributing notably to the quality and efficiency of our code deployment process.
- Created and administered our virtual data warehouse using Denodo, maintaining and expanding over 50 Kubernetes Manifest files. Collaborated closely with vendor support to troubleshoot and report bugs. A Gitlab CI/CD template was provided which used Ansible to template the files using Jinja2. This was done in order to publish to multiple environments. The files were output to a Gitops Gitlab repository. ArgoCD was used to autosync the cluster with the Kubernetes manifest files. This included identifying the required databases for each environment. I recommended and we used Kerberos for our authentication when using JDBC to connect to the SQL Server databases. Authentication to use Denodo is managed per application using folders and active directory groups we helped set up. I built a Gitlab CI/CD VQL test harness to run queries in Gitlab pipelines.
- Innovated a Python Data Access Library that utilizes Denodo metadata to dynamically construct python classes for data models and readers for each view in Denodo. Used secure connections leveraging SQLAlchemy, with credentials for applications securely managed via Hashicorp Vault. I used pydantic, Jinja2 and dataclasses to build the code from templates using metadata from Denodo stored procedures. This is executed as a submodule and removed when uploading to Artifactory. This allows business users to easily connect to Denodo views with python without needing to install any external drivers.
- Setting up Dagster, a python oriented, ETL orchestration tool. I maintain and extend the helm charts and build any extra Kubernetes manifest files. This includes Hashicorp Vault external secrets and pod cleanups for runs. The provided Gitlab CI/CD template is used to output the manifest files to be synced by ArgoCD.
- Provided the Dagster Python Utilities library functions to connect to any AWS/MinIO S3 bucket using an S3 Client. We also provided functions to push source files into parquet files in a S3 bucket which could be executed after picking up files using an S3 or SFTP Dagster sensor. I provided functionality to add the modified date from the S3 file as a column which could be used for partitioning and also the ability to filter files in the source directory using regex.
- Built example Dagster Gitlab repositories which were published to docker images in Harbor using the provided Gitlab CI/CD scripts. The Docker images are used by the Dagster system where master branch merges create images to be used for the development environment and tags are used for UAT and production environments. The task involved adding Dagster assets around business code that used Denodo views to query a database and transform the data and load the Json response to an API endpoint which updated a database. This involved working closely with the business user and the provider of the API during testing.
- Built the AWS to MinIO S3 syncs Dagster Gitlab repository which uses a shell script and a Dagster resource to run MinIO CLI commands to mirror the S3 buckets from AWS cloud to on-premise MinIO. I also provided examples of using the SFTP sensor to download files to S3 source bucket. Once the data is in MinIO source, I use the Dagster Utils library to push the files into parquet format in our 'gold' standard bucket. The files are then automatically created as Dremio views using the APIs and Denodo views are created from these views using the Dremio JDBC driver via a deployment process we created for new Denodo views.
- Built the Qlik Sense Service library, published to Artifactory, which creates python functions for interacting with the Qlik Sense Repository and Engine APIs. I built a Qlik Sense report which used incremental refresh to provide example to the Qlik developers.
- Led weekly training sessions for the business users to learn how to use the systems I built. This involved documenting heavily on Confluence and onboarding and mentoring a new team member.

IT | BI Developer - Ladbrokes Coral, *December 2019 -> December 2022*, London.
Working in a big data environment, implementing data solutions for the Finance, Compliance and Trading teams. Responsibilities include:

- Built and maintained a SQL Server Analysis Services solution which hosts billions of rows of data. SQL Server Integration Services and SQL Server is used to monitor and refresh the database. Logging tables are used to monitor the processes and reasons for failures. The SSAS database is used by the Compliance, Retail Finance (UK and ROI) and Retail Trading teams to extract insight from the Retail betting activity and interactions within shops. The server is configured to use the Entain SMTP relay to send notifications and data check emails through C# scripts.
- Designed and created a Python tool which currently executes 34 SQL scripts daily and uploads them to SharePoint. The compliance team was doing lots of manual work extracting data from Power BI so I built this to free up resources. The Python script first checks data availability and moves scripts to the end which don't have data available. The files are GPG encrypted before uploading to SharePoint. Emails and a Power BI report have been created so the compliance team can monitor the availability of the files.
- Maintained the architecture that allows the movement of data from source using the Airflow server to the Hive Hadoop data solution. This involved creating and maintaining shell scripts working on the Linux servers. The shell scripts use sqlworkbench and I manage the Drivers and Profiles used. The BI team use these scripts to build DAGs that migrated data from other areas of the business to the data lake. Trino sits on top of hive to read from the ORC files. Ladbrokes Retail use GCP hosted Presto and Entain, the parent company, use on-premise Trino.
- Created Python packages using the Google My Business and Microsoft Power BI REST APIs. This automated the process of updating shop information, triggering Power BI refreshes and building pipelines for usage, meta, and refresh data. The PowerBIService package is also used by Entain who copied the Retail architecture.

Business Intelligence Developer - Xenomorph, *January 2019 -> December 2019*, London.
Working in a small company, implementing solutions with Power BI for multiple clients within the finance industry. Responsibilities include:

- Lead developer working with HPS, mentoring 2 graduate analysts to move data from SFTP into a object oriented data warehouse, through to the Power BI Service. This involved understanding and confirming client requirements and then leading and providing guidance for the data analysts as well as building the complex Power BI reports.
- On-site Power BI training with Helaba and HPS. The clients learnt how to use and create new Power BI reports.
- Creating a new azure tenant for the external clients.

Data Analyst - Calastone, *October 2017 -> January 2019*, London.
Developing a pipeline of BI solutions on the Calastone Data Services Product, within the Mutual Funds Industry. Responsibilities include:

- Created the analytical reports using Microsoft Power BI including implementing Row Level Security.
- Using Thomson Reuter Lipper feed with Python.

Control Systems Designer - UK Quantum Technology Hub Sensors and Timing, *June 2016 -> August 2016*, University of Birmingham.

Skills

Programming Languages: Python, Shell, PowerShell, PowerQuery;
Cloud Services: AWS S3, Azure AD, Power BI Service, GCP (BigQuery, Server hosting);
Data Engineering: Airflow, Dagster, SQL Server, SSIS, SSAS, Teradata, MinIO S3, Dremio, HDFS, Hive, Presto, Trino, ORC, Parquet, Kafka, APIs (REST, gRPC), Vim;
Visualization: Power BI, Qlik Sense, Dash; **Authentication:** Kerberos, JWT, GPG, PGP, OAuth 2.0, Basic;
DevOps: Kubernetes, Docker with Harbor, Gitlab CI/CD, Ansible, Artifactory, ArgoCD, Grafana;
Project Management: JIRA, Confluence, YouTrack, Trello, Excel;
Degree: 2013 – 2017 Msci Physics, University of Birmingham, Final Grade: Class II (Division I) 69%

About

Confident and self-motivated person who enjoys working in agile environments. Loves taking advantage of all opportunities available and takes great enjoyment from overcoming challenges. Looking to further a career as the Data Architect and seek out new opportunities to develop knowledge and experience. Sociable and enjoys working with others to develop strong relationships with colleagues.