

# Earl Cahill

*LinkedIn - earlcahill.com - Pleasant Grove, Utah - cahille@gmail.com - PDF*

## Experience

Weave

**September 2024 to present**

***Semantic Search Engineer / MLOps / AI / Principal Engineer***

- *De facto* principal engineer over AI for the company
- Trying to do some great things with semantic search, generative AI and all the rest!
- Helped introduce semantic search and Elasticsearch throughout Weave
- Helping on a low latency AI Receptionist that integrates with speech to text, LLMs and text to speech to help front office staff with common tasks like location FAQs and appointment scheduling
- Helped integrate Salesforce data into Elasticsearch to power a customer help knowledge base that uses RAG to help answer questions

Whistic

**October 2022 to September 2024**

***Senior Architect***

- Helping architect and individually contribute code to our generative AI project, which aims to answer security assessment questions using Python, OpenAI APIs and embeddings . Using PostgreSQL, pgvector, and cosine similarity to find best matches and generate responses via Retrieval Augmented Generation (RAG).
- Wrote a system for handling our rather complex local development environments, which brings up Docker, npm and Java services locally, using Python and two simple commands for bringing the environment up and down. The system tries to allow for easy diagnosis of issues, self-healing where possible and even manages git across the many incorporated services. Started virtually from scratch of my own initiative, with the system now used daily by our entire engineering department, saving man years of development time.
- Used AWS Step Functions, DynamoDb and TypeScript to help handle an AI workflow in a more scalable way, easily allowing for failures and retries
- Completed upgrading around 150 repositories to using Node 18, using CDK and CodePipelines from AWS. The repositories varied and required individual code changes, making for a challenging project.
- Added caching to several database queries, allowing for a significant number of queries to not hit our production database.
- Named a member of AI Utah 100

ModelOp

**June 2020 to June 2022**

***Director of Engineering / Principal Engineer***

- Worked hands-on with Kubernetes, Python and Jenkins to help install, update and manage ModelOp Center (our company product) largely in AWS and EKS, but with the flexibility to work with other flavors of Kubernetes or cloud provider. The services were Java Spring boot microservices that (depending on customer setup) interacted with MongoDB, PostgreSQL, MySQL, S3, ServiceNow, Jira, Jenkins, Git and more. ModelOp Center helps companies manage their machine learning models. Completed four-part Kubernetes certification, offered by Google
- Worked directly with customers and our sales team to help install ModelOp Center, ensure quality. Had shell access to customer systems to perform tasks and help diagnose issues
- Wrote Java Spring-based code, which ran as a Kubernetes job that could help diagnose issues. Contributed Spring Cloud Config Server to our fleet, helping allow for team-specific config to the customer
- Helped get suites of integration tests back in good health and helped to improve the framework. Brought new hire documentation more up to date. Helped with using Helm to deploy ModelOp Center
- Volunteered and worked as the company Scrum Master, working to help define our software development lifecycle and plan and track the development work of more than twenty engineers
- Manager of Field Engineering, QA and DevOps teams, including HR responsibilities

Banjo

**April 2020 to June 2020**

***SVP - Engineering***

**Systems team**

- Led investigation into using KEDA to auto-scale Apache Kafka consumers. Completed Terraform training
- Researched and proselytized (with some success) switching from GlusterFS to S3 for storing incoming video streams
- Demonstrated using inexpensive Wyze cameras and RTSP for development purposes

Domo

**October 2015 to April 2020**

***Technical Manager, Principal Engineer***

**Machine Learning Platform team lead**

- Led an effort to get more and better machine learning available throughout Domo
- Designed and wrote code in Java for a system that allows for data scientists to iterate on models and see results in production without further developer assistance. The system feeds user activity data into SageMaker training jobs, and generates artifacts to be consumed by users. Several models can be in production concurrently with the system tying a user to a model. The served artifact and the subsequent user interactions are recorded, allowing for further experimentation and improvement
- Wrote Java code to advance work on *rich types*, which are an abstraction layer above database types, with city, state, zip code, currency, with many more possibilities. My code inspected data on ingest and recorded the rich types, which were later used to populate defaults and allow for easy visualization creation

