German Neironi

Senior AI Engineer (FDE) at Marvell | Founder at CloudPruneAI and Quira | Solutions Architect at Argeniss

g.neironi@gmail.com | linkedin.com/in/gneironi | gneironiar.github.io/personal_website | github.com/gneironiar
Cordoba, Argentina (Open to Remote - US/EU/LATAM)

Professional Summary

Cloud and AI Solutions Engineer with 15+ years of experience designing and building cloud and AI systems in production - from architecture and IaC to enterprise deployment and multi-vertical AI enablement.

Currently Senior AI Engineer - Forward Deployed Engineer at Marvell Technology (via InterEx Group), owning 5 concurrent enterprise AI engagements across Finance, HR/People, Global Workplace Services, Video Creators Conference Support, and Claude Cost Monitoring & Alerting. Marvell is Nvidia's NVLink Fusion partner.

Founder of two SaaS products in production: CloudPruneAI (B2B2B FinOps platform with AI-powered IaC code generation, 18 analyzers, 31+ checks, real-time anomaly detection, 271-test suite, partner channel across 3 countries) and Quira (multi-tenant industrial IoT platform for environmental monitoring, launched April 2026). Both built solo from architecture to production, integrating Claude API for code generation, business context enrichment, and customer-facing features.

Expert in enterprise AI enablement (LLM Gateway design, RAG, Prompt Engineering), full-stack cloud development with AWS (CDK, Lambda, ECS Fargate, Bedrock, RDS, S3, Glue, Athena), FastAPI, Next.js, PostgreSQL, and Generative AI integration (Claude API, AWS Bedrock, MCP). Skilled in Python, TypeScript, Docker, and designing hybrid serverless architectures for production systems.

Also Solutions Architect / Pre-Sales Engineer at Argeniss Software, leading technical assessments and architectural proposals across travel tech and healthcare.

Professional Experience

Marvell Technology | Senior AI Engineer - Forward Deployed Engineer

June 2026 - Present | Contractor via InterEx Group | Remote

- Enterprise AI engineer within Marvell's Enterprise AI team (semiconductor company, Nvidia NVLink Fusion partner)
- Own 5 concurrent verticals and engagements: Finance (dashboards and analytics), HR/People (governance-first configuration on Workday HRIS), Global Workplace Services (real estate, facilities, EHS, sustainability), Video Creators Conference Support (AI video tools for presenter enablement), and Claude Cost Monitoring & Alerting
- Designed end-to-end solution architecture for Claude Cost Monitoring: LLM Gateway (source of truth) -> warehouse -> BI dashboards + alert engine + Copilot Kit self-service governance UI
- Author bilingual delivery documentation (ES working copy + EN VDI-ready) with vertical-specific training and templates, following a reusable playbook for onboarding new verticals
- Operate under a Forward Deployed Engineering model: scoping AI enablement per role and user rather than universal tooling

Tech Stack: LLM Gateway, Claude API, AWS Bedrock, Snowflake/BigQuery, Databricks, PowerBI, Copilot Kit, Anthropic Admin API, ServiceNow, Glean, Workday, MCP, RAG, Prompt Engineering

CloudPruneAI | Founder and Technical Lead

November 2025 - Present | Independent SaaS Startup | Remote

- Founded production FinOps SaaS platform analyzing AWS accounts and auto-generating Infrastructure-as-Code (CDK Python, Terraform HCL, CloudFormation YAML) for cost optimization using Generative AI and LLMs
- Built and deployed 18 resource analyzers covering 31+ optimization checks across compute, storage, networking, databases, commitments, extended support, and AI/ML services
- Shipped a production real-time anomaly detection engine (Q3 2026): 3-sigma against a trailing 30-day baseline, EventBridge daily cron, Lambda worker, alert delivery via Slack and Teams webhooks, per-account settings + history UI
- Added an AI cost tracking module for Bedrock, OpenAI, and Anthropic spend visibility (token-level tracking on the Q3+ roadmap)

