Joseph Haaga
Principal Software Engineer
Washington, D.C. US

🔗 http://joehaaga.xyz   ✉ haaga.joe@gmail.com
○ josephhaaga   in joseph-haaga
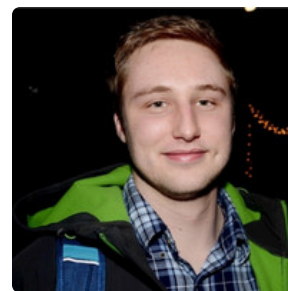
SKILLS

Web Development

html5 | css | lamp stack | mean stack |
django | reactjs | nodejs | jquery |
pimcore | drupal 7/8 | wordpress |
508 compliance

Ops/Infra

serverless | docker | terraform | ansible |
kubernetes | eks | aws | gcp | heroku |
datadog | gitlab ci | argocd | knative |
istio | kubeflow

Software Development

tdd | oop | functional programming |
design patterns | algorithmic analysis |
sql/nosql | postgres | mongodb | neo4j |
elasticsearch & logstash

Machine Learning

llms/chatgpt | prompt engineering |
pandas | tensorflow |
huggingface transformers | spacy |
pytorch | numpy | nlp | classification |
regression/linear modeling |
clustering/unsupervised techniques |
tableau | arcgis

WORK EXPERIENCE

Principal Software Engineer at BuzzFeed November 2022 - Current

https://buzzfeed.com

BuzzFeed, Inc. is an American Internet media, news and entertainment company with a focus on digital media.

- Hired to formalize and scale the ML Platform across BuzzFeed properties including Tasty, Huffington Post, Complex and more.
- Built enterprise ML platform infrastructure serving BuzzFeed Engineering, including NSQ-based real-time feature inference architecture, and automated model deployment from KFP training pipelines.
- Led and managed cloud infrastructure consolidation effort, reducing spend by 50% for monthly savings of $10,000+
- Spearheaded incident response investigations, identifying the root causes of outages, recommending remediations, and providing cross-team coordination.
- Advised on modern Python tooling and DevEx through multiple platform modernization initiatives
- Consulted with chief architect on LLM/agent-friendly dev framework, shaping effective strategy for engineering org
- Mentored senior engineers across multiple teams while providing technical guidance, peer reviews, and leadership coaching.
- Drove strategic product vision for AI-powered content creation through cross-functional collaboration with Editorial, Analytics, and Engineering teams
- Delivered measurable business impact through Recommendation System pipeline optimizations achieving 4.1% CTR improvement for anonymous users
- Recognized for exceptional technical problem-solving, identifying root causes in complex distributed systems and implementing rapid fixes for production failures affecting multiple stakeholder teams

Lead MLOps Engineer at Interos December 2020 - November 2022

https://interos.ai

Interos is the operational resilience company — reinventing how companies manage their supply chains and business relationships — through our breakthrough cloud platform that uses artificial intelligence to model the ecosystems of complex businesses into a living global map, down to any single supplier, anywhere.

- Grew the ML Platform team from 1 to 6 in first year (4 Senior Engineers + 1 Mid-level Engineer + Lead)
- Built an ML Platform supporting 45 models across 4 environments, serving 1M predictions/day. (Terraform, EKS, Kubeflow, Istio, Knative, VPC, Oauth2-proxy, ArgoCD, Gitlab CI)
- Introduced development environments for 20+ engineers providing elastic compute + GPU to support experimentation. (Kubeflow Notebooks, EKS, Istio)
- Spearheaded the Interos Engineering Blog, establishing a content calendar and editorial board & authoring the first content.
- Provided architectural/design oversight to breakout team working on building inner-sourced shared solutions. CTO called result "the most elegant solution to this problem I've ever seen"

Senior Data Software Engineer at Interos July 2020 - December 2020

https://interos.ai

Interos is the operational resilience company — reinventing how companies manage their supply chains and business relationships — through our breakthrough cloud platform that uses artificial intelligence to model the ecosystems of complex businesses into a living global map, down to any single supplier, anywhere.

- Evaluated and implemented common capability for ETL jobs. (Prefect on EKS).
- Designed & built `prefect-deployer`, shared dev tooling used in 455 deployments of 66 ETL jobs (and counting).
- Advised on architecture for platform rebuild as part of a 7-person team chosen by CTO.
- Mentored and supported engineers across the company, and published educational repos to upskill others on Kafka, fixtures & dependency inversion for cleaner unit tests.
- Assisted in onboarding & orienting key engineering leadership hires (Director of DevOps, Director of Data Architecture, Chief Data Scientist)
- Selected by VP of Machine Learning to spearhead ML Operations practice, resulting in promotion to Lead MLOps Engineer.

