

James “Michael” McGarrah

Data Scientist & Cloud Architect

Email: mccgarrah@gmail.com

Phone: +1 (919) 807-1057

Resume: <https://www.mccgarrah.org/resume>

LinkedIn: <https://linkedin.com/in/michaelmccgarrah>

CAREER PROFILE

My goal is to engage my diverse set of skills to solve interesting problems with Machine Learning for your business. My background includes Cloud Engineering, Architecture, Security Controls, and Software Engineering. Along with these skills, I have over the last decade become proficient in Machine Learning practices in both an academic and professional setting. My area of expertise is building Data Science platforms, often in the cloud, and producing valuable insights for businesses. My overly detailed longer CV can be found at <https://mccgarrah.org/resume/print>

EDUCATION

Masters of Science in Computer Science

2014 - 2019

Georgia Institute of Technology

GPA: 3.63

While working full-time, I completed my Machine Learning focused curriculum. My focus areas were Interactive Intelligence, Machine Learning and Perception Systems. While progressing in the program, I pursued beyond the standard program guided independent studies with research projects and an associated conference and papers to extend my academic credentials. This was a cross-discipline Socio-Linguistics and Computer Science engagement..

- NWAV 46 Conference Presentation “155. Introducing NordFA: Forced alignment of Nordic languages. Nathan J. Young & Michael McGarrah”
- Abstract: Forced alignment for Nordic languages: Rapidly constructing a high-quality prototype. (PDF)
- LG-FAVE Software Repository

Bachelor of Science in Computer Science

1990 - 2011

North Carolina State University

GPA: 3.42

Completed my degree graduating with Honors while working full-time and raising a family.

- Member Association for Computing Machinery (ACM)
- Member Institute of Electrical and Electronics Engineers (IEEE)
- Member Association of Information Technology Professionals (AITP) (archive)
- Upsilon Pi Epsilon (UPE): International Computer Science honor society

PUBLICATIONS

Forced alignment for Nordic languages: Rapidly constructing a high-quality prototype

Nathan J. Young & Michael McGarrah

Nordic Journal of Linguistics, 2021

Lead Principal Platform Engineer (ML and Cloud Focus)

Oct 2021 - Present

Envestnet, Inc., Emerald Isle NC (Remote)

Building large scale cloud platforms for financial services and data science initiatives. A combination of AWS cloud and data center infrastructures are included in the architecture allowing for flexible integration between on-premise and cloud services. An automation first attitude pervades this work using Terraform, scripting and Ansible. Containerized cloud-first technologies (AWS & Kubernetes) to support new initiatives are current focus areas. This includes a variety of AWS Resources RDS/Aurora, vEMR (Hadoop on EKS), MWAA/Airflow on EKS, and EKS hosting microservices being primary contenders. We also have legacy tech stacks including MySQL and PostgreSQL installs on EC2/VM directly with integrations for BC/DR support for those direct installs.

My participation in and actively seeking out multiple internal projects with any machine learning or data engineering and/or data management as they become available. These projects have used a variety of methods including serverless machine learning (Lambda/Step) development and SageMaker with container deployment mechanisms. They include finding the data in the enterprise and providing repeatable data feeds which is a considerable challenge with our deeply siloed tech stacks and culture. Providing cloud compute platforms for data scientists to do evaluations of the data and promoting as production models continues to be an interesting and difficult process that we are continuously refining.

Principal Data & Analytics Platform Engineer

Apr 2019 - Oct 2021

Blue Cross and Blue Shield of NC, Raleigh-Durham NC (Remote)

Building a data science platform incorporating a data-warehouse, data-lake and data-execution environment using containerized cloud-first technologies (AWS & Kubernetes) to support our mission of data driven decision to provide for a value-based healthcare approach. The primary platform automation was done in AWS CloudFormation with Terraform being used later in the project. Ansible Tower was used for managing various configurations across the infrastructure both in the data center and cloud. My active participation in various AI/ML projects along with my infrastructure and platform roles was core to my efforts to bring the best capabilities to our Data Science Initiatives as an active participant on both sides.

My most direct work in the ML space was for CarePath. CarePath is a software framework developed by Blue Cross NC that acts as a deep learning model factory. It constructs state-of-the-art deep neural networks which learn and then recognize the patterns in sequences of claims which eventually lead to particular health events. This enables Blue Cross to accurately identify members at risk for a wide range of preventable events.

We used this in-house general framework to build specific health related models to identify member level risks and possible health impact to focus our care management programs on those members at greatest risk and improve their health outcomes.

