

Diego Nieto

Profile

- **Birth date:** 04/04/1976
- **Email:** dnieto@gmail.com
- **Web (in spanish):** <https://www.dnieto-it.es>
- **GitHub:** <https://github.com/lesandie>
- **LinkedIn:** <https://linkedin.com/dnieto>
- **Path:** /Spain/Pontevedra

I was born in the 8-bit era and grew up with a 128KB spectrum and then an AMIGA 500. I learned to program BASIC at the age of 12 and suddenly I was playing arcade games, SNES and Megadrive until I went to college to study computer science in the 90s: learnt Assembler, C, C ++, JAVA or SQL and then unknown Linux. The first distro I installed was a Slackware. I have worked in companies such as Sun Microsystems, Informix software or Ascential until I went to the public sector in 2002 to work for [CESGA \(Galicia Supercomputing Center\)](#). I currently work for [Tinybird](#).

I like to read, walk and enjoy a good day with my wife and kids. Continuous learning is in my DNA, and as a geek I value technology as a tool to improve our lives and change the world.

During my long experience I touched a lot of technologies, languages and paradigms, some still around, but my interests and skills are bet on DBMSs, SQL, PL/pgSQL, Python and sysadmin stuff, in which I have a strong technical background and experience.

If I have to choose a set of labels for my work they would be: data/database engineer with a pinch of sales support and data analyst: Understand business goals and technical requirements and turn complex problems into understandable, achievable solutions.

Education

1994-1999: BSc, Computer Science Engineering, Deusto University

Final Degree Project: Developed a front-end interface written in tcl/tk for the debian package manager (dpkg). It used the alien tool (<https://joeyh.name/code/alien/>) to convert packages between formats (.tgz, .rpm, .deb).

2004-2007 MSc in Data engineering Computer Science Joint PhD Programme [Santiago de Compostela](#) & [A Coruña](#) Universities

[Project MORFEO \(Spanish\)](#): Designed and implemented a datawarehouse for storing and analyzing data from patients with Cardio-Respiratory Sleep Disorders. In terms of data management and ETL, I had to deal with different types of data, from basic tests like a Glasgow Scale to binary files from a polysomnography and came up with a design based on Kimball's approach to analyze and study all the data in an OLAP way using Oracle 10g, with the possibility of using datamining techniques. I published an article at a national level congress (CASEIB2006 <http://www.unavarra.es/caseib2006/Download/abstracts.pdf>) that was a summary of the main work.

Experience

Ago2021-Now | Data Engineer, [Tinybird](#)

I work in a real time analytics environment, solving issues and adding new features to ETL pipelines and API endpoints for clients. Typically the workflow is: batching out data from Snowflake or PostgreSQL or using topics from kafka, into Tinybird (ClickHouse) and preparing different endpoints to query the data in real time. The batches were managed with GitHub actions (some every minute, others every day depending on the operational needs of the client) using the tinybird's CLI tool that has different connectors for different databases (BigQuery, Snowflake, PostgreSQL, etc). Also other clients used the streaming API of Tinybird and I've written a python benchmark to check the ingestion limits simulating a specific load environment.

I contribute to the product by doing some PRs to fix some issues detected working with the clients and helped with the content of some blog posts about ClickHouse.

The main tooling I use is SQL, Python, Tinybird API & CLI, ClickHouse, Google Cloud and Ansible for deployments.

Jan2002-Feb.2021 | Engineer/Analyst, Galicia Supercomputing Center (CESGA)

Since I moved to CESGA I've been using the latest tech and applying it to many projects, for example helping with the benchmarking and deployment of the first version of the CESGA's Hadoop cluster for scientists in 2012 that evolved to <https://bigdata.cesga.es>

In terms of computing, networking and storage infrastructure, CESGA runs its own datacenter so we built what we needed using our computing and networking resources (OpenStack & CloudStack & OpenNebula).

I've been working 15+ years at CESGA with many talented colleagues from different areas (networking, computing, apps ...) developing many activities and projects while managing the technical team of the e-learning area. Some remarkable ones:

- Design and implementation of backend infrastructures with different OSS technologies and tools for CESGA's main services:
 - *GaleraDB and PostgreSQL/PostGIS/TimescaleDB cluster infrastructure sysadmin/DBA*: due to my strong background in RDBMSs (nearly 10TB storage). Many of CESGA's services (like the ones described next), projects and critical infrastructure (Accounting for HPC resources) share both of these backends:
 - I wrote shell or python scripts to automate some monitorization metrics (Zabbix and Nagios), database backups or sysadmin stuff and sometimes I have to update or add new stuff to them.
 - Also I helped users to model E/R schemas to incorporate their datasets.
 - Tune/rewrite their queries.
 - Solved performance issues or wrote ETL pipelines in PL/pgSQL or Python (Airflow)