**Search squad manager**

- Grew system from nearly nothing to being stable in production. System provided faceting, multiple languages, query logging, and custom relevance; and is used by many teams throughout Domo as backend for storing and querying data. The system included Elasticsearch, Java, Spring, MySQL and Reddi. Started as Team Lead, made Technical Manager in September of 2017
- Wrote documentation for ingesting and querying documents, which was used by the search team and other teams when integrating with the system. The hope was that the documentation would be viable on its own, allowing technical employees to quickly add value
- With the team, owned the system into production, with team members being on call. Used Grafana, Datadog, PagerDuty to help monitor and alert on performance and reliability
- System was one of the most used in Domo, handling millions of queries weekly and more than a billion operations monthly. Wrote code that indexed user activity data into Elasticsearch and used that data to provide better relevance in our queries
- Director of back-to-back Domo Video Music Awards winners (Domo Love and Domo Land)

Workfront (formerly AtTask)

**November 2012 to October 2015**

***Principal Engineer***

**Ministry of Data - MOD Squad**

- Lead developer (only developer on squad) for transition from ActiveMQ to Kafka, building framework for message publication and consumption
- Helped provide Oracle database metrics through Kafka, into Hive and with Kibana visibility
- Team was owner of messaging, SOLR, ActiveMQ, HornetQ, Hazelcast and EclipseLink. Shared ownership of Zookeeper, Hadoop, ELK and Kafka

**backend architecture team lead (MSCL - My So Cloud Life)**

- Team covered stability, scalability and performance of our SaaS offering. Led hands-on development for backend work on new SOLR cloud, Zookeeper-powered search system
- Working with Java (JZEE), JMS (ActiveMQ), SOLR, Spring, Maven, Hibernate, Oracle, MySQL, Python and Perl. Led a cross-functional team, organized to investigate and improve slow running SQL queries
- Started at AtTask as a Senior Engineer, promoted to Principal Engineer and team lead in April of 2013

Church of Jesus Christ of Latter-day Saints, **Family Church History**

**March 2007 to November 2012**

***Development Manager, October 2008 to November 2012 - Senior Engineer, March 2007 to October 2008***

As a division, our mission was to organize the world's genealogical data. Our software aimed to aid in scanning microfilm, digitally capturing images worldwide, transcribing genealogical records, ingesting the records into a database and making the images and records available to the public

Personally accountable for my team's technology decisions, ensuring that solutions were implementable, scalable and maintainable. Aided in interviewing and hiring developers. Served as developers' direct report, being responsible for their reviews. Teams practiced agile development. Helped focus developers on getting things done, aiming for meaningful iterations with user-demonstrable features. Team unit test code coverage is generally around eighty percent. Regularly aided in presentations to upper management concerning team progress. Taught multi-month Head First Java course

**record affiliate solutions team dev manager** - managed team across Utah, Europe and South America developing software that handled digitally capturing images with prosumer cameras or scanners, and publishing images. Engaged with record affiliates on five continents. Used jclouds with OpenStack and Amazon EC2 for automated Drupal server install used for image publication

**data framework team dev manager** - managed team that owned the startings of a canonical datastore for digital pipeline data. Received data from various sources, transforming genealogical record data and loading it into Oracle. System used Java, Hibernate, RESTful service, fronted by a WinForms-driven client helps customers manage the collections. Data on FamilySearch for public access and searching

**phoenix team dev manager** - managed team that supported the Acquisitions Management System (AMS), an Oracle-based ERP system that managed 2.4 million rolls of microfilm stored in the Granite Mountain Record Vault. 100s of internal employees relied on AMS daily. Helped transition legacy hardware to VMs and from legacy software to a Linux, Java, Hibernate, RESTful service, fronted by a more modern client

**personal development** - Hadoop - personally scrounged together 26 nodes, writing a custom Perl script to help manage installing new nodes and cluster configuration. Cluster was used to analyze millions of rows of access logs and nearly a billion genealogical records, performing name frequency analysis using Pig. Prototyped a real-time log aggregation framework using Flume, HDFS, HBase and Hive

**personal development - Customer Relations Management system** - as the sole engineering resource, wrote script to install Amazon instances which host SugarCRM. Used script to rather easily change instance size during high traffic period. Wrote cronable script to snapshot volumes nightly. Wrote python monitoring script to run health checks, actually logging into the site, confirming database interaction

**cloud services team dev manager / scrum master** - investigated building internal clouds using Eucalyptus, VMWare and OpenStack. Supporting internal teams as they transitioned to deploying their software in clouds. Had a Eucalyptus system that allowed developers to launch / terminate virtual machines without bureaucratic interference or invoice

**Contract, Rights, Expressions (CRX) team dev manager** - managed team that handled role base access control for millions of images in the digital pipeline, managing permissions for an eventual multi-petabyte image store. Tens of thousands of contracts are managed using Oracle, Java, a RESTful web service and a WPF C# client

**Rosetta Stone team dev manager** - managed team that handled digital pipeline business process management using Lombardi Teamworks Software with help from custom Java and Flex with an Oracle backend. Virtually the entire digital pipeline relied on Rosetta Stone to complete their day-to-day activities. A Flex client was backed by Oracle, Java, and a RESTful web service

**dCamX (digital camera) tech lead** - Grew a team of engineers from myself to four, writing software that interacts with a digital camera and is used for capturing genealogically significant images. Driven by the Java Eclipse Rich Client Platform, C / C++ (via JNI) and Derby (SQL-driven database), dCamX planned for worldwide use, to capture hundreds of millions of images annually.

