

# Charlie Collier

Email: [collier.p.charlie@gmail.com](mailto:collier.p.charlie@gmail.com)

GitHub: <https://github.com/collier-p-charlie>

Website: <https://www.cpcollier.com>

## Skills

Expertise in **Python**, **SQL** and **Infrastructure as Code** as well as **Airflow**, **Bash**, **Testing**, **dbt**, **Databases**, **Docker**, **APIs** and **Spark**. I like to design solutions (from networking and permissions through to app development) and then implement with best practices. In particular, setting up the permissions / networking, architecting the pipelines, optimising costs and deploying with **IaC** tools.

## Work Experience

**Data Engineer**, Hard Rock Digital (Remote)

January 2025 to present

- **Optimising and Refactoring Suboptimal Processes**

- Refactored and optimised various **Airflow DAGs** to adhere to best practices, improving the efficiency of **DAG** parsing and debugging.
- Started the creation of documentation in **DBT Cloud**, utilising unit and data tests for audit purposes, as well as **Snowflake**.
- Designed and built out in **Terraform** the new gambling self-exclusion pipeline, creating a new data model for efficiency.
- Created several **CI/CD** pipelines for deployment of this pipeline, as well as testing including tools such as **mypy**.
- Bootstrapped the **Astronomer** (Airflow) environments for our *regulatory* team, creating utility operators for *logging* and *notifications*.
- Created a template for *unit testing* including mocking, as well as introducing data validation tools to the team.

**Senior Cloud Engineer**, MarketCast

November 2023 to January 2025

- **Designing and Implementing the Survey Database Solution**

- As part of the *data architecture* team, I scoped, designed and implemented our new data warehouse solution for survey data.
- The current *file-based* architecture was limiting use cases and our ability to do further lookbacks for historical normative calculations.
- Utilised **AWS CDK/CloudFormation** for the deployment (including several event-based **Lambdas** and **ECS** for processing the surveys).
- Used **Redshift** for the warehouse solution and **dbt** for modelling **ELT** layers within the database, orchestrated through **AWS MWAA**.
- Developed a table design to preserve the metadata in the survey **SPSS/SAV** files and for ease of metadata filtering.
- Built a generalised **API** with **API Gateway**, **Route53** and **FastAPI** for database access and aggregation calculations for dashboards.
- Implemented proper logging and alerting via **SNS**, **CloudWatch** and **SQS** for error catching / retries between services.
- Integrating the database with data science **ML APIs** to enrich the survey data.
- Created a standard for *networking* across the data architecture team, which was approved by our **AWS** account solutions architects.
- Collaborated with teams in the US and India to incorporate this new survey database into the backend of our new product offerings.
- Utilising **Redshift** as our data warehouse has allowed us to calculate normative values going back years, not just months, in seconds.
- It has allowed us to consolidate duplicate products which all had similar architectures, reducing **AWS** costs by thousands per month.
- Involved **AWS** in discussions as part of **MAP** program to help ensure best practice and design patterns for this architecture.

**Data Engineer / Scientist**, MarketCast

April 2021 to November 2023

- **Automating and Optimising our Attribution Engine**

- Singularly designed and implemented full **end-to-end system** for Excel report generation from client requests in **JIRA**.
- Utilised **GPUs** for nearest-neighbour **ML** algorithm, **CTAS** for **Athena** query optimisations and **VPC** optimisations (**NAT** and endpoints).
- Optimised various **SQL** queries to reduce data scan and execution time and migrated **Redshift** queries to **Athena**.
- Used **memory profiling** (memray) to track memory usage of jobs to gather training data for an **ML** model to predict memory usage.
- The above has significantly improved the reliability of client deliverables, reduced their delivery time and freed up resources.
- Moreover, we were able to reduce the cost of running a report down to a few hundred pounds from multiple thousand pounds.

- **Cost Savings, Technical Support and Implementation of Best Practices**

- Working with the SVP of Technical Architecture to help reduce our organisations **AWS** account costs.
- Point of contact for **AWS IAM** permission issues, **BitBucket** access and **Okta** through having admin access.
- Defined best practices for code development, unit testing, **IaC** and **CI/CD** deployments across the data science team.
- Implemented the **principle of least privilege** to roles in our production accounts.
- Conducted knowledge sharing sessions for upskilling the team, based on specific gaps in knowledge.
- Optimised a sub-optimal **ETL** pipeline using **Spark** on **Glue**, reducing the costs by 25% and execution time from hours to minutes.

## Certifications

AWS Certified SysOps Administrator - Associate

Awarded March 2024

AWS Certified Data Engineer - Associate

Awarded February 2024

AWS Certified Developer - Associate

Awarded January 2024

AWS Certified Solutions Architect - Associate

Awarded August 2023

AWS Certified Cloud Practitioner

Awarded February 2023

## Education

**Masters in Mathematics (MMath)**, University of Bath

September 2015 to July 2019

Awarded first class honours.