Caylin D. **Hickey**

Data Scientist · Cloud Engineer

1095 Veterans Drive, HSRB154, Lexington, KY USA

(859) 323-8255 | caylin.hickey@uky.edu | cayl.in | tsuyoi | caylin-hickey | @caylin_hickey

Experience

University of Kentucky

Lexington, Kentucky USA

Cloud Engineer - Clinical Laboratory

Jul. 2018 - Present

Oversee the development, management, and operation of the internally developed cloud-based genomics data processing pipeline system, Genomic Pipeline Management System (GPMS) used in the processing of both clinical and research genomic samples from an on-site sequencer and processed using cloud resources. Involved in the establishment of various efforts related to digital pathology and pathology informatics. Additionally develop or contribute to multiple projects related to efforts in the Clinical Lab, Department of Pathology, Division of Pathology Informatics, and other interdisciplinary collaborations.

- Upgraded many aspects of the Genomic Processing Pipeline Management System (GPMS):
  - Incorporate BagIt specification and TAR archiving encapsulation methods to verifiably archive and move data pieces to, within, and from the cloud environment
  - Refactor processing agents to be cloud-identity aware and report this to the central management system for automated deployment and termination of worker units
  - Augment the GPMS controller with the ability to spin up and down cloud-based computing instances to be more cost efficient
  - Provide an email-based notification system to alert of errors and notify of progress step completion
  - Interface with diverse modules to be run by the processing pipeline developed by colleagues in various departments
- Digital pathology efforts
  - Participate in the onboarding and workflow development around a new Philips Ultra Fast Scanner (UFS) and Image Management System with efforts related to upgrading digital barcoding and improving labeling until the establishment of a new EHR
  - Aid in the administration of a feasibility study on using web-based whole-slide image (WSI) services for diagnosis
  - Research into white-blood cell (WBC) counting and ratio calculation in specific tissue types for aiding in diagnosis
- Pathology informatics efforts
  - Create multiple applications to streamline efforts in the sign-out process for clinical genomics cases
- Clinical lab efforts
  - Upgrade email-based report processing script to work across multiple users to improve delivery efficiency while correcting some processing errors and including processing/delivery notification to ensure functionality
- Medical education efforts
  - Develop an electronic attendance tracking system leveraging existing magnetic stripe ID cards and remote Raspberry Pi machines to allow for the digital tracking of CME credits and seminar attendance in the department
- Interdisciplinary efforts
  - Aid in the development of a pharmacogenomic report and data presentation phone application to aid in informing those who have undergone pharmacogenomic testing as to what further considerations and conversions would be useful with their primary care physician (PCP)

University of Kentucky

Lexington, Kentucky USA

Data Scientist - Department of Pathology and Laboratory Medicine

Nov. 2017 - Jul. 2018

Responsible for research, development, and operations of a complex agent-based distributed systems used in the monitoring and measurement of national and international research networks. Leveraging both internal (UK) and external data sources, develop complex models of data for analysis. Responsible for statistical analysis on large datasets to identify important relationships and trends, and then summarizing and communicating findings to stakeholders.

- Researched the effects of using NVDIMMs in conjunction with SSDs as a cache tier to speed up the performance on a 2.5PB Ceph storage cluster being constructed as part of KyRIC for the Major Research Infrastructure grant
- Developed graph-based web IDE for composing Cresco CADL pipeline files for use in deploying IRNC measurement tasks
- Refined development pipeline dashboard plugin for Cresco framework as well as unifying theme with newly developed demonstration site
- Initial work on unit testing of low-level Cresco framework systems
- Researched containerization of open-source Next-Generation Sequencing tools into a Docker container toolkit designed to run pipelines from YAML configuration files
- Researched image segmentation used in nuclear density of digital pathology images of bone and soft tissue slides using deep learning and fully-convolutional networks

University of Kentucky

Lexington, Kentucky USA

Data Management Specialist, Sr. / Intern Manager - Research Computing

Jun. 2015 - Nov. 2017

Aided in the development of genomic-processing pipeline distribution system by leveraging background in informatics to efficiently organize distributed sequence/sample processing, collect performance metrics of processing machines, and analyze/display overall system health, status, and processing progress. Aided in deployment, monitoring, and administration of various cloud computing and storage clusters utilized in research on campus.