- Architected hybrid serverless platform (FastAPI on Lambda + Mangum, ECS Fargate scan workers, async RDS PostgreSQL) with a 271-test suite and CI/CD via GitHub Actions OIDC (no long-lived credentials)
- Implemented AI-powered IaC code generation using Claude API (Anthropic) with multi-format output and business context enrichment via Claude Haiku
- Designed B2B2B partner-led GTM model with tiered pricing (Free / Assist EUR 4,900 one-time / Operate EUR 590/month): white-label PDF reports, gainshare (10% / 7.5% at scale) plus recurring subscriptions
- Onboarded 4 partners across Argentina, Spain, and Uruguay; first real scan with consulting partner detected USD 5,260/year in waste across a client AWS account
- GDPR-compliant architecture (Privacy Policy, DPA, ROPA, Security Posture Document, DELETE /users/me); USPTO trademark registration in progress
- MCP Server + Assistant Backend + Chat UI + Slack Bot bundle on the active roadmap

Tech Stack: AWS CDK, Lambda, ECS Fargate, RDS PostgreSQL (async), API Gateway, S3, SQS, Cost Explorer, EventBridge, FastAPI, Next.js 14, Material UI, Auth0, Stripe, Claude API, MCP, Python, TypeScript, Docker, Turbo, GitHub Actions OIDC

Quira | Founder and Technical Lead

December 2025 - Present | Independent SaaS Startup | Remote

- Founded multi-tenant industrial IoT platform for environmental monitoring (temperature, humidity), with real-time alerts, auditable PDF/CSV reports, and native Microsoft Teams integration via Power Automate
- Designed and deployed full production stack on AWS in under 4 months: 9 CDK stacks covering ECS Fargate, ALB, WAF, Route53, ACM, RDS PostgreSQL 15, Cognito, SQS, 6 Lambdas, and SNS
- Built FastAPI backend (47 endpoints, 103 tests) with HMAC-SHA256 + HKDF device authentication, async ingestion pipeline with idempotent client-side deduplication, alert engine with severity escalation, and async report generation via WeasyPrint workers
- Designed sensor-agnostic data model and driver registry to support multi-sensor expansion without schema changes (CO2, pressure, vibration on the V2 roadmap)
- Developed Next.js 14 admin panel with 25 pages, 6-section sidebar, i18n (ES/EN), and Recharts-based reading visualizations
- Validated end-to-end with real ESP32 + SHT31 hardware: device readings reach the platform, trigger alerts, and render in production charts

- Designed B2B2B revenue model: tripartite split between platform, partner, and vendor; first partner LOI signed in Colombia (aerospace warehouse pilot, ~30 devices)
- Production launch: April 19, 2026

Tech Stack: AWS CDK, ECS Fargate, ALB, WAF, Cognito, RDS PostgreSQL 15, SQS, Lambda, SNS, Route53, ACM, KMS, FastAPI, Next.js 14, Material UI, next-intl, ESP32 (Arduino C++), SHT31, HMAC-SHA256, Python, TypeScript, PostgreSQL, Resend

Argeniss Software | Solutions Architect / Pre-Sales Engineer

March 2026 - Present | Multiple Clients (Travel-tech, Healthcare) | Remote

- Technical advisor in pre-sales engagements supporting the COO during client discovery and proposal phases
- Lead technical assessments, architecture proposals, and stack recommendations for early-stage and pre-seed clients across travel tech and healthcare
- Created Technical Assessment Playbook (Discovery -> Assessment -> Proposal) standardizing the pre-sales workflow
- Travel-tech CRM: designed a Reviewer-of-Record engagement model for a product built with AI coding agents (architecture review of critical paths + production support with severity-based SLAs)
- Healthcare (homecare AR): authored a staffing-model proposal for phased ERP modernization on AWS (Postgres operational + Redshift analytical domains, Cognito + SSO, PWA-first)
- Established reusable proposal patterns: AWS-first, Cognito over custom JWT, ETL batch (Glue) vs CDC managed, dual operational/analytical domains
- Generate additional engineering scope for delivery teams (PM + Architect + Engineer) by anchoring trust through technical depth

Tech Stack: AWS, Cloud Architecture, Solution Design, Technical Discovery, Stack Selection, Healthcare Interoperability, Multi-tenant SaaS Design

Argeniss Software | Senior Cloud and AI Engineer

January 2026 - Present | Multiple US Clients (confidential) | Remote

End-to-end AWS + AI systems for US enterprise and startup clients across pharma AI, consumer finance, healthtech, and industrial verticals (client identities under NDA).

