

JUAN MONSALVE

INGENIERO EN SISTEMAS E INFORMÁTICA | SOFTWARE ENGINEER — BACKEND, CLOUD E IA APLICADA

Bucaramanga, Colombia | +57 3175259997 | JuanMonsalve.23@hotmail.com | github.com/Moonsalve | linkedin.com/in/jmonsalve

PERFIL PROFESIONAL

Ingeniero con base sólida en Python, Java y JavaScript, especializado en desarrollo backend, APIs REST y sistemas con IA aplicada. Experiencia diseñando arquitecturas escalables, optimizando bases de datos e implementando soluciones en contenedores con Docker y AWS. El trabajo reciente se centra en productos de IA que corren en el propio equipo: integración de modelos de lenguaje, ruteo de intenciones por embeddings, y las decisiones de latencia, costo y seguridad que determinan si un modelo debe estar siquiera en la ruta crítica.

COMPETENCIAS TÉCNICAS

Lenguajes de programación

Python, Java, JavaScript/TypeScript, SQL, Bash

IA y aprendizaje automático

Integración de modelos de lenguaje (Ollama, inferencia local), embeddings de frases y ruteo semántico, reconocimiento de voz (Whisper), llamada a herramientas con esquema validado, ONNX Runtime / CUDA

Nube y DevOps

AWS (EC2, S3, Lambda), Azure, Docker, Git/GitHub, CI/CD, Linux

Backend y APIs

Node.js, FastAPI, Express, APIs REST, arquitectura de microservicios

Bases de datos

PostgreSQL, MySQL, MongoDB, Redis

Prácticas

Desarrollo guiado por pruebas, contratos tipados (Pydantic, TypeScript), documentación técnica, Agile/Scrum

PROYECTOS DE SOFTWARE

APOLLO — Asistente de voz en español, 100 % local

Proyecto personal

Python 3.11 · Whisper · Sentence Embeddings · Ollama · Pydantic

2026

- Construí un asistente de voz en español que funciona sin conexión y por completo en el equipo del usuario: **Silero VAD**, **faster-whisper large-v3-turbo** sobre CUDA, **Piper TTS** y un modelo de lenguaje local servido con **Ollama**. Ningún audio sale de la máquina.
- Diseñé un **ruteo de intenciones en cuatro etapas** (literal → patrones → similitud por embeddings → modelo como último recurso) que resuelve la mayoría de las órdenes en **~2 ms** y deja fuera de la ruta crítica la llamada al modelo, de 250 a 500 ms; medí un **turno completo de 0,135 s** (voz a texto 0,121 s / ruteo 0,006 s / ejecución 0,008 s) en el equipo de destino.
- Diagnosticué y corregí una falla de ruteo semántico en la que los puntajes de casos válidos e inválidos **se solapaban** bajo cualquier umbral absoluto; lo resolví centrando los embeddings e introduciendo una clase negativa `_fallback`, que convierte la decisión de absoluta en relativa: de inservible a **24/24 positivos y 8/8 negativos**.
- Limité la ejecución del modelo con una **lista blanca de herramientas y esquemas Pydantic estrictos** (`extra="forbid"`): solo puede emitir llamadas validadas, nunca comandos de shell ni rutas libres. Cubierto por **307 pruebas unitarias**, con el ruteo y las habilidades comprobables fuera del equipo final.

NovaTrack — API backend escalable para gestión de tareas

Proyecto personal

- Diseñé y desarrollé una API REST de alto rendimiento con **Python (FastAPI)** y **PostgreSQL**, con autenticación de usuarios (JWT) y control de acceso por roles.
- Empaqueté la arquitectura de microservicios con **Docker** y orquesté el despliegue en **AWS EC2**, con 99,9 % de disponibilidad.
- Construí canalizaciones de CI/CD automatizadas con **GitHub Actions** para agilizar las pruebas y el despliegue, reduciendo en un 30 % el ciclo de publicación.

- Documenté la API con Swagger UI para facilitar la integración desde el frontend.

EM TORNEOS — Plataforma de gestión de ligas (en producción)

Proyecto personal

Node · TypeScript · Prisma · PostgreSQL · Electron · Docker · Caddy

2026

- Construí y **opero en producción** un sistema en dos repositorios, una **API en Node/Express** tipada y un cliente de escritorio en **Electron**, que una liga de fútbol amateur usa a diario; **22 modelos de dominio** que cubren torneos, partidos, deudas, multas, abonos y nóminas.
- Desplegado con **docker compose** en un servidor propio: **Caddy** resolviendo y renovando el certificado sin intervención, **PostgreSQL** sin puerto publicado y alcanzable solo por la red interna, y migraciones que corren al arrancar el contenedor.
- Implementé **respaldos diarios y un script de prueba de restauración** que carga el último en una base aparte y verifica los conteos: un respaldo que nunca se restauró no es un respaldo.
- Publico **versiones de escritorio que se actualizan solas** desde ese mismo servidor, con la dirección de la API compilada en el instalador para que el operario instale e inicie sesión sin configurar nada; los permisos siguen al oficio, no a la jerarquía.

EXPERIENCIA PROFESIONAL

Team Manager — Operaciones y análisis de datos

Nov 2025 – Ago 2026

WW Funcrafters JWA LLC

Remoto

- Lideré un equipo multidisciplinario de más de 15 personas, a cargo de las operaciones técnicas, el aseguramiento de calidad y el cumplimiento de indicadores.
- Automaticé el reporte semanal de desempeño con **Python (Pandas)** y SQL para analizar métricas de productividad, aumentando en un 40 % la eficiencia del proceso.
- Gestioné casos técnicos escalados con análisis de causa raíz, manteniendo por encima del 95 % el cumplimiento de los objetivos operativos.

Customer Sales & Support

Feb 2024 – Nov 2025

WW Funcrafters JWA LLC

Remoto

- Atendí ventas y soporte a clientes de Estados Unidos en inglés, en un entorno remoto y de alto volumen.
- Ascendí a Team Manager en noviembre de 2025, asumiendo las operaciones y el reporte del mismo equipo.

Customer Technical Support Engineer

Dic 2023 – Sep 2024

Concentrix Web Help

Remoto

- Diagnostiqué y resolví errores complejos de aplicación mediante análisis sistemático de causa raíz y depuración de registros.
- Trabajé con los equipos de ingeniería L2/L3 documentando pasos reproducibles y registros técnicos de los fallos, acortando el tiempo de resolución.
- Cumplí de forma sostenida los indicadores de calidad y tiempo de respuesta en un entorno de alto volumen, adaptándome a actualizaciones y cambios de arquitectura continuos.

FORMACIÓN Y CERTIFICACIONES

Ingeniería en Sistemas e Informática

2023 – 2026

Universidad Pontificia Bolivariana

Bucaramanga, Colombia

Certificaciones y cursos: AWS Cloud Practitioner Fundamentals, fundamentos de Linux, programación en Python, SQL y bases de datos, Git y GitHub, Agile y Scrum.

Idiomas: español (nativo), inglés (avanzado, C1+).