United Online, Mega Web Services division (formerly about.com and PRIMEDIA)

**January 2000 to April 2007**

***Engineering Manager, May 2006 to April 2007 - Principal Engineer, March 2004 to May 2006 - Senior Engineer, January 2000 to March 2004***

FreeServers **SiteBuilder Integrations** - provided a simple API for auto-installing a SiteBuilder site for sister sites including Classmates. Thousands of daily installations with millions overall

**blog** - drove process to establish a simple YUI-driven blog system with posts, tags, threaded comments and role based access control. Personally wrote and maintained initial API

**log parsing** - helped architect and write a plugin-based system to process billions (two) of monthly hits

**signup / upgrade process for FreeServers** - architected and wrote system for signing up and upgrading customers. CGI::Path-based. Handled millions of hits and dollars through the years

**search** - integrated with Overture / Yahoo! XML API to serve search results

**Lead or sole (\*) engineer on the following projects**

**parallel processing system** \* - loosely based on my masters thesis, my system was used on over fifty distinct projects with uses including sending mail, path tracking and indexing html pages.

**keyword indexer** \* - wrote a scoring algorithm for finding keywords on millions of arbitrary html pages, which was used to serve millions of targeted ads in real time, greatly increasing revenue.

File::CacheDir \* - wrote system that takes a ttl and rather arbitrary files, quickly, automatically and cronlessly cleans up itself

**engineer hiring and training** - performed technical interviews, trained new hires and oversaw continued training, taught a twelve week Perl class

Brigham Young University Mathematics Department

***Mathematics Instructor, September 1996 to January 1999, Fall 2003***

Instructed classes of thirty-five plus students in college algebra and first year calculus. Responsible for writing and grading tests, lecturing, and assigning grades

## Education

Brigham Young University

***Master of Science, Mathematics*** • *Thesis: "Computer Implementation and Experimentation of the Quadratic Sieve Factoring Algorithm"*

***Bachelor of Science, Mathematics - Bachelor of Arts, Honors Philosophy***

***University Honors Graduate*** • *Honors Thesis: "A Refutation of Thomas Morris's Two-Minds Model of the Incarnation"*

***Minor: Music***

## Training

**Architecting with Google Kubernetes Engine (2020)**

- Google Cloud Platform Fundamentals: Core Infrastructure
- Architecting with Google Kubernetes Engine: Foundations
- Architecting with Google Kubernetes Engine: Workloads
- Architecting with Google Kubernetes Engine: Production

**Machine Learning - University of Washington (2016)**

- Machine Learning Foundations: A Case Study Approach
- Machine Learning: Regression

**Data Science - Johns Hopkins / Coursera (2015)**

- The Data Scientist's Toolbox
- R Programming
- Getting and Cleaning Data
- Exploratory Data Analysis
- Reproducible Research
- Statistical Inference
- Regression Models
- Practical Machine Learning
- Developing Data Products

## Select Side Projects

CameraPanda • Awesome camera / lens shopping website, uses React-Redux, hosted Elasticsearch and Amazon hosting. Java and MySql for offline processing

fantasy sports • Performed real time analysis of schedules and results, and in fully automated fashion, nightly completed trades to maximize predicted points via a least squares regression system

MusicInsights.org • Looking for patterns in music, analyzing the Well-Tempered Clavier and Goldberg Variations of Bach, and hopefully finding many insights!

## Personal

- CPAN author
- accomplished pianist
- novice screenwriter
- Golden Age of TV fan
- blogger
- "funny"
- videographer

Served two years in England as a missionary for the Church of Jesus Christ of Latter-day Saints. Intramural champion in basketball, softball, volleyball, wallyball, badminton and sports trivia

Competed on the nationally ranked Brigham Young University Academic Quiz Team for four years. Earned national collegiate awards for team and individual debate

Enjoy traveling (48 states, 23 countries), photography, podcasts, games, watching movies, listening to music, singing, memorizing classical piano music and performing in piano recitals