- Co-Developed distributed genomics processing control system leveraging Cresco edge computing framework and background in informatics related to organization of distributed workflows as well as information representation and summarization
- Developed automated account deployment system for research computing and storage infrastructure
- Directed intern work involving upgrading and refining computing infrastructure monitoring systems as well as research efforts in migrating active Docker containers from one system to another
- Leveraged open source software stacks relating time-series data and data analytics to build base storage and computing infrastructure monitoring system
- Researched performance comparisons between various clustered filesystems such as CephFS, GlusterFS, GPFS, etc. to investigate possible upgrade of existing storage cluster
- Deployed production-ready OpenStack cluster used by researchers requiring fixed computing resources for their research projects
- Deployed 3 Petabyte geo-replicated, production-ready Object (S3) storage cluster built on the Ceph open source software
- Aided in testing and management of existing OpenStack testbed to provide resources to researchers and IT department groups
- Worked on IRNC distributed network monitoring grant project to implement RESTful interface, and generic modules as well as general code maintenance
- Continued development of GENI distributed monitoring system:
  - Separated functionality into separate agents on distributed cluster
  - Added analysis of incoming data through pattern recognition queries
  - Added report generation system to collate and summarize historical data
  - Continued to refine overall system architecture to improve collection stability and presentation

University of Kentucky

Lexington, Kentucky USA

Data Management Specialist - UKIT

Jun. 2013 - Jun. 2015

Utilized knowledge of distributed systems and background in informatics to develop, operate, and maintain monitoring cluster in order to provide insight into the utilization, status, and operation of the GENI (Global Environment for Network Innovations) distributed research testbed cloud infrastructure. Extended analytics capabilities of high-performance computing utilization system to incorporate costs and project grant awards with associated cluster usage.

- Developed distributed system and usage monitoring Collector for GENI networking research project designed to:
  - Collect computational and network utilization data from distributed sources
  - Group, analyze, and summarize current information
  - Store and quickly retrieve historical records for troubleshooting
- Investigated using the Apache Hadoop implementation of the Big Table paradigm to solve large-scale data problems in a distributed cluster computing environment
- Developed High-Performance Computing usage analytics reporting web service in conjuncture with researcher management system
- Co-Developed attendance policy compliance system prototype using mobile web and location data compiled from distributed networking data
- Explored virtualization through the OpenStack platform coupled with the JuJu rapid deployment architecture
- Leveraged Play Framework for rapid development/deployment of Java Web Services

University of Kentucky

Lexington, Kentucky USA

Web Development Intern - Academic Technology Group

Oct. 2011 - Jun. 2013

Employed knowledge in general informatics techniques to gather, organize, summarize, and analyze researcher utilization of the DLX high-performance computing cluster as well as researched techniques in data science used in building a text machine learning classification system from existing taxonomy data. Aided in testing and learning of open-source cloud infrastructure software Open-Stack.

- Programmed high-performance computing (HPC) researcher management website to collect and present usage statistics for distributed HPC cluster
- Investigated and implemented strategies for exploring using machine learning techniques and categorization informatics to classify new data into pre-existing categories using existing data in a supervised learning system deployed on a distributed computing platform
- Developed generalized graphical data flow management GUI using web technologies
- Implemented JavaScript libraries into existing web systems
- Assisted in the overhaul of several internal websites
- Documented usage manuals for several internal websites

Commonwealth Tool & Machine, Inc. *Stamping Ground, Kentucky USA*

Workflow Engineer - Engineering Department *Apr. 2010 - Sept. 2010*

Worked to improve distributed materials informatics systems, efficiency automation systems, as well as provide some IT consultation related to storage devices and professional-grade graphics cards used by the engineering department. Continued to refine private cloud storage system design.

- Expanded development of cost matrix processing program to work on distributed data sources in order to reduce time used by the engineers and purchasing department
- Continued development of SolidWorks part schematic extraction program to improve performance and functionality
- Continued to refine private cloud storage system design
- Adjustments to the prototype website

Commonwealth Tool & Machine, Inc. *Stamping Ground, Kentucky USA*

Workflow Engineer - Engineering Department *Jun. 2005 - Jul. 2008*

Developed efficient programs for collecting, managing, summarizing, and reporting of project material requirements for the purchasing department from lists extracted and compiled from engineering specifications and part schematics on distributed systems. Aided in design and deployment of migration to private cloud storage for centralized engineering file management. Responsible for research, development, and maintenance of automation operations utilized by the engineering and purchasing departments.