Production examples of this include:

- Complex Case Management (CCM) models for predicting both potential hospital initial admissions and separate readmissions after care using claims and other historical data to target members at risk with the goal of improving health outcomes. These two models have had a measurable impact on member health.
- Hospital to Home (H2H) model to identify members at high risk for readmission who need support transitioning from inpatient care to their homes.
- Additional models for Cardiovascular disease, Diabetes

My role was to validate the framework and models along with assisting in making the code-base production ready. I was also the primary implementer for creating an enterprise production capable platform for the final product. The platform implementation was initially on local laptops and later a set of AWS EC2 instances. I performed code reviews and made changes for the framework source code, individual model data-prep and code, enterprise processes and data management as we moved out of the internal POC. I rewrote sections of the framework and model code to allow migration to scalable cost-effective containerized managed Kubernetes (K8S) deployment using our multi-stage secure cloud-native development platform that provided for a fully self-managed developer experience via a cross-account CI/CD Pipeline. This allowed for fast cycle times for the development. Further improvements I added included zero-scaled K8S GPU nodes instantiated via resource taint/toleration, scalable and highly available scoring access for the daily public model releases and secure access to managed internal only datastores for the model generation. This was a complex and very powerful environment for Data Science investigations. It was later simplified to an AWS SageMaker and OpenShift environment post-2022.

Data Engineer

Nov 2017 - Apr 2019

United States Postal Service, Raleigh NC

As part of the Advanced Analytics Group in Enterprise Analytics Program of the USPS, engaged across multiple platforms requiring moderate to high controls with federal content from FBI and DEA. Primary focus on a SAS 9.4 and SAS Viya in-memory cluster with linkage to Hadoop Datalake inching up on 1Pb of data. Exploring machine learning and big data. We work with Tb not Gb of data in a high security environment.

The installation of the multiple clusters used Ansible in a stand-alone configuration without an Ansible Tower. This required extending the SAS Ansible installer to provide for higher level controls for security than provided. Tech stack included Ansible 2.2.1/2.3.2 automation for platform installation and configuration with a custom quickly built Python Fabric 1.14.x extension for system administration automation and management to provide a standardized method for administration tasks.

I was a participating member in the Computer Vision rapid response project team to track the pipebomb packages from the USPS 2018 mail bombing https://en.wikipedia.org/wiki/October_2018_United_States_mail_bombing_attempts through the USPS logistics process. This heavily optimized image pipeline was used to find and track suspected packages. Rapid reconfiguration of platforms while building new AI/ML capabilities was taxing. We worked around the clock with lives on the line.

Some less life threatening work was building refined GIS models to enable better Feature Engineering for logistics path analysis of mail delivery in the rural United States. My contribution was both Data Engineering in Hadoop, and developing a custom library and API for interfacing with refined ESRI ArcGIS to be uploaded into SAS Viya and our local AI/ML processing systems.

Cloud Architect

Jul 2016 - Sep 2017

American Kennel Club, Raleigh-Durham, North Carolina Area

Oversee the design and execution of the cloud computing strategy including the cloud adoptions plans, cloud application design, and cloud management and monitoring. Provide expertise in the definition, design, implementation, adoption and adherence to enterprise architecture strategies, processes and standards. This role at AKC was focused on expertise in the Amazon cloud technologies. Used Gitlab, Phabricator, SONAR, Ansible, and Jenkins platforms in AWS for existing development. Built extensions using Docker for containerized development and testing platforms on local machines.

Enterprise Architect for Strategic Initiatives

May 2015 - Jul 2016

North Carolina Department of Information Technology, Raleigh NC

Accountable for developing, maintaining and overseeing the execution of formalized technology, application, platform, and systems integration strategies. Focus on technology of software defined networks (SDN) and software defined infrastructures (SDI).

Senior Systems Programmer & Administrator (AI Technologies)

Apr 2013 - Jul 2015

Measurement Incorporated, Durham, NC

Systems Administration for Linux and Windows on Cloud platforms. Implement integration of Active Directory with UNIX authentication & authorization. Provide support for PostgreSQL, MySQL, & SQLServer database environments. Systems programming with optimization of low-level machine learning support libraries such as CBlas, LPSolve, Shogun and others. Expand network topology. Implement virtualization with VMware ESXi, AWS, KVM, & Xen. Implemented a AWS clone using the Eucalyptus 3.4 platform compiled from source with iSCSI shared storage on FreeNAS. AWS Services cloned included: EC2, EBS, IAM, S3, AMI, Autoscaling, Elastic Load Balancer, and Cloudwatch with a local Ceph & RiakCS S3 test.

