Rodrigo Palacios

Personal
rodrigopala91@gmail.com

GitHub
https://github.com/rodricios

Phone
(559) 930-0910

Senior Software Engineer | Solutions Architect | Web Data & Platform Systems

Solutions Architect and Senior Software Engineer with 7+ years of experience designing and scaling web data, crawling, and ETL systems for startups and large platforms (Medium, Localyze, TruePlan). Proven track record translating ambiguous customer and business requirements into production-ready architectures, onboarding clients, and supporting high-throughput data pipelines in distributed environments. Strong background in web data acquisition, analytics, and stakeholder-facing technical leadership.

Experience

Senior Consultant / Solutions Architect (Freelance). 2024 - current
- Designed a declarative Python-based web crawling framework to simplify large-scale web data acquisition
- Consulted for a startup building an ag-tech AI-first SaaS product
- Conducted whitehat security research for a major public gaming company (NDA); identified critical privacy vulnerabilities and secured bug bounty

Senior Backend Engineer, Data Engineer, Localyze. August 2022 - August 2023
- Designed and owned the data team's ETL architecture and local development workflows, enabling faster iteration and higher reliability across analytics pipelines
- Implemented an in-house billing and auditing system on top of Stripe, supporting finance and operations stakeholders with transparent reporting
- Partnered closely with non-technical teams to translate business KPIs into actionable dashboards and data products
- Technologies: RoR, PostgreSQL, Heroku, Google Cloud, Metabase, Airflow

Founding Engineer, TruePlan, Inc. (*Acquired by Localyze*). April 2021 - August 2022
- Acted as a founding engineer responsible for system architecture, data models, and customer onboarding for a growing SaaS platform
- Onboarded and supported 20+ customers, designing custom integrations and data workflows to meet client-specific requirements
- Led architectural migrations of ETL and observability systems to improve scalability and maintainability
- Mentored and onboarded new engineers as the team scaled to eight developers
- Technologies: RoR, PostgreSQL, GraphQL, Redis, Heroku, Datadog

Software Engineer, Medium, Inc. July 2019 - April 2021
- Designed and operated high-throughput data pipelines supporting recommendation and content delivery systems serving 100K+ requests per day
- Designed news-crawling system for an experimental Medium app that ingested 1K news articles per hour
- Collaborated across product, data, and engineering teams to ensure reliable data flow between services and customer-facing applications
- Technologies: React, Node.js, GraphQL, SQS, DynamoDB, Snowflake

Software Engineer, Evie, Inc. (*Acquired by Medium*). July 2016 - July 2019
- Built large-scale web crawling and ingestion systems processing tens of thousands of articles daily
- Built ML pipeline for news recommendation, entity extraction, and ranking
- Technologies: Python, Scrapy, Cassandra, Redis, Spark, Spark SQL, Mesos, Marathon, Chronos, Consul, Kubernetes, Protobuf

Research Experience

Research Intern, University of Notre Dame, summer 2015, advisor: Prof. Tim Weninger
- Conducted full-time research on stochastic graph generation under faculty supervision
- Designed and evaluated novel algorithms; results accepted for publication at CIKM

Research Collaborator (Remote), Fresno State & University of Notre Dame, Spring 2015, advisor: Prof. Tim Weninger
- Collaborated remotely with faculty on web content extraction research
- Designed experiments and datasets evaluating modern and legacy extraction systems
- Co-authored peer-reviewed publication accepted at ACM SIGKDD Explorations

Publications

Web Content Extraction - a Meta-Analysis of its Past and Thoughts on its Future[1]
Tim Weninger, Rodrigo Palacios, Valter Crescenzi, Thomas Gottron, and Paolo Merialdo
ACM SIGKDD Explorations Accepted. 2015.

Growing Graphs with Hyperedge Replacement Graph Grammars[2]
Salvador Aguiñaga, Rodrigo Palacios, David Chiang, Tim Weninger
CIKM Accepted. 2016.

Projects
Selected open-source projects focused on web data extraction and content analysis:

libextract[3] (500+ stars) - Libextract is a Python-based data extraction library that works on HTML and XML documents.
eatiht[4] (400+ stars) - Implemented novel data extraction algorithm to extract the main content from web pages automatically.
autocomplete[5] (400+ stars) - Lightweight Python autocompletion library focused on practical developer tooling.
wxpath[6] - A declarative Python-based web crawling and scraping library with XPath semantics.

Education

Fresno State - Computer Science, 2013 to 2016, 2025 - Present

Willow International, Fresno City & Fresno State (OpenU), 2010 to 2013

---

[1] https://arxiv.org/abs/1508.04066
[2] https://arxiv.org/abs/1608.03192
[3] https://github.com/datalib/libextract
[4] https://github.com/rodricios/eatiht
[5] https://github.com/rodricios/autocomplete
[6] https://github.com/rodricios/wxpath