

Simone Benitozzi

Milan, Italy

✉ inbox@simonebenitozzi.com

🌐 simonebenitozzi.com

🌐 simonebenitozzi

EXPERIENCE

Tecnocasa

Data Engineer

Sep 2024 - Present, Milan, IT

- Joined as an **external consultant** and was subsequently offered a **permanent position** in recognition of results delivered and continuous improvement initiatives.
- Design and implement **complex ETL and real-time/batch data pipelines** across heterogeneous environments (Oracle DB, Salesforce, AWS S3, REST APIs).
- Architect and maintain **scalable data integrations** between relational databases, CRM platforms, and cloud storage, supporting both operational and analytical workloads.
- Propose and drive **performance refactoring initiatives**, consistently delivering measurable throughput gains on legacy systems.

Red Reply

AI & Data Consultant

Nov 2023 - Jun 2025, Milan, IT

- Developed **ETL pipelines and cloud architectures** using Oracle, Python, Docker, and Kubernetes (OKE); contributed to AI-driven automation projects.
- Recognised as **top-3 performer** in first-year appraisal (8.33/10), admitted to the firm's **Excellence Path** for top academic achievers.
- Active member of the **General AI Community of Practice**, collaborating with international Reply teams across Europe on GenAI and agent-based solutions.

PROJECTS

Listings Publishing - ETL Refactoring

Python, ODI

- Legacy ODI-based pipeline caused **8-hour peak backlogs**; re-architected to a **Python-orchestrated concurrent solution** leveraging in-memory processing.
- Designed full solution, built an independent **MVP that convinced the CTO** to adopt the new approach; then led a **team of 3** to production delivery.
- **97% processing time** - average per-listing latency from **~50 s** **1.5 s**; queue from **8 h** **10 min**.

Mail Spooler - High-Performance Rewrite

Python, ODI, Oracle SQL

- Self-proposed migration of a PL/SQL email system to a **Python high-concurrency solution**; preserved legacy spooling table for zero-disruption rollout.
- **5,000x throughput gain** - from **~12 emails/min** **1,000+ emails/sec**; eliminated delivery backlogs that previously lasted **days**.

Cadastral Data Import

AWS S3, ODI, Python, Salesforce API

- End-to-end design of a pipeline ingesting **Italian cadastral data** (Agenzia delle Entrate) from AWS S3, applying complex transformations, and syncing to Salesforce CRM.
- Automated property and client registration: **manual CRM input fully automated**, generating direct business value for agencies.

PROFILE

Data Engineer specialising in performance-critical ETL pipelines and scalable data architectures across heterogeneous environments.

Known for identifying systemic inefficiencies, proposing targeted refactors, and delivering measurable, production-grade results often self-initiated. Fluent across the full stack: from Oracle and Salesforce integrations to cloud-native Python orchestration.

Mensa member. Strong analytical and problem-solving mindset.

SKILLS

Languages Python, SQL, Bash

Frameworks FastAPI, Flask

ETL / Int. ODI, Informatica

Databases Oracle, MySQL

Cloud OCI, AWS S3

CRM / API Salesforce, REST

DevOps Docker, Kubernetes, Git

AI / ML PyTorch, XGBoost

OS Linux (Oracle/RHEL)

EDUCATION

MSc - Computer Science (AI & Computer Vision)

University of Milano-Bicocca

110/110 with Honours - GPA 4.0/4.0

Thesis: Vision Transformers

BSc - Computer Science

University of Salerno

110/110 with Honours - GPA 4.0/4.0

CERTIFICATES

- **OCI Generative AI** | Oracle
- **Oracle AI Foundations** | Oracle
- **OCI Data Science** | Oracle
- **Oracle Analytics Cloud** | Oracle
- **Intro to Machine Learning** | Duke

LANGUAGES

- Italian - Native
- English - Fluent
- Polish - Conversational