# 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.

## Professional 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 of these responsibilities below.

**1. Kubernetes platform development**

**Object storage for platform users**

* Created an opt-in s3 application to provide platform users with an s3 interface to Azure Blob Storage 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).
* Decreased the amount of custom code required to provide s3 object storage to our users, reducing implementation complexity and maintenance burden.
* Wrote several middleware components to handle networking concerns such as a 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.

**Network connectivity to internal services for authorized users**

* Created a feature to allow authorized users network access to important internal services, unblocking hundreds of authorized employees who require access to these services.
* Wrote a Kubernetes controller to create several Istio resources to handle routing of certain requests to an Istio egress gateway.
* Wrote Kubernetes manifests 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, an Azure subnet, and Azure firewall rules.

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

* Unblocked several high priority data science projects by enabling Data Scientists to access machines with sufficient compute resources.
* 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 deep learning 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.

**Deemed employee access control**

* Unblocked a Government department by enabling the department’s employees to access to certain critical features on our platform.
* Wrote a Kubernetes controller to implement capability-based access control by applying labels to certain Kubernetes resources.

**2. Putting data science projects into production**

**Custom PDF Document Information Extraction Application**

* Led an effort to deploy a large custom-built PDF table extraction application into our organization’s internal Kubernetes environment.
* Enabled dozens of employees to greatly speed up their work by searching an indexed database of extracted PDF information rather than searching for the information manually.
* Wrote a Helm chart to deploy several components such as a deployment for the custom web application, several cron jobs, Elasticsearch and Kibana Kubernetes Operators, and [s3Proxy](https://github.com/gaul/s3proxy) backed by Azure Blob Storage.
* Wrote Terraform configuration to provision resources required by the application from Azure such as Storage Accounts and Storage Containers.
* Onboarded the original application developers to our organization’s internal Kubernetes environment

**3. Horizontal Technical Advisory**

**Cloud-Native Redesign of a Large Data Processing System**

* Provided technical advisory services to a project team engaging in a cloud-native redesign of a data processing system for a large statistical program at Statistics Canada.
* The redesigned system significantly reduces comptue and storage costs, human error, and maintenance burden.
* Provided technical recommendations around developer experience, data processing workflows, code organization, and version control.

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

* Chaired weekly meetings to provide technical advisory services related to DevOps and Data Engineering.
* Unblocked multiple data science projects by providing recommendations around product deployments, software design, and cloud infrastructure decisions.

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

#### Statistics Canada · Data Science Division

In this position, I was 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, 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**

* Wrote a Helm chart, which specified the deployment of the system including components such as Argo Workflows, a custom web application, 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 Storage Accounts, Storage Containers, Object Replication Policies, and Storage Management Policies.
* Wrote 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 multiple Data Scientists with various Data Engineering and Software Development issues through both direct help and presentations to larger audiences. I list below several technical support services I provided in this position.

* 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 to onboard their project source code and documentation to Gitlab and showed them how to apply version control best practices.
* Consulted various Data Scientists about how to write application code to interface with various platform services (e.g. Elasticsearch, s3 object storage)

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

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

In this role, I worked on both conventional data science projects (e.g. data processing, machine learning) and web development for the purpose of prototyping and creating proof-of-concept data science products.

**Web Development**

* Led the development of several proof-of-concept web applications to alleviate the administrative burden associated with common government tasks.
* Created programs to automate the processing of complex data into clean usable data artifacts (e.g. a CLI program that converts raw XML files into formatted and populated Microsoft Excel workbooks).
* 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 interactions could be recorded to implicitly label training data (e.g. created a simple client-side javascript library to anonymously track how users interact with the predictions of machine learning models in a web application).

## Skills, Tools, and Technology

**DevOps/Cloud**

Kubernetes development · packaging Kubernetes applications (e.g. Helm, Kustomize, Jsonnet) · writing Kubernetes controllers (Golang) · building Docker images · Istio service mesh · shell scripting · ArgoCD · Gitlab CI · Github Actions · Terraform · Azure · s3 object storage · Azure Blob Storage

**Web Development**

Python · JavaScript · React · Vue · d3.js · Flask · FastAPI · ORM (e.g. SQLAlchemy) · SQL databases (e.g. SQLite, Postgres) · document oriented databases (e.g. Elasticsearch)

**Documentation**

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 · Argo Workflows · schema management and data validation in Python (e.g. Pandera, Pydantic)

**Data Science/Machine Learning**

text data analysis (e.g. spaCy) · web scraping (Scrapy, Beautiful Soup) · deep learning frameworks (e.g. Tensorflow) · deep learning theory (e.g. sequence models, convolutional neural networks)

**Statistics and Economics**

Stata · MATLAB · R · econometric modeling (e.g. regression analysis, panel data methods) · time series econometrics · empirical microeconomics · research design

**General**

public speaking · leadership · consulting · communication · academic and non-academic writing · presentations

**Languages**

English (fluent) · French (beginner/intermediate)

## Professional Honours and Awards

### Branch Award - Most Innovative Project · October 2022

#### Statistics Canada · Modern Statistical Methods and Data Science Branch

Received the annual branch award for the most innovative project. I was given this award for leading the team that delivered the organization’s first successful cloud-native redesign of a large scale data processing system that is being used in monthly data production at Statistics Canada.

## 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 Coursework, Computer Science (part-time) · (2020 - Present)

**Athabasca University · 4.00/4.00 cGPA**

Taking core courses in the undergraduate Computer Science curriculum in my spare time.

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

### 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 Sustainability](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