

Bill Coffman
bill.coffman@gmail.com 408.461.5708

SUMMARY

Sr. Data Engineer with very good experience in software development and database engineering using Python, SQL, NoSQL Databases, and Cloud technologies. Expertise include:

- Data Architecture, Database design & optimization, Data Mining, ETL & Data Warehousing
- Expert in Analytics and machine learning using Python NumPy, Pandas, Arrow, SciPy & Octave
- Design, construct, install, test and maintain highly scalable data management systems
- Implement & maintain automated large scale enterprise data ETL processes
- Experienced Software Engineer & Test Engineer: developed DevOps CI/CD platform
- Data Governance, Big Data Solutions and ETL tools. Deployed solutions business environments

EDUCATION

University of California Davis, Davis, CA

PhD work -- All but dissertation, Computer Science, with emphasis in Systems and Networks

Dissertation Title: "Graph Coloring Heuristics and Compiler Optimization".

MS, Computer Science

California State University Fresno, Fresno, CA

MS, Applied Mathematics

BS, Applied Mathematics

TECHNICAL SKILL

Primary Programming Languages: Python3, Python2, C++/C, Perl5, Bash

Python Libraries: NumPy, Pandas, Arrow, flask, sqlalchemy, Airflow

Secondary Languages: Java, R, octave, JavaScript/Nodejs

Relational Databases: SQL, Hive, Presto, PostgreSQL, MySQL, Redshift, Oracle

NoSQL Databases: ElasticSearch, MongoDB; Hadoop, Kafka

Containerization Tools: Docker, Kubernetes

Machine Learning Tools: Jupyter, Pandas DataFrames, Tensor Flow

BI Tools: Tableau, Excel, Splunk dashboard

Code Repositories: Github, Github Actions, Git & SVN

Cloud Technologies: AWS, Elasticsearch, Google Cloud

Operating Systems: Debian, Ubuntu & RedHat Linux, MacOS, & FreeBSD

Development methodologies training: Agile, (scrummaster experience)

Misc: Scalable algorithms, Analytics, Data Science, optimization, Architecture, and DBA skills

PROFESSIONAL EXPERIENCE

ASML, Inc - Wilton CT

1/2022 - present

Software Design Engineer

- Maintained Python 3 code to run CI/CD for Devops
- Built and led Docker containerization as way of running code tests (checks) in an isolated env
- Solved problem of building and running docker containers in GitHub Actions
- Used Polarion (which is a requirements management tool) to capture requirements
- Negotiated with other silos in ASML to set company wide standards for code checks

Lyft, Inc - San Francisco, remotely **5/2021 - 11/2021**

Python Data Engineer - contractor via Randstad

- Architected new pipeline for Lyft financials, which began with CSV files stored in google docs
- Built & Maintain with Python 3 code to for airflow ETL, with Hive/Presto/Spark, with s3 storage
- Created hive & Postgres SQL as well as SparQL for analytics on new Data Science projects
- Automated Google-Drive data ingestion to airflow via python NumPy Arrow Pandas dataframes

Facebook, Inc - Menlo Park, remotely **3/2021 - 4/2021**

Python Data Engineer - contractor via US Tech Solutions

- **Designed and implemented multimedia data ingestion pipeline for data scientists**
- Learned the PHP/HACK language, the FB data graph api, and Dataswarm
- Success, but legal dept killed the project & contract terminated early

Apple, Inc - Cupertino, remotely **8/2020 - 2/2021**

Python Data Engineer - contractor via TEKsolutions

- Maintained airflow pipelines with Python 3 to orchestrate, s3, parquet and Postgres jobs
- Optimized code for speed in creating Splunk dashboards
- Trouble-shot and tested big data infrastructure
- Designed Database; optimized data architecture; did Data Mining, ETL & Data Warehousing
- Built and maintained SQL queries for data insight on various projects
- Implemented, automated and maintained large scale enterprise data ETL process.

Magic Leap - Sunnyvale, CA **1/2019 - 9/2019**

Infrastructure Engineer, Contractor via Swoon, Inc.

- Tooled Docker containers and submitted jobs to Kubernetes on Google Cloud.
- Worked with Linux Debian packaging to package binaries, plus dependencies.
- Hand-coded Python in micro service architecture.
- Provided reporting analytics with Python "matplotlib".

Crypto-Money Soft - Sunnyvale, CA **6/2018 - 7/2020**

Entrepreneur, Lead Software & Data Engineer - worked for equity as time allowed

- Architected infrastructure for Beta Release (aka MVP - minimum viable product)
- Constructed REST API interfaces with flask & python3 for mobile and web applications
- Interfaced with crypto-currency exchange APIs as data source for pipeline
- Created ETL process and data models to populate PostgreSQL, for Machine Learning
- Implemented greenfield automation for medium scale enterprise data ETL processes
- Utilized APIs to Cryptocurrency wallet for transactions.

Elastica - San Jose, CA **9/2014 - 6/2017**

Staff Software Engineer

- Implemented anomaly detection software, called "User Behavior Analyzer".
- Maintained first Perl version; translated to Python; also debugged in Hadoop.
- Cleanroom re-architected a new Python version, working closely with data scientists.
- NoSQL database modeling & design: Elastic Search, Kafka in AWS cloud.
- Developed data collection for Web App training, used for machine learning.
- Created framework to collect data via MITM-Proxy (traffic sniffing).
- Used Python on a Linux server; occasionally modified AngularJS code.
- Worked with high performance C++ multithreaded scanner (Symantec) server.

Facebook **12/2013 - 9/2014**

Data Engineer Contractor

- Worked in FB Data Warehouse group.
- Utilized Hive (and presto) data pipelines that send data to MySQL analytics.
- Used Oracle data warehousing pipelines that move data to excel spreadsheets.
- Assisted with Oracle to Vertica migration.
- Deep data analysis of log tables from hive, requiring python transforms in hive.

WHITEHAT SECURITY, INC.

2/2012 - 11/2013

Sr. Data Engineer, DBA & Data Analyst

- DBA duties: built Oracle DB for Birst data warehouse; maintained over 20 Postgres servers.
- Dimensional modeling, ETL design for Salesforce, Splunk, Postgres, and to Birst.
- Researched Hadoop for big data projects at WhiteHat & reporting with Tableau and NodeJS.

GOOGLE, INC.

6/2010 - 12/2011

Test Engineer

- Wrote test plans and built testing infrastructure to test Google Analytics backend.
- Built python framework to test queries in Google Analytics, resulting in automated tests.
- Worked on Google Analytics Release, validating "big tables" and other Map-Reduce monitoring.

WHITEHAT SECURITY, INC

2006 - 2010

Sr. Software Engineer

- Planned and lead team meetings as ScrumMaster (CSM).
- Tuned accuracy of scans with statistical analysis as a Mathematician/Computer Scientist.
- Wrote text classification algorithms, using Perl, PDL, also C and XS Built.

CONTINUING EDUCATION

- Machine Learning, Stanford University, Online Course by Andrew Ng
- Linear and Logistic Regression, Neural Networks, Support Vector Machines
- Clustering, Anomaly Detection and Dimensionality Reduction, 2012
- Advanced Dimensional Modeling, The Data Warehouse Institute (TDWI), 2012
- Data Mining Camp, (ACM, et al) Two day Big Data and Data Mining seminar. 2012

CERTIFICATIONS

- Certified ScrumMaster, Agile University by Rally Software, 2007

ACADEMIC PUBLICATION

- Bounds for the Chromatic Number of Graphs with Partial Information", Discrete Mathematics, July 2002, Article 4815