Senior Data Engineer at Capital One August 2019 - July 2020

https://capitalone.com

Capital One Financial Corporation is an American bank holding company specializing in credit cards, auto loans, banking, and savings accounts, headquartered in McLean, Virginia with operations primarily in the United States

- Supported efforts to achieve compliance with enterprise Cloud security mandates
- Automated reporting for cloud infrastructure vulnerability scans
- Led assessment & remediation of difficult cloud modernization efforts (Terraform, Ansible, boto3)

Data Scientist at Tylor Data Services October 2018 - August 2019

http://tylordata.com/

TDS continues to deliver data science consulting, big data analytics, and custom computer programming services that are proven to address business challenges for government agencies, non-profit organizations and private corporations.

- Led development of a graph analytics tool (React, Flask, Neo4j, Docker) for fraud pattern detection in financial data graphs.
- Supporting research in applying novel deep learning techniques (RNN embeddings, Tensorflow, graphwave) to elicit graph embeddings.

Data Scientist at Redhorse Corporation June 2018 - October 2018

https://redhorsecorp.com

Redhorse Corporation, a program management and technology services company, provides integrated solutions in environmental consulting, energy and climate change, information technology, and intelligence.

- Designed and implemented a image recognition application for environmental science use case.
- Experimented with different convolutional neural network architectures to maximize F1 score.
- Built custom sklearn Pipeline components to take advantage of OpenCV preprocessing techniques.
- Developed web-based utility resulting in a 10x decrease in time spent labelling training data.

Full Stack Developer at Grafik August 2016 - June 2018

https://grafik.agency

Grafik is a branding agency focused on crafting digital brands for B2B, B2C, Nonprofit and government clients worldwide.

- Led architecture and development of CMS-driven solutions for clients like Strayer & Global Automakers, using Drupal 8 and Pimcore 5.
- Contributed to the architecture and development of marketing websites for brands like Strayer, Ligado, Politico, Merritt Clubs, JBG and EYA. Developed on Drupal 7, Pimcore 4/5, and Wordpress.
- Drastically reduced development timelines by building internal tools to streamline internal processes (structured QA spreadsheet with Slackbot integration to close developer-designer feedback loop and closely track live performance/completeness metrics). Built using Google Apps Script, Google Sheets, and Slack Webhooks.
- Delivered customer persona insights by applying data mining & visualization techniques to client CRM & transactional data. Performed analysis using Tableau and Anaconda toolkit (Seaborn, Pandas, NumPy).

Developer at Gray Street Solutions May 2015 - July 2016

https://graystreetsolutions.com

Gray Street Solutions is a boutique, digital-first creative agency, founded on the principle of bringing a "Fearlessly Agile" approach to online communications consulting, open-source web development, and social media campaign management.

- Led development of a cross-platform mobile application to track a brand's social media campaigns, using Apache Cordova and Adobe PhoneGap.
- Developed and maintained digital presences for a variety of trade organizations, defense companies, and nonprofit clients. Developed on Wordpress and maintained on WPEngine managed hosting.

Jr. Solutioneer at Eastport Analytics May 2014 - August 2014

https://www.eastportanalytics.com/

Eastport Analytics is an analytics consulting group applying advanced analytics to uncover valuable insights and drive measurably better mission outcomes.

- Developed automated spreadsheet parsers in Java 7.
- Merged GIS shapefiles with political ad spend data to determine campaign efficacy in Designated Market Areas.

Developer at @HomeRehab May 2013 - April 2014

https://gust.com/companies/homerehab

@HomeRehab designs exercise devices and software to improve the continuum of care related to falls among the aging population.

- Prototyped IoT orthopedic rehabilitative device on Arduino hardware in a dialect of C.
- Developed games to be played with the device to aid in gamification of proprioceptive rehabilitation exercises.
- Implement wireless modules (XBee), LCD screens, & various sensors/actuators within I/O constraints of a development board.

VOLUNTEER

Coordinator & Volunteer at Grafik Gives

August 2017 - Current

https://grafik.agency

Introduce potential initiatives and coordinate CSR events.

- Raised >$1000 for AIGA, providing scholarships for underrepresented design students in the DC area.
- Developed spreadsheet solution for long-term initiative planning, enabling tracking of the organization's impact on local, regional, and national levels.

EDUCATION

Solutions Architect Associate at AWS

2019 - Current

M.Sc. Computer Science at Georgia Tech

2019 - Current

Major courses: software analysis and testing | software architecture and design

M.Sc. Data Science at The George Washington University

2017 - 2018

Major courses: data science | data warehousing | data mining | visualization of complex data | machine learning i | machine learning ii | time series analysis | cloud computing | bayesian methods | member of graduate consulting club | member of mba business analytics club | president of ms data science student council

Grade: 3.92

B.Sc. Computer Science at The George Washington University

2012 - 2017

Major courses: circuit theory | bioinformatics | algorithms | software engineering | database systems | systems architecture | operating systems | machine learning | systems programming | artificial intelligence algorithms | statistical nlp | deepquest ibm watson (essentially a supplement to stat. nlp)

LANGUAGES

Python
advanced

BASH
advanced

SQL
intermediate

Javascript
intermediate

PHP
beginner

Java
beginner