Implement a custom MIT Star Cluster for distributed computing to a combined AWS and Eucalyptus environments. Used lower level tooling to optimize base toolchain libraries such as ATLAS, LPSolve, CBlas, Eigen, ColPack, ARPREC, Ccache, and LAPack for optimizing the non-GPU accelerated machine learning code. Library optimization was done to improve AWS EC2 virtualization runtime operation which was successful with between 5%-20% improvements depending on data size and content. Had to become proficient in machine learning algorithms to resolve runtime issues in production workloads from linux core dumps.

Build an analytics server providing Octave, R, Shiny, and rStudio server, SAGE Notebook / IPython Notebook, and MySQL for the AI researchers. Developed an interactive interface and shared environment for use between applications improving the researchers self-service capabilities in the data analysis steps of their NLP investigations.

Systems Administrator

Aug 2011 - Mar 2013

SAS Institute, Inc., Cary, NC

Systems administration for Windows, Solaris, AIX and Linux for SAS Solutions OnDemand with a focus on validated systems for the pharmaceutical industry. SAS Drug Development (SDD), Clinical Data Integration Studio (CDI) and SAS Clinical Standards Toolkit are my primary product focus. SAS Retail and Fraud are a secondary focus area. Provided customers with automation solutions and POC extensions as necessary. Certification in several SAS product lines. Automated installations and deployments with Chef (similar to Ansible). Used VMware ESXi with vFabric tcServer integration to SAS Drug Development 4.2. Built a VMware ESXi 5.1 VM only system for a full three-tier platform as a demo of tech stack that included databases, web-servers, and storage.

Application and Network Security Specialist

Mar 2007 - Aug 2011

NC Department of Revenue, Raleigh NC

Review compliance to required state and federal policies. Review security requirements and adherence for the agency. Provide risk assessments. Respond and document security incidents. Document security procedures. NIST, FISMA, FIPS, and ISO are some standards I actively reviewed during my tenure. As a security specialist, provided network engineering services in cooperation with network engineering with management of firewalls, NAC and SIEM log management.

IT Engineer (IT Director)

Mar 2006 - Mar 2007

BD Biosciences, Durham, NC

Provide for all IT services at the BD Treyburn manufacturing plant located in Durham NC. Skills utilized include Microsoft Windows admin, Lotus Notes admin (email), Oracle DBA (8.x and 9i), Microsoft SQL Server 2000 DBA, Sun Solaris 8 and UNIX admin, network admin (Nortel and Cisco devices), system programmer and some work with Allen & Bradley ControlLogix PLC.

Systems Programmer

Jan 2005 - Mar 2006

NC Community College System, Raleigh NC

Sun Solaris and IBM AIX systems administrator, UNIX programming in Perl and shell, and write technical documentation for colleges on system tasks. Network and OS support for colleges.

Systems Engineer

May 2004 - Jan 2005

Hosted Solutions (Springboard Hosting), Raleigh, NC

A senior UNIX & Windows systems and network administrator with an emphasis on storage technologies. I also provided guidance to off-shift personnel in three separate operation centers.

Systems Engineer (DevOps/Agile)

Jun 2002 - May 2004

NetIQ, Inc. (Ganymede Software, Inc.), Morrisville, NC

Brought a broad background in UNIX and Windows programming and administration to the new UNIX agent services for the flagship AppManager product. Managed the lab and was the primary systems administrator and quality engineer for the UNIX services and also maintained and extended network testing facilities.

Development Manager / Operations Manager

Jun 1998 - Jun 2002

North Carolina State University, Raleigh, NC

Managed development and primary architect of NC LIVE web services and College of Engineering financial systems. Managed Cisco 4700 using Distributed Director product to provide load balancing and redundancy of service. Provide senior level guidance to UNIX systems administrators. Acting Operations Manager for NC LIVE server sites. Implementation included Java, C, Perl, and integration with early LDAP services tied to Microsoft Kerberos.

Systems Programmer (Benchmark Developer)

Jun 1994 - Jun 1998

Ziff-Davis Publishing, Raleigh, NC

Provided consulting services for internal project development for magazines. Developed industry standard benchmarks for public distribution. Wrote benchmarks in a combination of C/C++, Assembly, Java, Perl and SQL.