- Collaborated with the purchasing department to develop cost matrix processing program to analyze project electronic bills of material and populate material ordering quotes and purchase orders
- Developed and maintained automated standard and isometric drawing extracting add-on for SolidWorks designed to streamline part schematic delivery from the engineering department to the machining department
- Migrated and improved previously developed printing automation program to be a SolidWorks add-in programmed in C# along with improved functionality and more accurate performance
- Aided in design and deployment of migration to private cloud storage for centralized engineering file management
- Developed prototype website for the company using ASP

Commonwealth Tool & Machine, Inc. *Stamping Ground, Kentucky USA*

Workflow Engineering Intern - Engineering Department *Aug. 2002 - Jun. 2003*

Developed task repetition automation program for the engineering department to speed up printing of part schematic files manually generated by engineers for manufacturing. Also assisted IT department with routine management and consulted on computing and storage expansions.

- Developed program to automate the printing of part file schematics using Visual Basic
- Assisted in the troubleshooting and minor hardware and software repairs of the computers in the engineering department

Education

University of Kentucky *Lexington, Kentucky USA*

Pursuing Ph.D in Computer Science *Jan. 2015 - Present*

University of Kentucky *Lexington, Kentucky USA*

B.S. in Computer Science with a Minor in Mathematics *May 2013*

Skills

Technology	Informatics, Distributed Computing, Edge Computing, Cloud Computing, Cloud Storage, Software-Defined Networking, Containerization, Virtualization, Machine Learning, Deep Learning, Reinforcement Learning
System Administration	Linux, Unix, Ceph, OpenStack, Ubuntu, OS X, Windows, Docker
Programming	Java, Python, C/C++, C#, R, Perl, Objective-C, Scala, Node.JS, LaTeX
Web Development	PHP, HTML, JavaScript, jQuery, Bootstrap, CSS3, Apache, Nginx, ReSTful architecture
Languages	English, Japanese (N4)

Publications

Di, J., Hickey, C., Bumgardner, C., Yousif, M., Zapata, M., Bocklage, T., ... & Qasem, S. A. (2024). Utility of artificial intelligence in a binary classification of soft tissue tumors. Journal of Pathology Informatics, 100368.

Bocklage, T., Cornea, V., Hickey, C., Miller, J., Moss, J., Chambers, M., Bachert, S. (2024). Ki-67 testing in breast cancer: Assessing variability with scoring methods and specimen types and the potential impact on therapy eligibility, Applied Immunochemistry & Molecular Morphology (AIMM).

Bumgardner, V. K., Klusty, M. A., Logan, W. V., Armstrong, S. E., Hickey, C., & Talbert, J. (2024). Institutional Platform for Secure Self-Service Large Language Model Exploration. arXiv preprint arXiv:2402.00913

Cody Bumgardner, V. K., Klusty, M. A., Vaiden Logan, W., Armstrong, S. E., Hickey, C., & Talbert, J. (2024). Institutional Platform for Secure Self-Service Large Language Model Exploration. arXiv e-prints, arXiv-2402

Jarrells, M., Armstrong, S., Hickey, C., Bumgardner, C., & Yu, M. (2023). B-198 Web-based Application of an Automated Quality Assurance Monitoring System for a Pain Management Panel by Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS). Clinical Chemistry, 69(Supplement_1), hvad097-527

Bumgardner, V. K., Mullen, A., Armstrong, S., Hickey, C., & Talbert, J. (2023). Local large language models for complex structured medical tasks. arXiv preprint arXiv:2308.01727

Bumgardner, V. C., Armstrong, S., Virodov, A., & Hickey, C. (2023). Automated Curation and AI Workflow Management System for Digital Pathology. AMIA Summits on Translational Science Proceedings, 2023, 71

Hickey, Caylin, Bumgardner, V.K.C, Miller, Justin 2022, Leveraging edge computing for workflow tracking and management to improve academic and healthcare security, efficiency, and auditability. Presentation at 2022 ISCB Rocky Mountain Bioinformatics Conference, Aspen/Snowmass, CO, USA, December 2022.

Hickey, Caylin, Bumgardner, V.K.C, Miller, Justin 2022, Leveraging edge computing for workflow tracking and management to improve academic and healthcare security, efficiency, and auditability. Poster session presented at 6th Annual Commonwealth Computational Summit, Lexington, KY, USA, October 2022.

