

# Héctor Eduardo Neri Cano

## Big Data Engineer

<https://www.linkedin.com/in/hectorneri>

[neri.cano@gmail.com](mailto:neri.cano@gmail.com)

<https://nericano.mx/cv>

## Summary

I'm a Big Data Engineer with over 7 years of experience with Big Data Technologies. I worked as a Representative for GCP where I Certified as Google Cloud Architect and Data Engineer. After that, I've been performing and leading Data Engineering efforts for achieving scalable, near-real-time systems and processes to create and harness analytical data products.

I enjoy to dig deeper and cover a wide range of solutions, from building data pipelines to building frameworks, from automating data transformations to designing the layout of a data lake, from connecting components within an architecture to defining infrastructure as code, from applying regex seeks and replacements to managing the workflow orchestrator, from encrypting specific fields in nested JSONs to be the data steward for many data products, from designing DAGs to using data mesh enablers, from performing root cause analyses to analyze the performance of clusters and data streamings.

## Core Skills:

- **Data Cloud** (Terraform, GCP, Bigquery, Cloud Storage, Cloud Pub/Sub, Cloud Composer, Cloud Dataproc, Cloud DataStream, Cloud Run, Cloud Functions, Cloud SQL, Firestore)
- **Data Engineering** (Database Modeling, Pipelines, Data Streaming, ETL, ELT, RegEx, Workflow Design & Orchestration, DAGs, Data Lake, Data Warehouse, Data Mesh)
- **Big Data** (Apache Airflow, Apache Beam, Apache Spark, Apache Hadoop, Apache Hive)
- **Coding / Programming** (Coding Standards, Data Structures, Python, Git, Java)
- **Systems Engineering** (Linux, Unix, Bash, Docker, Kubernetes, Distributed Systems, Clusters)
- **Data Analytics** (SQL, PL/SQL, NoSQL, OLAP, Business Intelligence, KPIs, MoS, OKR)
- **Data Quality** (Data Catalog, Data Governance, Data Security, Personal Data Protection, Data & Metadata Management, Data Ownership, Data Lineage, Architecture Guidelines, Business Processes, Monitoring, Alerts)

## Principles:

- Belief in Freedom. Advance and prosper. Challenge the status quo. Do something big.
- Passion may need structure to survive, but for structure to grow, it needs passion.
- The higher the goals, the greater the performance.
- View this year's failure as next year's opportunity to try it again.
- It's not what computers can do, but why we need them.
- Coordinating is about helping people understand what needs to be done.
- Find ways to remove obstacles to ensure that everyone can work to their greatest potential
- Go down rabbit holes, drill down for the pure joy of geeking out.
- Follow your leaders' curiosity, rather than merely be amused by it.
- Well-defined objectives generate convergence, clarity and satisfaction at work.

## Work Experience



**falabella.com**

**Lead Data Engineer**

Jun 2022 – Dec 2023

Mexico · Remote

After the integration of the teams of Linio and Falabella.com, I became a Team Lead for one of the Data Engineering Cells responsible to create Data Products for internal use among all the business units of Falabella and its providers. A team being integrated by Data Engineers and Data Analysts working along Data Architects, Data Governance team and Stakeholders.

- Delivered and led the Development of scalable and reliable Data Products with focus on business priorities and impact.
- Designed the Data Lake & Data Warehouse within Google Cloud following guidelines from Data Analytics Platform, Architecture, Governance, Security and Compliance.
- Created a framework for encrypting specific nested fields on Nested Data Structures (Such as JSONs) containing sensitive data, by using recursive algorithms that worked fast in clusters (Cloud Dataflow with Apache Beam)
- Created a Data Product that resolves / conciliates deltas from multiple entity's payloads sent as data streams through Pub/Sub.
- Along with my team, we used and improved a Data Migrator framework built on top of Spark in order to move data from Postgresql into Bigquery.
- Ownership of the Big Data components and jobs to consolidate data assets in the Data Warehouse (BigQuery) such as Cloud Composer (Airflow), Cloud Dataflow jobs (Apache Beam), and the related repositories in Gitlab which provisioned infrastructure using Terraform and Gitlab CI pipelines.
- I learned a lot from different people in the company and also had the opportunity to share knowledge with my team and others within the company.