Technical Services Manager

Jul 1992 - Jun 1994

Pioneer Software, Inc (Q+E Software / Intersolv / Merant / Data Direct / Serena), Raleigh, NC

OEM services and international support manager. Supported over 30 commercial DBMS. Worked, on an as needed basis, for development, quality assurance and technical-writing departments.

CERTIFICATIONS

Certified in Cybersecurity (CC) International Information System Security Certification Consortium (ISC2)	Nov 2024 - Nov 2027
ITIL Foundation in IT Service Management v2011 (2016) AXELOS Global Best Practice (02354900-01-E6V6)	2016 Axelos
Foundation Certificate in IT-Service Management (2006) itSMF EXIN	2006
AWS Certified SysOps Administrator - Associate Amazon (ZMS5T5T11N1E1XK5)	2017 - 2020
AWS Certified Solutions Architect - Associate Amazon (QD3RGLB1KNQQQ295)	2017 - 2020
USPS Analytics University Bronze Certification for Delivery Operations United States Postal Service	2018
Learning the OWASP Top 10 LinkedIn	2022 LinkedIn
SAS Certified Base Programmer for SAS 9 SAS Institute (BP029145v9)	2012
SAS Certified Clinical Trials Programmer Using SAS 9 SAS Institute (CTP000286v9)	2012
SAS Certified Data Integration Developer for SAS 9 SAS Institute (DID000615v9)	2012
SAS Certified BI Content Developer for SAS 9 SAS Institute (BICD000693v9)	2012
SAS Certified Platform Administrator for SAS 9 SAS Institute (PA000207v9)	2012
SAS Certified Statistical Business Analyst Using SAS 9 - Regression and Modeling SAS Institute (SBARM000177v9)	2013
Solaris Certified Systems Administrator - Solaris 8 (2003) Sun Microsystems	2003 - 2005
Red Hat Certified Engineer (2001) Red Hat, Inc.	2001 - 2006
Microsoft Certified Professional - 70-215 & 70-210 - Windows 2000 (2003) Microsoft Corporation	2003
Nortel Networks Certified Design Specialist (2002) Nortel Networks	2002

PROJECTS

Certification Timed Multiple Choice Quiz WebApp (Full Stack Development)

This application is a fully-fledged, dynamic quiz platform built on the lightweight Flask web framework for Python. The code demonstrates both front-end and back-end development using modern tools such as Bootstrap5 and dynamic rendering in the pages. Forgive me if the demo site is down... free tier hosting.

Website: <https://plain-gaby-mcgarrah-a35e7264.koyeb.app>

Code: https://github.com/mcgarrah/legendary_quick_quiz

AWS Quicksight Demonstration (Terraform Automation)

The goal of this repository is to automate as much as possible, from the varied AWS posts on Federated QuickSight integration, the manual steps in Terraform and convert the CloudFormation to Terraform where possible. Also, updating the Python Lambda layers and code to more current versions of libraries and runtimes is in scope as well.

Code: https://github.com/mcgarrah/quicksight_demo

Phonetic Transcriptions (acoustics modeling)

My academic partner and I automated identifying the sounds of a non-English language, found the temporal boundary markers or segmentations, and fully automated output of segmentation results in a linguistically useful format. My part was generalizing the code to handle the language features and separate those into a configuration file. My partner created the non-English language dictionaries. The heavy lift was building the dictionaries for Swedish and Danish but he needed a platform to use them. I also built a website to present an interface.

Video: <https://www.youtube.com/@MichaelMcGarrah> (Phonetic Transcription)

Website: <https://lgas.phonemes.org>

Code: <https://github.com/mcgarrah/LG-FAVE>

Photosynth Point Clouds Extraction and Visualization (Computation Photography)

Extraction and visualization of point cloud data from the Microsoft Photosynth web service. Exporting the data into common formats for further analysis and processing.

Video: <https://youtu.be/WZPtunNaqVc> (<https://www.youtube.com/@MichaelMcGarrah>)

Code: <https://github.com/mcgarrah/PyPhotoSynthExport>

Computation Photography Video Stabilization Demos

Stabilize and automatic looping of video sessions for varied scenes finding keyframes. The video stabilization is enabled to assist in extracting individual frames for a keyframe analysis of the video.

Video: <https://youtu.be/WZPtunNaqVc> (<https://www.youtube.com/@MichaelMcGarrah>)

Artificial Intelligence for Robotics (CS8803-001) (AI for Robotics)

The goal for the final project in CS6475 AI for Robotics was to create a robotic platform to investigate computer vision technology. The platform included an Arduino with sensors and motors and a Raspberry Pi 2 for the vision and primary control system. A remote control web-ui was included as a cool extra for demoing to graders. The project URL is a video channel that shows the progress and challenges.