Di, J., Hickey, C., Bumgardner, C. and Qasem, S. "Cell-Specific Parameters Differentiate Benign from Malignant Soft Tissue Tumors." LABORATORY INVESTIGATION. Vol. 102. No. SUPPL 1. CAMPUS, 4 CRINAN ST, LONDON, N1 9XW, ENGLAND: SPRINGERNATURE, 2022.

Yousif, M., Hickey, C., Zapata, M., Bocklage, T., Balzer, B., Bui, M., Gardner, J., Bumgardner, C., Pantanowitz, L. and Qasem, S. "Curbside Consults for Soft Tissue Pathology Using Whole Slide Images." LABORATORY INVESTIGATION. Vol. 101. No. SUPPL 1. CAMPUS, 4 CRINAN ST, LONDON, N1 9XW, ENGLAND: SPRINGERNATURE, 2021.

Ismail, B., Benrajab, K., Hickey, C., Ahmad, M. S., Rosenau, J., & Bumgardner, V. C. "DEVELOPMENT OF A MACHINE LEARNING MODEL TO PREDICT UNFAVORABLE OUTCOMES AMONG PATIENTS WITH ACUTE HEPATITIS A VIRUS INFECTION." The Liver Meeting Digital Experience™. AASLD, 2020.

Bumgardner, V. K., and Caylin Hickey. "Huge Data in Medicine (White Paper)." (2020).

Bumgardner, Cody, Caylin Hickey, and Nima Seyedtalebi. "Agent Communications in Edge Computing." 2019 IEEE International Conference on Industrial Internet (ICII). IEEE, 2019.

Bumgardner, V. K., Nima Seyedtalebi, and Caylin Hickey. "Toward Edge-enabled Cyber-Physical Systems Testbeds." arXiv preprint arXiv:1910.01173 (2019).

Bumgardner, V.K.C., Hickey, Caylin, Watkins, Brad, Seyedtalebi, Nima 2018. Edge computing in AI and Big Data applications. Poster session presented at the Defense and Smart Cities TechConnect, Tampa, FL, USA, October 2018.

Chen, XI[†], Hickey, Caylin[†] "Parallelized Interactive Machine Learning on Autonomous Vehicles in Simulated Environment" 2018 70th National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, July 2018.

Bumgardner, V.K.C., Insko, George, Hickey 2018. Cresco Framework: Edge Computing in Homeland Security. Poster session presented at the 4th Annual Kentucky Intelligence Colloquium, Lexington, KY, USA, April 2018.

Chen, XI, Hickey, Caylin 2018. Parallelized Interactive Machine Learning on Autonomous Vehicles in Simulated Environment. Poster session presented at the 2018 NVIDIA GPU Technology Conference, San Jose, CA, USA, March 2018.

Bumgardner, V.K.C, Hickey, C.D., Marek, V.W. "An Edge-Focused Model for Distributed Streaming Data Applications" 2018 International Workshop on Pervasive Flow of Things (PerFoT). IEEE, 2018.

Bumgardner, V.K.C, Hickey, C.D., Marek, V.W. "Edge-enabled Distributed Network Measurement" 2018 2nd International Workshop on Smart Edge Computing and Networking (SmartEdge). IEEE, 2018.

[†]Co-first author

Bumgardner, V.K.C., Hickey, Caylin, Insko, George, Watkins, Brad 2017. Cresco Framework: Edge Computing in Defense Management. Poster session presented at the 2017 Defense Innovation Challenges, Tampa, FL, USA, October 2017.

Bumgardner, V.K.C., Hickey, Caylin, Shapiro, Scott, Cottle, Chad 2017. Cresco Framework: Edge Computing in Smart Cities. Poster session presented at the 2017 Smart Cities Connect Conference and Expo, Austin, TX, USA, June 2017.

Bumgardner, V.K.C, Marek, V.W., Hickey, C.D., et al. "Constellation: A Secure Self-Optimizing Framework for Genomic Processing." 2016 18th International Conference on E-health Networking, Application & Services (HealthCom). IEEE, 2016.

Bumgardner, V.K.C, Marek, V.W., Hickey, C.D. "Cresco: A distributed agent-based edge computing framework." 2016 International Workshop on Green ICT and Smart Networking (GISN 2016). IEEE, 2016.

