

Ryan Chapin

www.ryanchapin.com <https://github.com/rchapin>

rchapin@nbinteractive.com 301-996-3714

West Friendship, Maryland

Technical leader with over 25 years of hands-on experience building large-scale, distributed systems both on-prem and in the cloud. Typically working with C-level stakeholders to clarify high-level business goals and then architect the system, write the code, deploy and run it in production. Sometimes assembling teams and leading developers, or working with a few other Senior Engineers to build everything out.

Seeking a position where the career path is taking on deeper and more involved technical challenges in an organization with a solid engineering culture that follows industry best practices and values engineers who are committed to building and operating high quality, reliable software.

Technologies

- **Languages:** Go, Python, Java
- **OS:** Linux (RedHat/CentOS/AlmaLinux, Debian, Ubuntu), Windows
- **Cloud:** AWS, GCP, BigQuery, Cloud Storage, Compute Engine, Deployment Manager, Pub/Sub, Source Repository
- **Big Data:** Hadoop, Hive, Avro, Kafka, Elasticsearch, Redis, RabbitMQ, Zookeeper
- **laC and Automation:** Terraform, Ansible, Python Fabric, SRE Best Practices
- **Observability:** OpenTelemetry, InfluxDB, Grafana, Datadog
- **SDLC:** Git, Maven, Gradle, Junit, Bitbucket Pipelines, Cloud Build, Nexus 3, Jira, Bitbucket, Confluence
- **Virtualization:** VMWare, Docker, Kubernetes, Helm
- **Database:** Oracle Db, MySQL, PostgreSQL, SingleStoreDB, Liquibase

Professional Experience

Palo Alto Networks, Inc.

September 2022 - Present

Principal Software Engineer - Prisma Cloud Platform

Palo Alto Networks is a cybersecurity company that provides next-generation firewalls, comprehensive cloud security, and AI based SOC automation tools.

- Designed, developed, and deployed Spring Boot based, distributed, microservice applications to provide core Prisma Cloud Platform services. Enabled unified, platform level, Cloud Asset resolution by developing SingleStoreDB pipelines and stored procedures to ingest 1.5+B records a day into the asset management data store.
- Developed REST based services to enable consumers to push data into and query data from the Cloud Asset Management system via horizontally scalable services.
- Developed Helm charts for deployment of microservice applications onto K8s clusters.

Hughes Network Systems, LLC

February 2016 - September 2022

Principal Software Engineer, April 2020 - September 2022

Hughes is a satellite ISP which provides managed network solutions for enterprises and governments managing and monitoring millions of devices.

- Enabled the Enterprise-wide standardization of data security, lineage, and cataloging by designing and developing a scalable, cloud-based, ETL pipeline for heterogeneous, schema managed, data streams to be written to GCP BigQuery. Developed a series of observable, microservice applications written in Go and deployed on Kubernetes with a combination of Terraform and Python. Releases are orchestrated with a quality gated, CI/CD pipeline comprised of BitBucket Pipelines, custom Python, Kustomize, and Cloud Build.
- Lead the Hughes initiative to the cloud by migrating on-prem Hadoop data and an ETL pipeline to GCP. Developed Java and Python ETL applications to ingest Avro data into BigQuery. Designed and developed custom Python applications to automate and ensure the migration of approximately 250TB to GCP.

Staff Software Engineer & Hadoop Developer, February 2016 - April 2020

- Reduced cost, increased performance, and enabled the business to leverage big data analytics by leading an enterprise-wide big data initiative. Architected, deployed, and maintained an on-prem, HA configured Hadoop cluster. Developed an ETL pipeline in Java and Python to replace costly, proprietary RDBMS systems. 1B records per day from 120 different schema are normalized, converted to Avro, and written to partitioned Hive tables.
- Enabled the development of streaming analytics by architecting, deploying, and maintaining an on-prem, SSL enabled, Kerberized Kafka cluster. Designed and developed an automated system for deploying and maintaining the cluster in Python. Implemented a monitoring and alerting system with Datadog and EM7. Developed a Kafka Java SDK that abstracts the configuration of a Kerberized client connecting to Kafka over SSL.
- Developed a generic Kafka-to-HDFSWriter Java library to read from a topic and append to files on HDFS which enables the persistence of streaming data for analytics and solves the Hadoop “small file” problem.
- Spearheaded the adoption of agile methodologies for all of EchoStar/Hughes by architecting and developing a highly-available, DR enabled, Atlassian suite to include Jira, Confluence, and Bitbucket integrated with corporate AD for SSO. Developed an automated system for deploying and maintaining each of the clustered components with a combination of Ansible and Python.