Video: <https://goo.gl/QmKfW3> (<https://www.youtube.com/@MichaelMcGarrah>)

"Senior Design: NetApp iSCSI with RDMA/TOE" (Fall 2010 Capstone Project)

Develop a methodology for 10Gbps iSCSI that utilized RDMA and TOE to reduce primary CPU and memory usage. Document and manage the project as a team leader. The resulting iSER implementation was utilized in a related NetApp Phd project at the University of Bangalore India.

"Ecommerce & J2EE: Kodak Image Mapping (similar to Google Maps)" (Fall 2001 Graduate Class)

A web application similar to Google Maps but five years earlier and allowed for monetizing the service. Interviewed customers and defined specifications for and delivered a prototype web service for an online map delivery system. Service allowed for customers to purchase and schedule aerial imaging via aircraft or satellite imaging at varied resolutions.

SKILLS

Programming Languages:

General: Python (1.5, 2.7, 3.x) • Java • C/C++ (Windows/UNIX) • SQL (ISO/IEC portable) • Scala • TypeScript/JavaScript • Shell Scripting (Bash, Zsh, Bourne) • C# • Perl (4/5.005/5.6/5.8/6) • Smalltalk • PHP • Ruby • Base SAS 9 • Visual Basic • Pascal • FORTRAN • PowerBuilder

laC: Terraform • CloudFormation • Ansible • Chef • Puppet • pyFabric

Platforms, Cloud & DevOps:

OS Platforms: Linux (Ubuntu, Debian, Red Hat Enterprise, CentOS) • UNIX (Sun Solaris, AIX, HP-UX, FreeBSD, OpenBSD, NetBSD, Ultrix, SunOS) • Mac OS X (System 6.x – X) • Microsoft Windows (NT 3.1 - current) • VMware ESXi

Cloud Platforms: AWS S3, EKS (Kubernetes), Athena, QuickSight, EMR/vEMR, CloudFormation, SageMaker, MWAA (Airflow), Lambda, Step Functions, EventBridge, CloudFront, IAM, with Kubernetes (EKS, Istio, Calico, Cilium, ArgoCD, etc...) and Hadoop on EC2, EMR, vEMR. Azure: Varied integration for SSO and key management services. GCP: Apigee and Compute resources. Misc: Rackspace, IBM Cloud, and other hosting services

CI/CD & Automation & Collaboration: ArgoCD • GitLab CI/CD Pipelines • Jenkins • Git • Gitlab Pipelines • GitHub & GitHub Actions • Gitea • Helm • Docker & Rancher • Terraform • Ansible • Maven • Ant • Makefile • Maven • SVN & CVS & RCS • Perforce • SourceSafe • Jira • Redmine • Asana • Bugzilla

Serverless: Lambda • Step Function • EventBridge • API Gateway • Athena • CloudFront • CloudWatch • DynamoDB • SQS • Aurora Serverless

Databases & Data Management:

Databases: (SQL) PostgreSQL • MySQL • Microsoft SQL Server • Oracle • Sybase • IBM DB/2 • UniData • SQLite • HyperSQL (NoSQL) MongoDB • Cassandra • DynamoDB • Redis

Data Science & Analytics:

Libraries: Pandas • Polars • Matplotlib • Seaborn • Spark • Scikit-learn • NumPy • PyTorch • SciPy • Keras

Tools: Jupyter Notebook • Zeppelin (for Hadoop) • DataIKU • Snowflake • Hadoop • TensorFlow • Airflow

Skills: Machine Learning • Deep Learning • Natural Language Processing (NLP) • Computer Vision

Software Development:

Development Practices: Agile • Software Development Lifecycle • Full Stack Development • APIs (REST, WebSocket, gRPC, SOAP) • Microservices Architecture • Test-Driven Development (TDD) • Object-Oriented Programming (OOP) • Data Structures & Algorithms • HCI • UI/UX

Web Technologies: HTML5 • CSS (Bootstrap, Tailwind) • Node, Angular, React, Vue • Flask • Django • FastAPI • Ruby on Rails • Groovy on Grails • Gradio • Apache • Nginx • Caddy • Varnish • Tomcat • Traefik • HAProxy

Security & Compliance: NIST 800-53 • ISO/IEC 27001 • Cloud Controls Matrix (CCM) • SOC 2 • ITIL