Presentations

2022 ISCB Rocky Mountain Bioinformatics Conference

Viewline Resort, Aspen/Snowmass, CO USA

Abstract Presentation for <Leveraging edge computing for workflow tracking and management to improve academic and healthcare security, efficiency, and auditability> *Dec. 2022*

- Present the applications of edge computing as they relate to the automation, efficiency, and security of healthcare-related processes

2022 6th Annual Commonwealth Computational Summit

Gatton Student Center, University of Kentucky, Lexington, KY USA

Poster Presentation for <Leveraging edge computing for workflow tracking and management to improve academic and healthcare security, efficiency, and auditability> *Oct. 2022*

- Present the applications of edge computing as they relate to the automation, efficiency, and security of healthcare-related processes

2018 National Aerospace and Electronics Conference (NAECON)

Holiday Inn Dayton, Dayton, OH USA

Paper Presentation for <Parallelized Interactive Machine Learning on Autonomous Vehicles in Simulated Environment> *Jul. 2018*

- Present our work on incorporating interactive machine learning into a simulated driving environment for both demonstrative and critique-style training methods

4^{th} Annual Kentucky Intelligence Colloquium

William T. Young Library, University of Kentucky, Lexington, KY USA

Poster Presentation for <Cresco Framework: Edge Computing in Homeland Security> *Apr. 2018*

- Present the applications of edge computing as they relate to efforts in the area of Defense and the prevention of human trafficking

NVIDIA GPU Technology Conference

San Jose McEnery Convention Center, San Jose, CA USA

Poster Presentation for <Parallelized Interactive Machine Learning on Autonomous Vehicles in Simulated Environment> *Mar. 2018*

- Present our work on the effectiveness of utilizing interactive machine learning (IML) to increase training efficiency in deep reinforcement learning, specifically Deep Q-Learning Networks (DQN), in a virtualized training environment

2018 International Workshop on Pervasive Flow of Things (PerFoT)

Divani Caravel Hotel, Athens, Greece

Paper Presentation for <An Edge-Focused Model for Distributed Streaming Data Applications> *Mar. 2018*

- Present our work on an efficient model for describing distributed streaming data applications for deployment in an edge-computing-enabled system

Defense Innovation Technology Summit

Tampa Convention Center, Tampa, Florida USA

Poster Presentation for <Cresco Framework: Edge Computing in Defense Management> *Oct. 2017*

- Present the applications of edge computing as the relate to efforts in the area of Defense

Smart Cities Connect Conference & Exposition, cohosted by US Ignite

Austin, Texas USA

Poster Presentation for <Cresco Framework: Edge Computing in Smart Cities> *Jun. 2017*

- Presented our collaborative efforts with the Lexington-Fayette Urban City Government related to Smarter Cities initiatives.

OpenStack Birthday Event

University of Kentucky, Lexington, Kentucky USA

Presenter for <OpenStack Success Stories> *Jul. 2016*

- Detailed how OpenStack has aided in the development of the GENI Monitoring system employed at the University of Kentucky

24th GEC (GENI Engineering Conference)

Arizona State University, Tempe, Arizona USA

Presenter for <Monitoring Support for Experimenters and Developers> *Mar. 2016*

- Introduced GENIMon API development progress as well as utilization tutorial

23rd GEC (GENI Engineering Conference)

University of Illinois at Urbana Champaign, Illinois USA

Presenter for <Tutorial: GENIMonitoring: Exploring the System from the User Perspective> *Jun. 2015*

- GENIMon user interface hands on tutorial session to familiarize new users and expose monitoring tools

22nd GEC (GENI Engineering Conference)

Washington, D.C. USA

Demo for <Stream Query Building system using ESPER> *Mar. 2015*

- Demonstration on a graphical interface to define streaming queries using ESPER as well as piping results to either further queries or storage systems

22nd GEC (GENI Engineering Conference)

Washington, D.C. USA

Presenter for <GENI Operations> *Mar. 2015*

- Update on new data and new features available to GENI Operators. Collecting community input for new development and integration.

21st GEC (GENI Engineering Conference)

University of Indiana, IN USA

Presenter for <GENI Operations> *Oct. 2014*

- Covered GENI Monitoring Collector topics: 1) History (Last GEC) -> Current State : we show the existing reporting capabilities and current navigation in a ”reporting” mode. 2) Short-term (for next GEC) : Describe our transition from reporting to alerting and backend changes. 3) Longer term : goals related to user-defined queries, profile development, etc.

Media

Dell, Inc

Print

Case Study *2017*

- HPC tailored to research needs with OpenStack cloud

HPE

Palo Alto, CA

Case Study *2017*

- University of Kentucky researchers move massive data loads easily with an SDN-enabled campus network