- Multimodal clinical inference platform (pharma AI): event-driven pipeline (S3 -> EventBridge -> Step Functions) with 4 layers - multimodal normalize, FHIR encoding to ICD-10/RxNorm/SNOMED, LLM patient-journey canonicalization, and gold feature matrix - across 6 CDK stacks with governance-as-code (CDK Aspects, OIDC CI/CD, KMS CMKs, DLQs); HIPAA/PHI compliance-by-design (Bedrock BAA, zero data retention, PHI redaction, surrogate patient keys)
- Consumer-finance data platform: Terraform-based replication of SQL Server + MongoDB to S3, medallion architecture in Redshift (RAW via Spectrum / SILVER views / GOLD materialized), config-driven extraction (DynamoDB catalog, full/incremental/change-tracking) with per-row SHA-256 diff; heavy reporting cut from minutes to milliseconds; passed client security review with zero open findings
- Veterans-benefits conversational form-fill assistant (healthtech): multi-agent LLM system on AWS Bedrock AgentCore (Claude Sonnet) over WebSocket API Gateway + DynamoDB single-table; owned a 48-hour pre-demo hardening cycle with atomic PRs, smoke tests, and zero rollbacks
- Labor-forecasting SaaS (US mechanical contractor): full-stack app (FastAPI Lambda + React SPA + RDS Postgres in a private VPC + Cognito RBAC + SOAP ERP integration + custom MCP server for natural-language scenario planning + Athena BI) across 8 CDK stacks, replacing an Excel-with-macros workbook as the single source of truth
- Applied governance-as-code and least-privilege security across engagements: CDK Aspects, OIDC CI/CD, KMS CMKs, DLQ + review queues, SSM Parameter Store secrets

Tech Stack: AWS CDK (Python), Terraform, Lambda (Docker), Step Functions, S3, EventBridge, DynamoDB, API Gateway (HTTP + WebSocket), Cognito, Bedrock AgentCore, Bedrock (Claude Sonnet), Comprehend Medical, Textract, Glue (PySpark), AWS DMS, Redshift (Spectrum, medallion), RDS Postgres, Athena, KMS, CloudFront, WAFv2, VPC + NAT, custom MCP servers, FastAPI, React 18, TypeScript, GitLab CI/CD, GitHub Actions OIDC, FHIR / ICD-10 / RxNorm / SNOMED

Argeniss Software | AWS Data Integration Engineer

October 2025 - Present | Fintech Client | Remote

- Architected automated DMS system replicating 10-30 SQL Server databases to Redshift with CDC (30-120 sec latency)
- Built reusable Terraform modules (DMS instance, endpoints, KMS) following Infrastructure as Code best practices
- Designed multi-source architecture supporting SQL Server (DMS), MongoDB (Change Streams), REST APIs (Salesforce, NetSuite)

- Developed orchestration system with Step Functions and Lambda for zero-touch database onboarding (3 commands)
- Implemented DynamoDB-driven configuration catalog and S3 data lake for semi-structured data
- Deployed multi-AZ, Security Hub compliant infrastructure with automated KMS rotation
- Created Redshift Spectrum integration for querying S3 data alongside relational tables
- Delivered 60-70% infrastructure reusability, reducing onboarding time from weeks to days

Tech Stack: AWS DMS, Lambda, Step Functions, DynamoDB, S3, Redshift, KMS, Terraform, Python, SQL Server, MongoDB, REST APIs

Argeniss Software | AWS Cloud and AI Engineer

February 2025 - October 2025 | Multiple Clients (E-learning, Healthcare) | Remote

Concurrent client engagements covering data engineering, AI/LLM integration, and machine learning forecasting.

