

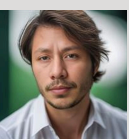
Erick Manuel Sang Cifuentes

Data Engineer | 5+ Years Experience in ETL/ELT Pipelines, Cloud Architecture & AI/ML Integration

Data Engineer with 5+ years of proven experience designing, building, and optimizing scalable ETL/ELT pipelines, cloud data architectures, and enterprise data warehouses. Expert in Python, SQL, Apache Spark, and Apache Airflow with demonstrated success in cloud migrations, process automation, and AI/ML integration. Proven technical leader with experience managing cross-functional teams and delivering data solutions for financial services, education, and enterprise clients.

Technical Expertise:

- ETL/ELT Pipeline Architecture & Optimization • Multi-Cloud Solutions (AWS, Azure, GCP) • Data Warehouse Design & Implementation
- Real-time Data Processing & Stream Analytics • RAG Architecture & Vector Databases • Machine Learning Operations (MLOps)
- Database Performance Tuning & Optimization • Financial Data Compliance (PCI-DSS) • Team Leadership & Project Management



📍 KR 23 A Sur # 1 - 140, Madrid, Colombia 📞 +57 314 272 8379 @ ericksang@gmail.com 🌐 Portfolio

[esangc](#) [leonsang](#) [ericksang.dev](#)

Professional Experience

SohoHumantech

Data Engineer - Consulting
Bogotá, Colombia

June 2025 - August 2025 (3 months)

- Developed and implemented comprehensive analytics panel for GASCO Chile using GCP Looker, React, Next.js, and BigQuery, resulting in 40% improvement in data query response times
- Designed scalable data integration architecture and built interactive frontend for corporate API consumption, serving 500+ daily active users
- Established CI/CD workflows using Git and SonarQube, reducing code defects by 60% and accelerating deployment cycles
- Collaborated directly with GASCO Chile technical teams to gather requirements and deliver enterprise-grade Analytics Panel solution

Corporación Unificada Nacional (CUN)

Data Engineer - Technical Lead
Bogotá, Colombia

February 2025 - June 2025 (4 months)

- Led implementation of Single Source of Truth (SSOT) architecture on AWS using Apache Airflow, dbt, Great Expectations, and S3, reducing data access time by 75%
- Architected and deployed CUN 360 conversational AI bot using RAG architecture with pgVector, N8N, and Docker microservices, serving 20,000+ university users
- Built end-to-end data pipeline using Airbyte for ingestion, Apache Spark for processing, PostgreSQL for storage, and S3 for data lake operations
- Managed technical team of 4 engineers and coordinated with Software Factory team to ensure seamless integration with existing university systems

FACTECH

Data Engineer - Project Lead
Bogotá, Colombia

August 2024 - January 2025 (6 months)

- Led 6-person engineering team in complete automation of MEN self-assessment and qualified registry processes, reducing manual processing time by 80%
- Implemented robust ETL pipelines using Apache Airflow, Python, and SQL, achieving 95% reduction in data processing errors and 50% improvement in processing speed
- Successfully migrated legacy systems to AWS cloud architecture (RDS/Aurora), improving system uptime to 99.9% and reducing infrastructure costs by 35%
- Deployed containerized applications using Docker and established CI/CD pipelines, reducing deployment time from hours to minutes
- Delivered end-to-end solution from data capture to automated reporting with built-in data quality validations using Great Expectations

Teleperformance

Data Specialist
Bogotá, Colombia

December 2021 - July 2024 (2 years 8 months)

- Designed and implemented enterprise data warehousing pipelines processing 10M+ daily transactions for multiple financial sector clients, improving data processing efficiency by 60%
- Built scalable ETL/ELT solutions using SQL Server, Python, and Apache Airflow, supporting high-volume transactional data with 24/7 production monitoring and 99.95% uptime
- Developed Power BI dashboards and automated reporting systems serving 200+ business users across multiple client verticals
- Ensured PCI-DSS compliance and implemented data encryption protocols for sensitive financial data, maintaining 100% regulatory compliance
- Performed advanced data analysis and created KPIs that directly influenced strategic business decisions for C-level executives

Nexabpo

Senior Visualization Architect
Bogotá D.C. Metropolitan Area

October 2019 - March 2022 (2 years 6 months)

- Developed custom data models and scalable architectures for enterprise clients, supporting data volumes growing 300% annually while maintaining optimal performance
- Implemented ETL pipelines integrating 15+ disparate data sources, eliminating information silos and creating unified business intelligence platform
- Led legacy system migrations from on-premise to cloud platforms, achieving 99.5% data integrity and reducing operational costs by 45%
- Generated actionable KPIs and business metrics that enabled data-driven decision making for executive leadership across multiple client organizations

Gaviria & Borbón Asociados

Systems Coordinator
Bogotá, Colombia

June 2018 - July 2019

- Coordinated and maintained enterprise technology services supporting 200+ users, ensuring 99.8% system uptime for critical business operations
- Implemented comprehensive backup and disaster recovery protocols, reducing potential data loss from hours to minutes
- Established IT governance framework and business continuity procedures, improving incident response time by 70%
- Managed software and hardware platform operations across multiple locations, optimizing resource allocation and reducing IT costs by 25%

Featured Projects

CUN 360 Bot - Conversational RAG System

Conversational AI system with operational RAG architecture

2025

Technologies: Python, pgVector, PostgreSQL, AWS, Docker, N8N, Apache Airflow

Impact: Deployed production RAG system serving 20,000+ university users with 95% query accuracy and sub-2-second response times

Achievement: First implementation of enterprise RAG architecture in Colombian educational sector, reducing student service response time from 24 hours to real-time

