Carmine DiMascio

Technical Leader and Principal Engineer

Hands–on tech–leader, architect, and inventor. Always learning

Email: cdimascio@gmail.com
Web: https://cdimascio.github.io
Git: https://github.com/cdimascio

PROFESSIONAL EXPERIENCE

Principal Engineer   2020–present
Amazon

Leads high-investment technology initiatives across the Amazon Pharmacy (AP) organization, aligning with director- and VP-level objectives. Owns systems supporting the AP buying experience, providing technical leadership to a group of ~30 engineers across 3 teams. Sets and aligns the technical vision, driving technical architecture decisions, and ensures the delivery of scalable, reliable, and customer-centric solutions.

Leads AP's customer-service and customer-facing generative-AI-chatbot initiatives, spearheading a multi-agent, Retrieval-Augmented Generation (RAG) solution. The system uses Elasticsearch vectorstores, LLMs, such as Anthropic Claude and Amazon Nova, and text embeddings powered by Titan models.

Drives critical and high-impact AP initiatives. including serving as the chief architect of RxPass, where I led 100+ engineers across 16 teams to deliver a HIPAA-compliant healthcare-subscription system on AP, expanding affordable prescription access for customers. Additionally, led 20-50 person teams on key initiatives such as integrating pharmacy benefit manager (PBM) using FHIR standards, building cash discount card solutions with 3p partners, delivering accurate claims-based insurance pricing, and developing machine-learning models for estimated insurance prices.

Principal Engineer   2019–2020
Tripadvisor

Led a 15-person cross-functional team to architect and implement a modern, scalable restaurant search solution for Tripadvisor. The system employs the CQRS pattern to independently scale reads and writes, leveraging multiple real-time data streams and pre-computed machine-learning recommendation models to deliver personalized results. It provides both RESTful and GraphQL interfaces and is powered by a robust backend, including Kafka, PostgreSQL, Memcached, and Elasticsearch.

Led the development of a custom service mesh leveraging go-control-plane, Envoy xDS, and gRPC, enabling teams to adopt off-the-shelf technologies while ensuring seamless integration with Tripadvisor's proprietary service stack and monolith. Later transitioned the solution to Istio, enhancing scalability and maintainability.

Team Lead / Architect   2014–2019
IBM Research

At IBM, I led several high-impact initiatives spanning research, product development, and strategic acquisitions.

IBM Watson Advertising
Led a 15-engineer team to create IBM Watson Advertising,, a conversational- A

PROFILE

I am a software engineer, mentor, and tech-leader with a focus on distributed systems, cloud engineering, and AI/machine-learning. With **30+ patents** and active contributions to open-source projects, I maintain **express-openapi-validator** (1.2M monthly downloads) and **dotenv-java** (500k monthly downloads), I bring deep expertise in building scalable, high-performance systems that drive technological innovation.

As a leader and mentor, I am passionate about empowering teams to succeed and fostering a collaborative, growth-driven environment. My expertise spans backend and frontend web and mobile development, cloud infrastructure, and distributed systems design. I'm a generalist with strengths in microservices, scalable architecture, and AI/ML, leveraging technologies such as Java, Python, Node, AWS, Kubernetes, SQL, NoSQL. I focus on delivering impactful solutions for customers, driving innovation, mentoring others, and building high-performing teams.

PATENTS

- IBM Plateau 8 inventor who has filed over 30 patents, including:
- System enabling audio-based navigation and presentation of a website
- Bootstrap a conversation service from a REST API

ads platform that delivers personalized premium ads on Weather.com. Under my leadership, Watson Ads evolved from concept to a multi-million-dollar product, earning recognition as a finalist for the MMA Technology and Innovation Award and MediaPost OMMA. Built on AWS, the platform serves ~5 million weekly requests, using a microservice architecture with Node.js, Java, Cloudant, Elasticsearch, and Watson.

IBM Research: Sesame Workshop & Bon Appetit Partnerships
For Sesame Workshop, I led the architecture and implementation of Watson Tutor for Early Childhood Vocabulary Learning, an API platform enabling personalized, adaptive learning experiences for young learners. I led a collaboration with Applied Scientists to develop an AI-driven phased-learning model using Python and Scikit-learn, EWMA, and historical performance data to guide students toward vocabulary mastery.

Led a team to build *Word Power Ups*, a Watson-Tutor-enabled iOS app featured on the Discovery Channel's *This is A.I.*, showcasing the platform's capabilities through vocabulary games.

I was a technical team lead for Chef Watson, leveraged by Kellogg and showcased on the *Today Show.* Chef Watson was an AI application capable of generating novel recipes optimized at the molecular level. The project leverages classic machine learning techniques, including SVMs and KNN alongside symbolic rules-based systems. It also leveraged NLP techniques to generate natural language recipe steps. Chef Watson, a winner of a Fast Company Design Award, was implemented in Java and Python, deployed as a CloudFoundry service on IBM Cloud.

Acquisition Technical Lead
Nominated by IBM Distinguished Engineer Steve Abrams and IBM Fellow Rob High, I led technical due diligence for IBM's acquisition of Alchemy API, evaluating its strategic fit. My work was crucial in influencing the acquisition decision. I also co-led technical due diligence for IBM's review and acquisition Blekko.

Full history <2014 available on https://www.linkedin.com/in/carmine

EDUCATION

Master of Arts, Computer Science	2004–2006
Boston University	

Bachelor's of Science, Computer Science	1996–2000
Providence College	
Magna Cum Laude	
Minor in Mathematics & Western Civilization	

- Generate a chatbot from an FAQ
- Context-specific view of a hierarchical data structure
- Cognitive Advertising triggered by weather data
- Direct payment system for web consumers
- Document Recommendation Method Based on Skill
- Digital content generation based on user feedback
- Determine logical breaks in a digital text

OPENSOURCE SOFTWARE

- express-openapi-validator (1m)
- dotenv-java (500k)
- dotenv-kotlin (150k)
- uuid-mongodb (40k)
- generator-express-no-stress
- py-readability-metrics
- essence
- react-native-camera
- More…

CONFERENCES

- IBM World of Watson 2016
- IBM Insight 2014 speaker
- IBM Innovate 2014 speaker
- Bluemix Meeting 2015 speaker
- Strata Hadoop 2015 lightning talk
- Vegas Uncorked 2015 (demo)
- IBM Innovate 2013 speaker
- EclipseCon 2012 (demo)
- IBM Innovate 2011 speaker