# Collin Brown

|  |  |  |  |
| --- | --- | --- | --- |
| github | [github.com](https://github.com/collinbrown95) | linkedin | [linkedin.com](https://ca.linkedin.com/in/collin-brown-499a4580) |

I do DevOps and Kubernetes platform development for a Data Science team in a large organization. I am passionate about bringing data-driven projects into production in an enterprise setting using modern open source tools.

## Experience

### Lead Software Developer · (Jun 2022 - Present)

#### Statistics Canada · Data Science Division

In this position, I have three primary responsibilities: (1) Kubernetes platform development for our organization’s analytics environment (the [Advanced Analytics Workspace](https://statcan.github.io/daaas/en/)); (2) putting Data Science projects into production; and (3) Providing horizontal technical advisory/consulting services to various employees in the organization. I elaborate on each area of responsibility below.

**Kubernetes Platform Development**

**S3 object storage for platform users**

* Created an S3 application to provide platform users with an S3 interface by combining two open source tools: [aws-js-explorer](https://github.com/StatCan/aws-js-s3-explorer) (client-side) and [S3Proxy](https://github.com/gaul/s3proxy) (server-side).
* Wrote several middleware components to handle networking concerns such as a custom service worker to intercept and modify requests from the browser and an Istio Virtual Service to apply service mesh routing logic.
* Wrote helm chart templates and kubernetes manifests to automatically deploy this S3 application to multiple users’ namespaces with ArgoCD.

**Cloud Environment network connectivity for employees**

* Wrote a kubernetes controller to create several Istio resources to handle routing of certain requests to an Istio Egress Gateway.
* Wrote ArgoCD deployment to deploy abovementioned controller alongside other required Kubernetes resources.
* Modified Terraform configuration to provision necessary cloud resources for this feature such as an Azure Kubernetes Service (AKS) node pool, Azure subnet, and firewall rules.

**Add specialized node pools to the kubernetes cluster**

* Wrote Terraform configuration to add specialized node pools to our Kubernetes cluster such as Azure’s [Fsv2-series](https://docs.microsoft.com/en-us/azure/virtual-machines/fsv2-series) VMs for CPU intensive workloads and [NCv3-series](https://docs.microsoft.com/en-us/azure/virtual-machines/ncv3-series) VMs for data science workloads that require multiple GPUs.
* Wrote logic in a mutating webhook to dynamically configure pod tolerations so that specific users have their pods scheduled to the correct specialized node pool.

**Putting Data Science Projects into Production**

**Custom PDF Document Information Extraction Application**

* Led an effort to deploy a custom-built PDF extraction application into our organization’s internal Kubernetes production environment.
* Wrote a helm chart to deploy several components such as a deployment for the custom web application, several cron jobs that perform actions such as an SFTP pull and data ETL, Elasticsearch Kubernetes Operator, Kibana Kubernetes Operator, and [S3Proxy](https://github.com/gaul/s3proxy).
* Wrote terraform configuration to provision resources from Azure such as Storage Accounts and Storage Containers to be used by the application.
* Onboarded the original developers of the application to the Kubernetes environment

**Horizontal Technical Advisory**

**Cloud-Native Data Processing System Redesign**

* Provided technical advisory services to a project team engaging in a cloud-native redesign of a data processing system.
* Provided recommendations around user experience, data processing workflows, code organization, and version control.

**DevOps / Data Engineering Centre of Expertise (CoE)**

* Chaired weekly meetings to provide advisory services related to DevOps and Data Engineering.
* Advised Data Science project teams on topics such as product deployments, software development design considerations, and cloud infrastructure decisions.

### Senior Software Developer · (May 2021 - Jun 2022)

#### Statistics Canada · Data Science Division

In this position, I was acting as the senior software developer on an 18-month cloud-native redesign of the data processing system for a large statistical program Statistics Canada. From May 2021 to July 2021, I worked directly under the project technical lead, and from August 2021 to March 31, 2022 I acted as the technical lead for the project. The project was delivered successfully (feature-complete, on-time, and at a cost significantly lower than the next best alternative) on March 31, 2022. Listed below were my primary duties in this position.