**General Skills:** Data Engineering · Big Data · Cloud Computing · Big Data Analytics · Data Products · Data Warehousing · Datalakes · Google Cloud Platform

**Technical Skills:** Workflow Design · Data Pipelines · Data Streaming · Batch Processing · Extract, Transform, Load (ETL) · Data Modeling · Data Quality · Data Warehouse Architecture · Data Analysis · Google BigQuery · PostgreSQL · SQL · PL/SQL · NoSQL · JSON · Parquet · Python · Apache Beam · Apache Spark · Apache Airflow · dbt · Cloud Dataproc · Cloud Dataflow · Cloud Firestore · Identity & Access Management (IAM) · Git · Gitlab CI · Gitops · Terraform · Scripting · Bash · Regular Expressions · OLAP · Linux · Hadoop · Docker · Kubernetes

**Soft Skills:** Love of Learning · Communication · Teamwork · Team Leadership · Easily Adaptable · Problem Solving · Agile Methodologies · Research Skills



## **Linio Group**

### **Big Data Engineer / Tech Lead**

Aug 2019 – May 2022

Mexico – Remote

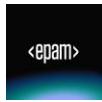
Big Data Engineer for Linio business unit within Falabella Group, responsible for creating the Data Lake layout and Data Warehouse best practices for Linio to integrate and refactor its Data Products into GCP as required by Falabella Group with the benefit of becoming scalable and serve a wide range of purposes.

- Designed the Linio BI's Data Lake & Data Warehouse layout for Data Engineers and Analysts to order and place their Data Products and to deliver them automatically to different analytical departments within the organization.
- Developed a Framework for automatically creating authorized views or for Analysts to create ad-hoc views that were automatically authorized according to guidelines.
- Created Dataflow Jobs for ETL purposes in order to move data in a scalable and faster way.
- Created Airflow workflows to automate the entity generation and update for data products that previously were on Jenkins.
- Consolidated the Transactions and Catalog products on Bigquery that contained data from multiple transactional databases following the analytical data modeling already created in a transactional database for compatibility and legacy purposes.
- I also designed my own toolbox using python client libraries for performing automatic actions (triggers) through the use of GitHub Actions by only pushing changes to config files to GitHub, allowing coordinated teamwork and ease of use.
- This was an important step in my career where I could experiment and learn by doing, while also improving with the feedback from my teammates and stakeholders.

**General Skills:** Data Engineering · Big Data · Cloud Computing · Big Data Analytics · Data Products · Data Warehousing · Datalakes · Google Cloud Platform

**Technical Skills:** Workflow Design · Data Pipelines · Batch Processing · Extract, Transform, Load (ETL) · Data Modeling · Data Quality · Data Warehouse Architecture · Data Analysis · Google BigQuery · MySQL · SQL · PL/SQL · NoSQL · Python · Apache Beam · Apache Airflow · dbt · Cloud Dataflow · Identity & Access Management (IAM) · Git · GitHub Actions · Scripting · Bash · Regular Expressions · OLAP · Linux · Hadoop · Docker · Kubernetes

**Soft Skills:** Love of Learning · Communication · Teamwork · Problem Solving · Agile Methodologies · Research Skills



**EPAM Systems onsite at Google**  
**Systems Engineer / Senior Systems Engineer**  
Dec 2017 – Aug 2019  
Mexico City Area – Onsite

I've worked as a Big Data Support Engineer or Technical Solutions Representative for Google Cloud Platform's catalog of Big Data and Machine Learning products.

Some general tasks were:

- Platform delivery support
- Reproduce issues and find root causes
- Drive issue resolution
- Mitigate availability issues and offer workarounds
- Test & troubleshoot configurations

During my first months, I achieved two Google Cloud Certifications: Cloud Architect and Data Engineer. I also always were among the top performing agents as measured by the defined KPIs.

