Christian Herrera Jiménez

Data Platform Architect · Databricks Lakehouse, Data Mesh and Data Contracts

Barcelona, Spain cherrejim@gmail.com linkedin.com/in/cherreraj cuadernodatos.es

Profile

Data platform architect with 19 years in software and the last 9 building cloud data platforms. I spent seven years on the team that built Adevinta Spain's multi-tenant data mesh platform, which processes more than 38 million events per hour, where I designed the source data contracts and moved the existing tables to data products without disrupting consumers. Today I shape the Azure Databricks lakehouse behind a loyalty product for a European pharma group: contract-driven ingestion, Lakeflow declarative pipelines, observability and data quality.

I am at my best where architecture meets people: turning business needs into contracts, standards and reusable platform patterns, and helping product squads adopt them. I also teach data engineering in master's programmes and in public.

Core strengths

- **Platform architecture:** multi-tenant lakehouse and data mesh platforms, data products, source-aligned data contracts, medallion design, self-service ingestion.
- **Governance and compliance:** GDPR by design, automated detection of sensitive fields, schema evolution, data quality expectations, observability.
- **Technical direction:** engineering standards (Terraform, Databricks), contract definition with product squads, client-facing delivery, hiring interviews.
- **Teaching and communication:** course author, university course materials, public talks and livestreams.

Experience

Data Platform Architect · TheHubs

`November 2025 – present · Barcelona (hybrid) · official title: Senior Data Engineer`

Data platform for a loyalty application built for a European pharma group. Azure, Databricks, Python.

- Own platform design decisions with wide autonomy, in a cross-functional data team of seven (data engineering, ML, analytics and BI).
- Automated bronze ingestion from source-aligned data contracts, with schema evolution, so that a new source is onboarded by declaring a contract instead of writing a pipeline.
- Run weekly sessions with every product squad to define their input contracts, and work directly with the client on scope and priorities.
- Build the silver and gold layers with Lakeflow Spark Declarative Pipelines (Spark 4), governed in Unity Catalog and deployed with Databricks Asset Bundles.
- Created the platform's observability layer and am now adding data quality rules with pipeline expectations.
- Built the first version of the data portal, the entry point for discovering and consuming data products.
- Introduced generative AI to review data contracts before they reach the platform.
- Currently designing how the platform handles personal and sensitive data under GDPR.
- Helped define the team's standards for working with Terraform and Databricks, and take part in hiring interviews.
- In the first months also owned the analytical data model and the Power BI dashboards, including real-time views.

Data Engineer, Data Platform · Adevinta Spain

`July 2018 – November 2025 · Barcelona (remote)`

Online marketplaces group. Central data platform team serving every marketplace and data domain in Spain; more than 38 million events per hour.

- Core engineer in the team that designed and built the company's data platform and its move to a data mesh, where each domain owns and publishes its data as products.
- Designed and implemented the source data contracts: the agreement between producers and the platform that drives ingestion, schema validation and access to data.
- Led the migration of the existing tables to data products, transparent to the teams consuming them.
- Built privacy into the architecture: contracts carry the GDPR classification of each field, and I implemented an AI agent that detects sensitive fields automatically.
- Contributed to a cost-efficient storage and processing architecture on Delta Lake, with orchestration in Airflow.
- Represented the platform publicly: co-author of "Architecting Compliance: Cost-Effective Data Strategies for GDPR" on the Adevinta tech blog, and speaker on data contracts.
- Took part in hiring interviews for the data platform team.

Scala, Spark, Kafka, Databricks, Delta Lake, AWS, Airflow, Hive.

Big Data Developer · Stratio

April 2017 – June 2018 · Madrid

- Real-time audience prediction for a Spanish media group (Kafka, Spark Streaming, Cassandra).
- News recommender for El País, built as microservices on AWS and Google Cloud.
- Real-time advanced analytics on football matches (Spark, Kafka, Scala).

Big Data Developer · IBM

June 2016 – April 2017 · Barcelona

- Back-end development for m-RESIST, a European digital health project supporting patients with treatment-resistant schizophrenia.

Functional Analyst and Team Coordinator · Viewnext (IBM group)

January 2015 – April 2017 · Barcelona

- Coordinated a team of 12 delivering the Generalitat de Catalunya's online procedures portal: functional analysis, technical feasibility, delivery oversight and documentation.

Functional Analyst and Team Coordinator · Necsia IT Consulting

November 2011 – January 2015 · Barcelona

- Managed a team of 8 on AquaCIS, the billing and customer system used by water utilities: resource planning, client relationship and release management.

Earlier roles

2006 – 2011 · Barcelona

Analyst and Java developer at Ibermática (2008–2011), senior developer at Abast Systems (2007–2008) and Java developer at Sogeti (2006–2007), for clients in banking, insurance, education and public administration.

Teaching, writing and speaking

- **University teaching:** course materials for master's and postgraduate programmes at Universidad Europea, Universitat d'Andorra and Universitat Oberta de Catalunya.
- **CodelyTV:** author of an Apache Spark course, and guest on the livestream "How Adevinta handles 38 million events per hour" about domain events and data contracts.
- **Cuaderno de datos** (cuadernodatos.es): my public notebook, with a 17-lesson course on Lakeflow Declarative Pipelines.
- **Open source:** Spark Ecosystem Cluster, a docker-compose environment to practise with Spark, Kafka, Airflow and Hive locally.

Technology

- **Lakehouse:** Databricks, Delta Lake, Unity Catalog, Lakeflow Spark Declarative Pipelines, Asset Bundles, Spark (PySpark and Scala), Structured Streaming.
- **Streaming and orchestration:** Kafka, Airflow.
- **Cloud and infrastructure:** AWS, Azure, Terraform, Docker.
- **Programming:** Python, Scala, SQL, Java.
- **Analytics:** data modelling, Power BI.

Languages

Spanish and Catalan (native) · English (intermediate).

Education

- **Master in Big Data**, Universitat Ramon Llull, 2015–2016. Grade 9/10. Thesis awarded with honours: "Semi-supervised Sentiment Classification on Large Databases".
- **BSc in Computer Engineering**, Universitat Oberta de Catalunya. In progress.
- **Data Engineering Nanodegree**, Udacity, 2019.
- **Higher Technician in Software Development**, Centre d'Estudis Sant Francesc, 2004–2006.

Certifications

Databricks Lakehouse Fundamentals (2023) · Functional Programming in Scala, Habla Computing (2019) · Oracle Certified Associate, Java SE 8 (2018) · Oracle Certified Professional, Java SE 6 (2010).