RAG Architecture, Conversational AI, pgVector, PostgreSQL, AWS, Docker, Python

Automated ETL Pipeline - FACTECH

Complete automation of educational processes with Apache Airflow

2024-2025

Technologies: Apache Airflow, Python, SQL, Docker, AWS RDS/Aurora, Great Expectations

Leadership: Managed 6-person engineering team across full project lifecycle

Results: Delivered 95% error reduction, 80% faster processing, and seamless cloud migration with zero data loss affecting 10,000+ educational records

Apache Airflow, ETL Pipeline, Python, SQL, AWS, Cloud Migration, Team Leadership

Data Analytics Platform - GASCO Chile

Enterprise analytics panel with GCP and modern technologies

2025

Technologies: GCP Looker, BigQuery, React, Next.js, SonarQube, Git CI/CD

Client Impact: GASCO Chile enterprise analytics serving 500+ daily users

Performance: Achieved 40% improvement in query response times and built scalable frontend consuming corporate APIs with 99.9% uptime

GCP, Looker, BigQuery, React, Next.js, Data Analytics, Corporate Integration

Financial Data Warehouse - Teleperformance

Financial sector data warehouse with PCI-DSS compliance

2021-2024

Technologies: SQL Server, Power BI, Python, Apache Airflow, ETL/ELT Pipelines

Scale: Processing 10M+ daily transactions across multiple financial verticals with 99.95% uptime

Compliance: Maintained 100% PCI-DSS compliance with encrypted data processing for sensitive financial information serving 200+ business users

SQL Server, Power BI, ETL/ELT, Financial Data, PCI-DSS, Data Warehouse, Python

Education

Politécnico Grancolombiano

Bogotá, Colombia

Eighth semester approved

Software Engineering

January 2023 - Present

Specialized training in **Software Architecture, Database Systems, Cloud Computing, Machine Learning**. Focus on modern technologies and enterprise software development.

ECCI

Bogotá, Colombia

Graduation option

Technology in Automation and Robotics

March 2017 - June 2021

Technology focused on **Control Systems, Industrial Automation, Robot Programming** and technical fundamentals in automation.

Technical Skills

Data Engineering & Analytics

5+ years continuous experience

████████████████████

Python, SQL, Apache Spark, Apache Airflow, ETL/ELT Pipelines, Data Pipeline Architecture, Data Modeling, Data Warehousing, RAG Architecture, Vector Databases, pgVector, Machine Learning, N8N Workflows, Google A2A Framework, SDK Development, pytest, ETL Testing

Cloud Platforms

Multi-cloud practical experience

██████████████████

Amazon Web Services (AWS), AWS S3, AWS RDS, AWS Aurora, AWS Redshift, AWS Lambda, AWS EC2, Microsoft Azure, Azure Data Factory, Azure Synapse, Google Cloud Platform (GCP), BigQuery, Looker, DataProc, Great Expectations, Airbyte, dbt, Docker, Kubernetes, Terraform, Infrastructure as Code

Databases & Storage

5+ years specialized

████████████████████

SQL Server, PostgreSQL, Oracle Database, MongoDB, pgVector, Amazon RDS, Amazon Aurora, MySQL, Snowflake, Database Performance Tuning, Database Administration

BI & Visualization

3-4 years experience

██████████████████

Power BI, Tableau, Looker, LookerML, SAS VIYA, SAS Guide, Data Visualization, Dashboard Development, Business Intelligence, Reporting Automation

Specialties & Leadership

5+ years main focus

████████████████████

Technical Leadership, Project Management, Team Management, Agile/Scrum Methodologies, Process Automation, Performance Optimization, System Architecture, Problem Solving, Financial Compliance, PCI-DSS, Data Governance, Quality Assurance

Soft Skills & Communication

Professional competencies

████████████████████

Cross-functional Collaboration, Stakeholder Management, Requirements Gathering, Technical Documentation, Mentoring & Coaching, Client Communication, Presentation Skills, Critical Thinking, Adaptability, Time Management, Conflict Resolution, Strategic Planning

Certifications

Associate Data Engineer in SQL

DataCamp (DE101)
2024

Comprehensive certification covering database design principles, ETL process implementation, data quality frameworks and advanced data visualization techniques

Data Engineer in Python

DataCamp (DE201)
2024

Advanced certification in cloud-native data engineering tools, scalable pipeline architecture and production-level code implementation practices

SQL Associate

DataCamp
2024

Specialized certification in complex query optimization, performance tuning and enterprise database architecture design

Scrum Fundamentals Certified (SFC™)

SCRUMstudy
2024

Agile methodologies and project management

Python for Data Engineering

LinkedIn Learning
2024

Advanced Python programming, data structures, and libraries for data engineering applications

SQL Complete Certification Suite

SoloLearn
2024

Comprehensive SQL training covering basic to advanced concepts

Microsoft Power BI Certification

LinkedIn Learning
2024

Business Intelligence, data visualization, and dashboard development

Big Data Analytics Certificate

Universidad de los Andes
2024

Advanced data analysis methodologies and big data processing techniques

Integrated Data Analysis Program

TalentoTech MINTIC
2024

Government-certified comprehensive data analysis and visualization program

Languages

Spanish

Native

████████████████████

English

Intermediate (B2)

██████████████████

Technologies in Development

Data Science & ML Integration

Machine Learning Pipelines, MLOps & Model Deployment, Data Science Automation, ML Model Integration

Automation & Optimization

Process Automation, Performance Optimization, Scalability Solutions, Problem Solving Frameworks

Cloud & Advanced Architectures

AWS Solutions Architect (in progress), Azure Data Engineer Associate (in progress), GCP Professional Data Engineer (in progress), Lambda & Kappa Architecture, Real-time Analytics, Distributed Systems