**General Skills:** Big Data · Cloud Computing · Data Warehousing · Datalakes · Google Cloud Platform · Troubleshooting · Technical Support · Customer Focus

**Technical Skills:** Google BigQuery · Cloud Dataflow · Cloud Dataproc · Data Studio · Cloud Datalab · Cloud Pub/Sub · Cloud Composer · Cloud Dataprep · Vision API · Natural Language API · Speech API · Translation API · Cloud Machine Learning Engine · Cloud Firestore · BigTable · Cloud SQL · Google Kubernetes Engine (GKE) · Cloud Functions · VPC Networks · SQL · PL/SQL · NoSQL · Avro · Parquet · Python · Apache Beam · Apache Airflow · Tensorflow · Identity & Access Management (IAM) · Scripting · Bash · Regular Expressions · Linux · Hadoop

**Soft Skills:** Love of Learning · Communication · Teamwork · Problem Solving · Research Skills



**Softek Mexico onsite at DXC**  
**Systems Engineer**  
Apr 2017 – Dec 2017  
Guadalajara Area – Onsite

Member of the Big Data Services Team at DXC Technology (An Hewlett Packard company). The purpose of the Big Data Services organization is to support all Big Data platforms to enable the COE (Center of Excellence) organization to meet the needs of the enterprise.

Key Responsibilities:

- Platform delivery
- Platform stabilization and optimization
- Governance to ensure compliance of the security model

- Drive issue resolution
- Operational support and administration
- System capacity planning
- Tool development
- Automation
- Provide operational reporting
- Drive Short to Midterm Solutions for Strategic BI Initiatives

**General Skills:** Big Data · Cluster Management & Provisioning · Troubleshooting · Tech Support

**Technical Skills:** Scripting · Bash · Linux · Hadoop · HDFS · MapReduce · YARN · Apache Ambari  
· Apache Hive · Apache HBase · Systems Monitoring · Troubleshooting Guides & Handbooks

**Soft Skills:** Love of Learning · Problem Solving

## Formal Education



[Universidad de Guadalajara](http://www.udg.mx)

[Bachelor of Technology - BTech, Information Technology](#)

Grade: 96/100

Activities and societies: I was part of the Stimulus Program for Outstanding Students.

This bachelor's degree has a IT Business Management focus, especially in regards to:

- IT Project Management
- Requirements Analysis
- Software Engineering
- Systems Design using UML
- Implementation of Business Information Systems (ERP, CRM, etc)
- IT Service and Process Management following the best practices of ITIL
- Security and Systems Audit according to ITIL, COBIT & ISO27000
- Database Management Systems
- Application Development
- Computer Networking
- Big Data & Business Intelligence