**Supervise the work of other developers**

* Supervised the work of 6 different developers (not all at the same time), where each developer focused on implementing a different aspect of the system.
* Provided technical advisory for the project’s developers on issues related to software design, code implementation, debugging, testing, and documentation.
* Chaired daily stand-up meetings with the developers, the projects’ clients, and subject matter experts to ensure that the project’s deliverables were on track.
* Managed bi-weekly project milestones and project Gitlab issues, ensuring that project goals were met according to specifications in a timely manner.

**Design and implement several cloud-native components of the system redesign**

* Designed and Implemented a Helm Chart, which specified the deployment of the system including components such as Argo Workflows, a custom web application service, MinIO (running in Gateway mode to Azure blob storage), Elasticsearch, and Kibana.
* Wrote infrastructure as code using Terraform modules to provision cloud resources through the Azure provider, such as Azure Storage Accounts and Storage Containers.
* Designed and Implemented a Gitlab CI/CD pipeline that builds the project’s main Docker image, performs a security scan for known CVEs, runs our automated test suite, builds and publishes the project’s documentation site, and pushes the Docker image to a private image registry.
* Designed the release management strategy for our project that synchronizes source code and infrastructure changes with our deployment environments so that development, test, and production environments can be updated independently.

### Software Developer · (Oct 2020 - May 2021)

#### Statistics Canada · Data Science Division

In this position, I was a core developer on an 18-month cloud-native redesign of the data processing system for a large statistical program at Statistics Canada. I list below several core features and items I implemented for this project.

* Designed and implemented a framework to organize a significant amount of business logic required for data processing and validation across dozens of different data sources.
* Designed and implemented a framework to apply unit testing and integration testing to the business logic of the system.
* Designed and implemented other aspects of the system such as the data storage layer, the user interface layer, and the workflow orchestration (data pipeline) component.

In addition, I helped various Data Scientists and Data Analysts with Data Engineering and Software Development implementations through direct help and presentations to larger audiences.

* Delivered a presentation to ~50 Data Scientists on how to use Data Version Control (DVC) - a tool to synchronize large file versioning with the commit history of a Git repository.
* Guided Data Scientists and Data Analysts to onboard their project source code and documentation to Gitlab and showed them how to apply good version control practices.
* Consulted various Data Scientists on how to write application code to interface with certain platform services (e.g. Elasticsearch, S3-Compatible Object Storage, etc.)

### Data Scientist / Software Developer (Aug 2018 - Oct 2020)

#### Employment and Social Development Canada · Data Science Division

In this role, I have worked on both conventional Data Science projects (e.g. data processing, machine learning, etc.) and web development for the purpose of prototyping and creating proof-of-concept products that are used to consume the outputs of Data Science projects.

**Web Development**

* Led the development of several proof-of-concept web applications to alleviate the administrative burden associated with common tasks.
* Created programs to automate the processing of data (e.g. a set of scripts that automatically convert raw XML data into a formatted and populated Microsoft Excel workbook)
* Handled small-scale deployments of applications for prototyping and testing purposes, leveraging containerization (Docker), cloud (Microsoft Azure), and version control (Git/GitHub) technologies
* Created and contributed to various sources of instructions and software documentation

**Data Science and Machine Learning**

* Wrote several reports with literature reviews on various methodologies (e.g. how machine learning techniques can be used in ad-hoc information retrieval)
* Quickly tested and experimented with different machine learning techniques to assess project feasibility (e.g. applying pre-trained word embedding models to problems involving free-text data)
* Wrote programs to independently evaluate the performance of other teams’ machine learning models
* Tested and experimented with ways that user interaction could be recorded to implicitly label data (e.g. created a simple client-side javascript library to anonymously track how users interact with the predictions of machine learning models).

