

Daniel McNulty

HIGHLIGHTS:

Senior Software Engineer with +11 years of experience in distributed systems and large databases. A data-driven professional with high quality code base. Python developer, master of SQL queries. Enjoys a challenging work environment.

EDUCATION:

- Machine Learning at Online Stanford University, 2016 (1 year).
- Bachelor Computer Science at University of Leon, 2014 (4 years).
- Bachelor Software Engineering at University of La Laguna, 2009, (3 years).

SUMMARY:

- Data-driven Python developer with strong experience in large database systems.
- Co-founder of a startup.
- Worked in several SQL and NoSQL database technologies like Hadoop, Hive, MySQL, SQL Server, PostgreSQL and MongoDB at all project stages.
- Experience building RESTful APIs.
- Experience in Python ecosystem: Flask, SQLAlchemy, peewee, PyMySQL, pymongo.
- Worked in distributed systems and BigData: MapReduce, Mesos.
- Large experience in developing robust and smooth data pipelines and ETL applications.
- Experience with Test Driven Development (TDD), Waterfall and Agile.
- Experience in python testing modules: py.test, unittest, mock, monkey-patching, coverage.
- Extensive knowledge of Software Development Life Cycle (SDLC).
- Taken ownership of data Life Cycle.
- Proficient in performance tuning and optimization of queries.
- Problem solver and analytical mindset.
- Solid understanding of data structures, algorithms, object-oriented design (OOD) and patterns.
- Knowledge in Machine Learning and Artificial Intelligence techniques.
- Pythonista
- Linux Arch lover.

TECHNICAL SKILLS:

- Languages: Python, , PL/SQL, SQL, C++..
- Databases: Hive, MySQL, SQLServer, PostgreSQL, MongoDB.
- Big Data: Apache Hadoop, Hadoop Streaming API, MapReduce, Mesos.
- IDE: PyCharm, vim.
- Tools: GIT, Github, Jira.
- Management methodologies: Waterfall, Agile, OOA, OAD.
- Data modeling tools: Data Modeler, Reno, Dia, ERwin.
- Modeling language: UML.
- Job scheduler: Celery, ControlM.

PROFESSIONAL EXPERIENCE:

Employer: Corus Systems

Position: Data Engineer

Projects:

RIA (Advanced Analytics, Capital Group)

- Built ETL pipelines in Python and SQL Server Parallel Data Warehouse (PDW).
- Development of Fuzzy matching algorithms for data identification.
- Built APIs with Flask.
- Web scraping with BeautifulSoup and scrapy.

PROFESSIONAL EXPERIENCE:

Employer: Corus Systems

Position: Lead Python developer

Projects:

Real-time Customer Transactions Report (Santander Bank)

- Built ETL pipelines in Python and SQLAlchemy to feed a real-time money laundering detection system.
- Built APIs with Flask.
- Created development plan using Agile and TDD methodologies.
- Wrote automated tests for data quality and performance.

- Optimized an ETL batch reducing the duration from 20 hours down to 7 after analyzing dependencies between processes and creating more efficient concurrent and independent dataflow chains.

Employer: Efron group Inc.

Position: Lead ETL developer

Projects: Anti-money laundering data warehouse (Santander Bank)

Boston, MA (Jan 2015 – Jan 2018)

- I took ownership of the data flow and data pipelines across several projects to deliver the highest data quality needed.
- Consolidated all enterprise's financial data sources (including interacting with APIs), created and maintained data models and schemas in several data warehouses.
- Built robust data cleansing batch processes using Python.
- Built testing applications, unit test, sit testing...
- Developed high complex SQL code.
- Improved existing batch processes and SQL queries, applied multi-parallel processes. Improved overall batch time by 50%.
- Worked with multi-terabytes data warehouses.

Employer: Research Institute of Technological Studies, European Commission

Position: Data Engineer

Main Projects:

- **Nielsen Clickstream** **Brussels, BE (Apr 2013 – Jan 2015)**
- **Facebook/iTunes/Torrent music likeness** **Brussels, BE (Apr 2013 – Apr 2014)**
- I was the only Data Engineer in the Institute, giving support to +300 data scientists.
- Built Big Data pipelines using Python, Hadoop, Hive, PostgreSQL, and MySQL.
- Designed and built Big Data solutions for several projects.
- Helping data scientists on using Python numeric and scientific libraries: pandas, numpy, and NLTK.
- Implementation of several multi-petabyte Hadoop clusters for research analysis.
- Performed an NLP analysis over 4 Billion websites.
- Improved existing algorithm for matching music artists among several data sources: Facebook, iTunes, Torrent, MusicBrainz...
- Improved adBlock 'Easylist' regular expression performance for specific needs on the Nielsen clickstream project.

Employer: Tax Management, Santa Cruz City Council

Position: Senior Software Engineer

Projects: CRM

Santa Cruz, Spain (Feb 2009 – Apr 2013)

The company manages and collects taxes from more than 330k citizens and companies for the City Council of Santa Cruz de Tenerife, Spain.

- Developed from the ground up several high-performance CRM applications in C++.
- Implemented unit testing and data quality procedures.
- Developed database code for processing billions of rows.
- Developed SQL queries for reporting.
- Implemented database and virtualization systems from scratch.
- Built desktop applications capable of retrieving millions of records in seconds.

Employer: Computing Research department of the University of La Laguna.

Position: Backend Engineer

Projects: DESIC

Santa Cruz, Spain (Nov 2006 – Feb 2009)

I was the DevOp of the Computing Research department of the University and on the Physics faculty.

- Multi-parallel processing Research.
- Built MySQL database solutions to researchers for their heavily computational process.
- Implemented Nvidia Tesla Cluster for GPU processing.
- Developed Linux scripts for scheduling data pipelines and systems administration tasks.

Employer: Innova7.org

Position: Co-Founder

Projects: TLP Tenerife 2006, 2007 & 2008

Santa Cruz, Spain (Jun 2006 –Aug 2008)

I was one of the founders of the Innova7. A startup that promotes and builds e-Gaming events in Spain, like "TLP Tenerife", a LAN Party event visited by +50K gamers.

- Linux scripting with bash.
- Designing the event's network for around 1300 computers and devices.