Matthew J. Collins
1658 NW 22nd Circle
Gainesville, FL 32605
352-275-4027
mjcollinsresume@gmail.com

Professional Summary

I have delivered reliable software and infrastructure to industry and academia. My backgrounds in web development, system administration, and data management give me a functional perspective on the current challenges of data-centric organizations. I am focused on working in high-functioning teams to collaboratively solve data access and discoverability problems.

Professional Experience

Data Engineer   January 2019 – Current

Constant Contact via acquisition of SharpSpring, Gainesville, FL

Data engineer for leading online marketing automation company building cloud-based data systems.

Key Activities

- Design and maintain user-facing data aggregation pipelines that back the SharpSpring app's custom reporting feature based on MariaDB and Elasticsearch
- Author Python data pipelines to populate a Google BigQuery data lake with all SharpSpring application and business data from Vitess, MariaDB, MongoDB, and 3rd party APIs
- Maintain Dockerized Airflow deployment running in a Google Cloud Platform Kubernetes Engine (GKE) cluster
- Standardize and approve all application database schema changes to align software developers' data designs with data analysts' reporting needs and best practices
- Mentor business analysts in software development practices and manage a business intelligence engineer

Technical Operations Manager   May 2010 – January 2019

Advanced Computing and Information Systems Laboratory, University of Florida, Gainesville, FL

Infrastructure software developer for computer engineering academic research laboratory specializing in big data, cloud, and virtualization systems.

Key Activities

- Design, develop, and maintain open-source infrastructure for the NSF-sponsored Integrated Digitized Biocollections (iDigBio) project. Used PostgreSQL, Elasticsearch, and Ceph to aggregate 120 million records of natural history museum specimens with 30 million images consuming 160 TB of storage (https://www.idigbio.org)
- Provide training and outreach to biologists in software development, data APIs, and reproducible research practices to increase use of large aggregated data in research
- Collaboratively design, deploy, and use Spark-based data analytics infrastructure with 192 threads and 576 GB of memory for mining and analyzing biodiversity data sets including formatting common data sources into parallelized data frames (http://guoda.bio)

- Maintain the facilities and configurations of over 260 servers running cloud, distributed, and virtualization environments such as Elasticsearch, VMware, XenServer, Spark, Cloudera, and Ceph
- Create collaborations across research institutions including the Encyclopedia of Life, Duke University, Royal Botanic Gardens, Kew, San Diego Super Computing Center, and NARA Institute of Science and Technology to design, present, and publish novel data processing pipelines and systems configurations
- Author publications, presentations, reports, blogs, and other technical communications to disseminate the research of the lab to a global audience including organizing symposia, hackathons, and workshops in multiple countries
- Mentor computer engineering graduate students in the technical aspects of their research and supervise master's projects

System Administrator   January 2008 – May 2010
Florida Museum of Natural History, University of Florida, Gainesville, FL

Lead system administrator for research and business environment with 300 staff and 40 Linux and Windows servers.

System Administrator   April 2007 – January 2008
Bureau of Economic and Business Research, University of Florida, Gainesville, FL

System and workstation administrator for business call-center environment with 250 staff, Linux and Windows servers, and 120 Windows workstations.

Lead Programmer   September 2005 – April 2007
Florida Museum of Natural History, University of Florida, Gainesville, FL

Lead developer on NSF-funded BioCorder grant project. Quarter-time system administrator.

Consultant System Administrator   September 2004 – December 2011
Bear Code, LLC, Montpelier, VT

Part-time consulting AWS system administrator for international software development company.

System Administrator, Programmer   June 2001 – August 2005
Signal Advertising, Inc., Montpelier, VT

System administrator and software developer for Internet development and marketing business.

Database Specialist   May 1997 – June 2001
Stone Environmental, Inc., Montpelier, VT

Manage and present environmental data for EPA reporting at environmental consulting company.

Education

M.S. Industrial and Systems Engineering   May 2016
University of Florida, Gainesville, FL

B.S. Civil Engineering with Environmental Option   May 1997
University of Vermont, Burlington, VT

Skills

Experienced with: Amazon EC2, AWS, Google Cloud Platform, Docker, VMware vSphere, ESX, Citrix XenServer, MySQL, MariaDB, PostgreSQL, Google BigQuery, Elasticsearch, Riak, MongoDB, Ceph, Spark, Hadoop, HDFS, Cloudera, Salt configuration management, Puppet, Foreman, Python, R, Jupyter Notebook, Bash, Java, PHP, MATLAB, Git, Github, Gitlab, Subversion, Jenkins, Travis CI, Looker, Metabase, Apache, Lighttpd, Bind, HaProxy, Postfix SMTP server, Zimbra, Redmine, Drupal, Zabbix monitoring system, Markdown, NFS, ZFS, RedHat Enterprise Linux, CentOS, Ubuntu, OpenSolaris, Windows Server, TCP/IP networks, firewalls and NAT, proxy servers, load balancers, software defined networking, DNS, IPMI, PXE, SNMP, Active Directory, Windows Terminal Server, IBM fibre channel SAN, IBM blade centers, Dell blade centers, iDataPlex, Cisco, Dell PowerConnect, Brocade, Pica8, Force10 switches.