

Matthew J. Collins
2631 NW 29th PL
Gainesville, FL 32605
352-275-4027
mjcollinsresume@gmail.com

Professional Summary

Over my career I have been responsible for delivering reliable and robust software and infrastructure systems to industry and academia. My backgrounds in web development, systems engineering, and large data management give me a functional perspective on the current challenges of data science and engineering. I am focused on developing high-functioning teams to collaboratively solve technical problems.

Professional Experience

Technical Operations Manager

May 2010 – Current

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

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

Key Activities

- 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 re-formatting common data sources into parallelized data frames (<http://guoda.bio>)
- 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
- Design, develop, and maintain open-source infrastructure based on PostgreSQL, Elasticsearch, and Ceph to aggregate 115 million records of natural history museum specimens with 27 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
- Maintain the facilities and configurations of over 260 servers running cloud, distributed, and virtualization environments such as Elasticsearch, VMware, XenServer, Spark, Cloudera, and Ceph
- 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 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, Apache, Lighttpd, MySQL, PostgreSQL, Bind, HaProxy, Postfix SMTP server, Zimbra, Git, Github, Gitlab, Subversion, Jenkins, Travis CI, Redmine, Drupal, Zabbix monitoring system, Foreman, Puppet, Salt configuration management, PHP, Python, R, Bash, Java, MATLAB, Markdown, Jupyter Notebook, Riak, Elasticsearch, Cloudera, Spark, Hadoop, HDFS, Ceph, VMware vSphere and ESX, Citrix XenServer, Amazon AWS, EC2, NFS, ZFS, Active Directory, Windows Terminal Server, IBM fibre channel SAN, IBM blade centers, Dell blade centers, iDataPlex, Cisco, Dell PowerConnect, Cisco, Brocade, Pica8, Force10 switches.

Github account at <https://github.com/mjcollin>