- Architected complete ETL system extracting student performance data from MySQL to S3 Data Lake with Parquet optimization
- Built dynamic REST API with API Gateway, Lambda, and Athena for configurable analytics endpoints
- Developed LLM-powered analysis system using AWS Bedrock (Claude 3.5 Sonnet) and Prompt Engineering for personalized educational recommendations (RAG pattern)
- Implemented 6 production Lambda functions (containerized) handling ETL, API queries, LLM analysis, and NCLEX question generation
- Created real-time update pipeline with SQS for user-specific data refreshes
- Designed Vue.js frontend with CloudFront for NCLEX question generator with iterative AI feedback
- Deployed multi-stack CDK infrastructure with automated Glue Crawlers and EventBridge schedulers
- Optimized query costs with Athena partitioning strategy, achieving sub-second API response times
- Built API data extraction and transformation pipelines with automated JSON-to-Parquet conversion and partitioned storage
- Implemented secure credential management with Parameter Store and S3 encryption
- Implemented machine learning forecast infrastructure using SageMaker Canvas with dual prediction system for short-term (180 days) and long-term (12 months) forecasting

- Designed multi-model Generative AI pipeline using AWS Bedrock with 4 concurrent models (Claude Sonnet, Haiku, Nova Lite, Nova Pro), orchestrated by Step Functions in a 3-stage LLM workflow
- Implemented automated CSV processing handling 300+ records with EventBridge triggers
- Set up GitHub Actions CI/CD pipeline with multi-environment deployment

Tech Stack: AWS CDK, Lambda (Docker), API Gateway, AWS Bedrock, Claude API, Athena, Glue, S3, DynamoDB, SQS, CloudFront, EventBridge, Secrets Manager, Parameter Store, Step Functions, SageMaker Canvas, GitHub Actions, Python, MySQL, Parquet, Vue.js, Docker

Scale AI | AI Quality Assurance Engineer

August 2023 - September 2024 | Remote

- Trained generative AI models by performing code review of LLM output
- Created code snippets with reasoning to improve AI model responses
- Identified and fixed bugs to enhance model performance
- Supported prompt engineering and quality standards for Generative AI systems

Tech Stack: Python, NumPy, Pandas, Scikit-learn, Keras, PyTorch, Docker, AWS

NTB Solutions | Senior Data Analyst

August 2018 - February 2024 | Remote

- Worked on data analysis and integration projects for healthcare clients across multiple countries
- Built ETL processes using Azure Data Factory, Notebooks, Scala, and Python
- Created BI dashboards using Power BI, Pentaho, and Tableau
- Designed and implemented health information systems (HIS, LIS, RIS)
- Set up HL7-FHIR health interoperability processes
- Achievement: Built custom report recovering billions of pesos annually in previously uncontrolled funds

Tech Stack: Azure Data Factory, Python, Spark, PySpark, SQL, Pentaho, Power BI, MIRTH Connect, MySQL, Oracle, PostgreSQL, MongoDB

Anyone AI | Machine Learning Engineer (Part Time)

July 2022 - December 2022 | Remote

- Developed sentiment analysis, image classification, credit risk analysis, and salary prediction models
- Built and trained ML models using Python and various libraries
- Integrated models into API-based web services with Docker deployment

Tech Stack: Python, Scikit-learn, Keras, PyTorch, TensorFlow, Docker

Sanatorio la Entrerriana | Data Analyst and IT Manager

July 2008 - August 2018 | Parana, Argentina

- Led multi-year implementation of integrated HIS/LIS/RIS health information system across the institution, achieving full staff adoption and replacing legacy paper-based workflows
- Created executive dashboards and operational reports using Power BI and Google Data Studio
- Managed Oracle databases and led data-driven decision-making processes for clinical and administrative operations

Tech Stack: Oracle, SQL, Power BI, Google Data Studio, ETL processes

Core Technical Skills

Cloud and Infrastructure:
AWS (CDK, Lambda, Step Functions, ECS Fargate, ALB, WAF, Cognito, ACM, Route53, SNS, DMS, Bedrock, S3, Glue, Athena, Redshift, DynamoDB, SQS, KMS, EventBridge, Secrets Manager, CloudFront, SageMaker), Terraform, Azure

Enterprise AI and LLMs:
LLM Gateway Design, AWS Bedrock, Bedrock AgentCore, Multi-Agent LLM Systems, Claude API (Anthropic), Anthropic Admin API, Prompt Engineering, RAG (Retrieval-Augmented Generation), MCP (Model Context Protocol) Servers, AI Agents, Generative AI Multi-model Orchestration, Comprehend Medical, Textract, SageMaker Canvas, Copilot Kit

AI Enablement Tools (Enterprise Context):
Glean, Databricks, PowerBI, Snowflake/BigQuery patterns, ServiceNow, Workday HRIS context