## Online Education Credentials

- **Architecting with Google Cloud Platform Specialization**  
[Coursera Issued Feb 2018 Credential ID W8NR7LN2L8MU](#)  
**Related Skills:** Google Cloud Platform · Cloud Computing · Data Warehouse Architecture · Kubernetes
- **Google Cloud Certified Professional – Cloud Architect**  
[Google Cloud Certified Issued Feb 2018 Credential ID Iq7pSU](#)  
**Related Skills:** Google Cloud Platform · Data Warehousing · Cloud Computing · Data Warehouse Architecture · Kubernetes
- **Reliable Cloud Infrastructure: Design and Process**  
[Coursera Issued Feb 2018 Credential ID WS8S5KHKW9NG](#)  
**Related Skills:** Google Cloud Platform · Cloud Computing
- **Building Resilient Streaming Systems on Google Cloud Platform**  
[Coursera Issued Jan 2018 Credential ID 8XM2UV9SRTS5](#)  
**Related Skills:** Google Cloud Platform · Apache Kafka · Data Pipelines · Data Streaming · Cloud Computing · Apache Beam · Extract, Transform, Load (ETL)
- **Data Engineering on Google Cloud Platform Specialization**  
[Coursera Issued Jan 2018 Credential ID M54P6CGWDH45](#)  
**Related Skills:** Google Cloud Platform · Data Pipelines · Big Data · Cloud Computing · Extract, Transform, Load (ETL)
- **Leveraging Unstructured Data with Cloud Dataproc on Google Cloud Platform**  
[Coursera Issued Jan 2018 Credential ID N3DSPN69PJFX](#)  
**Related Skills:** Google Cloud Platform · Data Pipelines · Big Data · Cloud Computing · Big Data Analytics · Extract, Transform, Load (ETL) · Hadoop
- **Serverless Data Analysis with Google BigQuery and Cloud Dataflow**  
[Coursera Issued Jan 2018 Credential ID 5MGSPF2P7KYZ](#)  
**Related Skills:** Google Cloud Platform · Apache Hive · Data Pipelines · Data Warehousing · Big Data · HBase · Cloud Computing · Big Data Analytics · Apache Beam · Data Warehouse Architecture · Extract, Transform, Load (ETL) · Google BigQuery · Data Analysis · SQL
- **Serverless Machine Learning with Tensorflow on Google Cloud Platform**  
[Coursera Issued Jan 2018 Credential ID 5VK3LDC54DQJ](#)  
**Related Skills:** Google Cloud Platform · Machine Learning · TensorFlow · Cloud Computing · Big Data Analytics
- **Architecting with Google Cloud Platform**  
 Coursera Issued Dec 2017  
**Related Skills:** Google Cloud Platform · Data Pipelines · Data Warehousing · Cloud Computing ·

## Data Warehouse Architecture

- **Elastic Cloud Infrastructure: Containers and Services**  
[Coursera Issued Dec 2017 Credential ID EXJQC2KRJGBC](#)  
**Related Skills:** Google Cloud Platform · Big Data · Cloud Computing · Kubernetes · Docker
- **Elastic Cloud Infrastructure: Scaling and Automation**  
[Coursera Issued Dec 2017 Credential ID 2DW97G38F58X](#)  
**Related Skills:** Google Cloud Platform · Data Pipelines · Big Data · Cloud Computing
- **Essential Cloud Infrastructure: Core Services**  
[Coursera Issued Dec 2017 Credential ID 69HDA3HDFLNS](#)  
**Related Skills:** Google Cloud Platform · Cloud Computing
- **Essential Cloud Infrastructure: Foundation**  
[Coursera Issued Dec 2017 Credential ID TTSEHG4P4KHV](#)  
**Related Skills:** Google Cloud Platform · Cloud Computing
- **Google Cloud Platform Big Data and Machine Learning Fundamentals**  
[Coursera Issued Dec 2017 Credential ID VERVCQ9VW9WX](#)  
**Related Skills:** BigTable · Google Cloud Platform · Data Warehousing · Machine Learning · Big Data · TensorFlow · Cloud Computing · Big Data Analytics · Data Warehouse Architecture
- **Google Cloud Platform Fundamentals: Core Infrastructure**  
[Coursera Issued Dec 2017 Credential ID 7ZZP5DXPRMMH](#)  
**Related Skills:** Google Cloud Platform · Cloud Computing
- **Introduction to Big Data with Apache Spark**  
[edX Issued Jul 2015](#)  
**Related Skills:** Data Pipelines · Apache Spark · Big Data · Big Data Analytics · Data Analysis
- **Data Lakes for Big Data**  
[Dell EMC Issued Jun 2015](#)  
**Related Skills:** Modelado de bases de datos · Data Pipelines · Big Data · HBase · Data Warehouse Architecture · Extract, Transform, Load (ETL)
- **Machine Learning**  
[Coursera Issued May 2015](#)  
**Related Skills:** Machine Learning · TensorFlow · Big Data Analytics
- **Mining Massive Datasets**  
[Coursera Issued Mar 2015](#)  
**Related Skills:** Machine Learning · Big Data Analytics