## Skills, Tools, and Technology

**DevOps**

Kubernetes Development · Packacing Kubernetes Applications (e.g. Helm, Kustomize, Jsonnet) · Writing Custom Kubernetes Controllers (Golang)

**CI/CD**

ArgoCD · Gitlab CI · Github Actions · Bash Scripting

**GitOps and Cloud**

Terraform · Azure · S3 Object Storage · Azure Blob Storage · Istio · Managed Databases

**Web**

Python · JavaScript · React · Vue · d3.js · Flask · FastAPI · ORM (e.g. SQLAlchemy) · SQL Databases (e.g. SQLite, Postgres) · Document Oriented Databases (e.g. Elasticsearch)

**Docs**

Static site generators (e.g. mkdocs) · Writing Software Documentation · Implementing Coding and Docstring Standards in Projects (e.g. Google Python Style Guide)

**Data Processing**

Pandas · Numpy · Implementing Scalable Data Processing Jobs on Kubernetes with Argo Workflows · Schema Management and Data Validation (e.g. Pandera, Pydantic)

**Data Science**

Text Data Analysis (e.g. spaCy) · Web Scraping (Scrapy, Beautiful Soup) · Deep Learning Frameworks (e.g. Tensorflow, Pytorch) · Deep Learning Theory (e.g. sequence models, convolutional neural networks)

**Economics and Stats**

Stata · MATLAB · R · Econometric Modeling (Regression Analysis, Panel Data Methods) · Time Series Econometrics · Empirical Microeconomics · Research Design · Experimental Economics

**General**

Public Speaking · Leadership · Consulting · Communication · Academic and Non-Academic Writing · Presentations

**Languages**

English (fluent) · French (beginner/intermediate)

## Open Source Software Contributions

Doccano

[Doccano](https://github.com/doccano/doccano) is an open source text annotation tool to facilitate data labeling for machine learning projects. On this project, I implemented internationalization on user interface, miscellaneous small feature requests, and bug fixes.

## Research Publications

Economics

**January 2022**: [What Drives Bitcoin Fees? Using Segwit to Assess Bitcoin’s Long-run Sustainability](https://www.risk.net/journal-of-financial-market-infrastructures/7914886/what-drives-bitcoin-fees-using-segwit-to-assess-bitcoins-long-run-sustainability) · Published in the Journal of Financial Market Infrastructures

## Education

### Undergraduate Courses, Computer Science Core Curriculum (part-time)

**Athabasca University · 4.00/4.00 cGPA**

Taking the core courses in the undergraduate Computer Science curriculum to supplement my existing education and experience.

**Completed / In Progress Courses**: Object Oriented Programming · Data Structures and Algorithms · Operating Systems · Design and Analysis of Algorithms · Computer Networks

**Remaining Courses (Expected)**: Computer and Network Security · Database Management Systems · Distributed Systems

### Master’s Degree, Economics (Focus in Econometrics)

**Queen’s University · 4.17/4.30 cGPA**

**Courses**: Microeconomic Theory, Macroeconomic Theory, Quantitative Methods (Econometrics), Industrial Organization I, Empirical Microeconomics, Time Series Econometrics, Public Economics

**Thesis**: [What Drives Bitcoin Fees? Using Segwit to Assess Bitcoin’s Long-run](https://www.econ.queensu.ca/research/working-papers/1423)

**Scholarships and Awards**: Canada Graduate Scholarship · Tri-Agency Recipient Recognition Award

### Bachelor’s Degree (Honours), Economics (Minor in Mathematics and Statistics)

**McMaster University · 3.96/4.00 cGPA**

**Courses**: Advanced Economic Theory I & II · Advanced Econometrics · Statistical Methods and Applications · Game Theory · Mathematical Economics · Calculus I–III · Linear Algebra · Directed Research in Econometrics I & II

**Thesis**: Using Dynamic Probability Models to Estimate the Effect that Historical Tax Credit Claims Have on Current Period Claims

**Scholarships and Awards**: Maclean Family Academic Grant · Thompson Academic Grant · Chancellor’s Gold Medal Finalist · Hugh Clark Scholarship · Kneale Brothers’ 37 Academic Grant · Provost’s Honour Roll Medal (Third Year) · Undergraduate Student Research Award · McMaster Spectrum Competition (First Place in Social Innovation Category) · Dr. Lyman Hooker Scholarship · McMaster Entrance Scholarship · University Senate Scholarship