Programming and Frameworks:
Python, TypeScript, SQL (T-SQL, PL/SQL), JavaScript, FastAPI, Next.js 14, React 18, Vue.js, R, Arduino C++ (ESP32), Embedded Systems

Data Engineering:
ETL/ELT, Medallion Architecture (RAW/SILVER/GOLD), Redshift Spectrum, Parquet, JSON, Pandas, PySpark, Apache Spark, Data Pipeline Design, CDC / Change Tracking

Healthcare Interoperability:
HL7, FHIR, ICD-10, RxNorm, SNOMED, HIPAA/PHI compliance-by-design

Databases:
PostgreSQL, MySQL, SQL Server, MongoDB, Oracle, Amazon RDS, Amazon Redshift, PostGIS

DevOps and Tools:
Docker, Git, GitHub Actions (OIDC), CI/CD, Infrastructure as Code (CDK, Terraform), AWS Amplify, Vercel, Jira

Business Intelligence:
Power BI, Tableau, Pentaho, QuickSight, SSRS

Key Projects and Achievements

Marvell Enterprise AI Delivery:
Own 5 concurrent enterprise AI verticals at Marvell Technology (Nvidia NVLink Fusion partner). Designed Claude Cost Monitoring architecture (LLM Gateway -> warehouse -> BI + alerts + self-service governance UI). Author bilingual delivery docs and a reusable vertical onboarding playbook.

CloudPruneAI Platform:
Production B2B2B FinOps SaaS with 18 analyzers, 31+ optimization checks, and AI-powered IaC code generation in 3 formats (CDK, Terraform, CloudFormation). Real-time anomaly detection engine (3-sigma, EventBridge cron, Slack/Teams delivery). 271-test suite with GitHub OIDC CI/CD. 4 active partners across Argentina, Spain, and Uruguay. First real scan detected USD 5,260/year in client savings. USPTO trademark registration in progress.

Quira IoT Platform:
Multi-tenant industrial IoT platform for environmental monitoring, launched April 2026. 9 CDK stacks on AWS, 47 backend endpoints, end-to-end validated with real ESP32 hardware. First partner LOI signed in Colombia (aerospace pilot).

Enterprise AWS + AI Delivery (Argeniss, US clients under NDA):
Four production engagements across pharma AI, consumer finance, healthtech, and industrial verticals: a multimodal clinical inference platform (Bedrock + FHIR/ICD-10/RxNorm/SNOMED, 6 CDK stacks, governance-as-code, HIPAA/PHI by design); a consumer-finance data platform (Terraform, medallion Redshift, config-driven extraction with per-row diff, multi-thousand-x reporting speedups, security review with zero open findings); a multi-agent LLM form-fill assistant on Bedrock AgentCore; and a full-stack labor-forecasting SaaS with a custom MCP server (8 CDK stacks).

Multi-Source DMS Architecture:
Automated replication system supporting SQL Server, MongoDB, and REST APIs with 60-70% infrastructure reusability.

Educational Analytics Platform with LLM Integration:
Complete ETL system with RAG-based LLM analysis, dynamic REST API, and NCLEX question generator using AWS Bedrock.

Healthcare Data Recovery:
Custom reporting solution recovering billions of pesos annually in previously uncontrolled healthcare funds.

Education and Certifications

Bachelor's Degree in Business Management
Universidad Autonoma de Entre Rios (UADER), Parana, Argentina

Certifications:

- AWS Certified Cloud Practitioner
- AWS Cloud Services (Kinesis, EMR, Glue, Athena, QuickSight, Redshift)
- Azure Fundamentals and Data Engineering (Microsoft)
- Machine Learning Engineer - Anyone AI
- C1 English Language - EFSET
- Scrum Foundation Professional

- ITIL Foundation Certificate in IT Service Management

In Progress:

- AWS Certified Solutions Architect - Associate (SAA-C03)

Soft Skills

- Strategic Thinking and Problem Solving
- Team Leadership and Collaboration
- Continuous Learning and Growth Mindset
- Strong Written and Verbal Communication in English
- Detail-Oriented and Goal-Driven
- Remote Work Excellence
- Cross-Functional Team Coordination
- Stakeholder Management