RedOwl Analytics, Inc.

April 2015 - January 2016

Lead Software Engineer

RedOwl provides a Big Data, ML/AI sentiment analysis platform to mitigate insider threat and provide tools for regulatory surveillance teams by ingesting and analyzing millions of events per day including emails, texts, chats, key-card swipes, and network traffic.

- Designed and developed a streaming email and chat ingest pipeline implemented with a combination of a custom Python Postfix content-filter, RabbitMQ, and a custom Java Elasticsearch plug-in.
- Lead teams on multiple high-profile projects. Worked with product owners, VP, and C-level management to plan and manage sprints, define user stories and acceptance criteria, and estimate and schedule tasks.
- Increased observability of production systems and enabled the ability to troubleshoot and optimize sub-systems by designing, developing, and deploying a standardized logging, metrics, and monitoring system that included log aggregation via rsyslog, Ganglia, and Nagios.

Financial Industry Regulatory Authority

May 2014 - April 2015

Senior Software Engineer/Hadoop Developer

FINRA ingests approximately 80B stock exchange events a day from NYSE and Nasdaq and runs surveillance queries on a daily basis to ensure compliance with US trading laws and regulations.

- Migrated the core surveillance query system to AWS EMR from on-prem, legacy Oracle, Netezza, and Greenplum enterprise RDBMS appliances by writing and revising complex, multi-staged, SQL queries as Hive HQL queries or custom Java MapReduce algorithms.
- Developed custom Hive UDFs and UDTFs in Java.
- Boosted team productivity by developing custom tools in Java, Bash, and Perl. One of the tools enables a distributed diff on terabytes or more of data.

Lotame Solutions, Inc.

March 2013 - May 2014

Lead Software Engineer

Lotame is an online ad-tech Data Management Platform (DMP) that processes 6B events per day, categorizing online browsing activity and behaviors providing data to online advertisers.

- Designed and wrote custom MapReduce jobs in Java running against a 1.6PB Hadoop cluster of approximately 3,000 cores. This involved Servlet development, protocol buffer design, MapReduce code, RDBMS schema design, and writing SQL stored procedures.
- Increased developer efficiency by creating a standard development environment for writing and testing MapReduce code for the core statistical analysis subsystem to include automated replication of a complete mirror of the production HDFS structure, RDBMS schema, and SQL stored procedures.

- Designed and developed an end-to-end solution for browser-based data collection which dynamically adapts at runtime to collect data. Implemented with a combination of JavaScript, Servlets, and custom back-end clusters of Java servers.
- Led the backend, Big Data team working directly with product managers, client facing teams, and VP level engineering stakeholders to groom the backlog, develop themes, epics, user stories, and tasks, and provide day-to-day sprint management.

Northrop Grumman Information Systems
Senior Software Engineer

April 2010 - November 2012

- Designed and developed a high-profile R&D project to ingest up to 500MB/sec of data into RDBMS and HDFS datastores and enable concurrent queries against that data at up to 200 queries/second which lead to the award of a \$500M contract. Developed a observability platform with a combination of custom Java applications, JMS, and SWING to provide a real-time metrics dashboard for 120 servers. Developed high volume, multi-threaded data generators in Java, and multi-threaded, dynamic query generators in Java to execute up to 5,000 queries per second.
- Architected and developed a real-time, behavioral based, zero-day, malware detection framework designed to stop e-mail phishing attacks based on observed behavior of potential malware as opposed to binary hashes. Developed a web based dashboard to provide real-time system status, metrics, and web based tools for malware analysis. Implemented with a JavaEE MVC system that featured a high-concurrency server-side layer and Javascript client-side components.

Patents

Calvin H. Smith, Kenneth Maclean, Jason J. Liu, Stephen Mann, Wendy Mann, Ryan Chapin, 2015. "**System and method for advanced malware analysis**". U.S. Patent 9,106,692 B2, filed January 31, 2013 and issued August 11, 2015.

Education

- Bachelors of Arts, cum laude University of Maryland, College Park Major: Art Studio with concentration in Illustration and Painting
- Certificate, East Asian Studies University of Maryland, College Park Concentration: Japan