You can check my github repo with some configurations I documented, a failover script in Python and some backup shell scripts. I had to anonymize the code and pushed it to github from a private gitlab repo:

- [pgpool-clusterconfig](#)
- [pgpool-failover](#)
- *Videoconference and Streaming scalable service infrastructure admin (Jitsi & BigBlueButton)* <https://seminar.cesga.es> & <https://meet.srv.cesga.es> (150+ users)

- *Chamilo scalable VLE service sysadmin/devops* (<https://aula.cesga.es>) (10K users). In the e-Learnig area we also run a helpdesk to receive emails with problems from users and to solve them, just pointing them out how to interact with the webapp to do some tasks, or in case of a bug, solve it and push the fix to our private gitlab repo. When we have time, we try to PR our corrections and new functionalities into the [Chamilo](#) public repo.
- *Nextcloud scalable service infrastructure sysadmin* (<https://cumulo.cesga.es>) (200+ users).
- Helping different Universities research groups and spin-offs to setup and use the different cloud/computing infrastructures at CESGA.
 - [GeoForsk](#): configuring and managing their ShinyProxy Docker infrastructure with some dataops stuff.
 - [Project Djehuty](#): The excavation team uses a hardened Raspberrypi4 in AP mode with a PostgreSQL11 instance. They use a Qgis template specially designed to map all the tomb layout and objects they find during the excavation process. When they bring back the "raspi4" I import all the data into the main cluster at CESGA's premises.

Mar.2001-Dec.2001 | Sales support eng./consultant, Ascential Software, Madrid, Spain

Ascential was acquired by Informix prior IBMs acquisition of the later. My work mainly was the design, implementation and exploitation of data warehouses using ETL and BI tools like DataStage, Cognos or Crystal Reports. Worked for customers like Repsol or Xunta de Galicia.

Mar.2000-Mar2001 | Sales support eng./consultant, Informix Software, Madrid, Spain

HQ was in Madrid, but I was based in Santiago de Compostela working mainly with local clients but sometimes I had to travel to other spanish areas. Worked with Informix database technology and Informix content management solutions (iReach, Media360), for customers like Xunta de Galicia, Altia, Oviedo University, Tele5 and also doing some training sessions at on-site clients. Later that year IBM acquired Informix database business and some former Informix employees stayed at the renamed company: Ascential.

Feb.1999-Feb.2000 | Engineer, Sun Microsystems, Bilbao, Spain

Finishing my studies at the university I was offered a learning position in a Sun sales office in Bilbao working mainly with local clients. Worked with Ultra Enterprise hardware, Solaris, Netscape suite for customers like Euskaltel. I also developed a Java Applet for the Caja Rural Vasca to simulate mortgages.

Skills

- Programming:
 - Python
 - Shell Scripting
 - C++ (refreshing)
 - Rust (learning)
- Data/ETL:
 - PL/pgSQL
 - SQL
 - Apache airflow (basic)
 - DBA Level in PostgreSQL & MariaDB
- Sys/infra:

- System Administration level in Linux
- Docker
- Kubernetes (basic)
- Ansible
- Languages:
 - Galician - Native
 - Spanish - Native
 - English - Fluent
 - Portuguese - can read and understand it

Side jobs & Hobbies

- I'm a selfhoster with many IoT, sensors, 3D-Printer, raspberries and HomeAssistant stuff at home. I run my own network (OpenWrt), storage & monitoring (collectd/prometheus/influxdb/timescaledb/grafana) infrastructure and sometimes I consult to customers and local SMEs. You can check my website <https://www.dnieto-it.es> (Spanish)
- I sometimes get hired for training sessions of PostgreSQL best practices (spanish). [Guia PostgreSQL](#)
- I've developed a PostGIS backend for a Qgis project template, written in PL/pgSQL and using many capabilities of a PostgreSQL engine like materialized views, triggers and functions. You can check the code and rationale at [postgis-backend](#) in my github public repo. Again I had to anonymize the code and push it to a public repo.
- Some problem solving skills:
 - Using the analytics platform [Tinybird](#) at my repo [tinybird_nyc_taxi](#)
 - Python scripting for outliers detection [python-outliers](#)
 - Simple flask web app example and deployment to Google Cloud [flask-app](#)
- (2018-2020) I was temporal university lecturer at the USCs [International PhD School](#), teaching text image processing with tesseract.