+34 626 364 652
dsaiztc@gmail.com
Spain

DANIEL SAIZ
English | Spanish

github.com/dsaiztc
linkedin.com/in/dsaiztc
dsaiztc.com

Work Experience

Data Manager | Spotahome | Madrid, Spain   04.2019 – Now

- Define company-wide Data Governance proposal with the goal of laying up the guidelines for any future developments what shall guarantee a minimum data quality

- Explore different Data Architectures with a combination of S3/Redshift, with different data sourcing methods for doing CDC (*Change Data Capture*): AWS DMS, Kafka Connect.

- Hiring manager for (also) other data profiles: *Data Engineer*, *Data Analyst*, *Data Scientist*. Conduct technical interviews and develop test-cases.

Business Intelligence Manager | Spotahome | Madrid, Spain   05.2018 – 04.2019 (1 yr)

- Build and lead a team of BI Developers with different backgrounds

  ▪ Assess & define the recruitment process, including technical interviews and tests
  ▪ Organize and prepare training sessions for Python-oriented ETL development

- Define, prioritize & coordinate roadmaps and day-to-day task planning, including technology assessment

- Define company-wide data strategy in terms of event standardization & domain language

- Design and build DWH based on a myriad of different data sources (Domain Events, relational tables, MongoDB, APIs, etc.) on Amazon Redshift

Data Analyst / Engineer | Spotahome | Madrid, Spain   09.2017 – 04.2018 (8 mos)

- Exploratory data analysis with Python

- ETL development with Python and Pentaho / Kettle [1]

- Assessment and posterior deployment of Apache Airflow [2] for batch processing (ETLs, scheduled reports) on Kubernetes [3] (CI with brigade [4]).

  ▪ Own initiative that I previously tested, proposed and led in order to take advantage of the upcoming (at the time) Kubernetes integration in v1.10
  ▪ This allowed us to better allocate infrastructure resources according to real consume, launching each individual ETL on its own POD and thus being able to dimension based on individual necessities
  ▪ Development of custom KuebernetesOperator for setup different environments automagically

- Development of internal libraries for posterior ETL building with Python: data connections to different sources, API management, logging, etc.

- Assessment and posterior deployment of a Business Intelligence tool, including POC deployment of Tableau and Power BI

- Report building with Power BI for company-wide consume, which included managing stakeholders requirements, help in KPI definition and ensure the quality of the data

Skillset: Python, SQL, git, Docker, Kubernetes, Power BI, Apache Airflow, Apache Spark, Redshift, PostgreSQL, MongoDB, Google Analytics

[1] https://github.com/pentaho/pentaho-kettle
[2] https://airflow.apache.org
[3] https://kubernetes.io
[4] https://brigade.sh

Software Engineer & Data Analyst | CST International | Dublin, Ireland   08.2015 – 07.2017 (2 yrs)

- Design and development of an *event-driven* data processing system in AWS with **python**: collects different sources of data (**S3**, **API Gateway**), process and performs different actions based on the content. All is based on events and queues (**DynamoDB**, **DynamoDB Streams**) that are automatically processed by different **AWS Lambda** functions. This was the result of a migration/unification of many different client-based systems (different scripts, jobs and DBs).

- Full Stack development for enhance/renew a reporting system running on **Web2py**. Using **Bootstrap** 3 for building the interface, **jQuery** (with a few plugins) for DOM manipulation and AJAX requests as well as visualization libraries

like **Google Visualization**, **D3.js** or **C3.js**. Use of **Reportlab** for building PDFs. Running on top of DWH (Aurora/MySQL).

- Setup of **AWS Elasticsearch Service** integrated (proxied) within the Web2py reporting system using Searchkit UI Components. Before going for this UI kit I built a prototype from scratch with **Vue.js**.

- Data cleaning and analysis (EDA, Hypothesis testing, visualizations and some basic ML algorithm). Data retrieving & wrangling mainly with **python** (requests, BeautifulSoup, pandas). Data analysis with **R** (tidyr, dplyr, ggplot, lm). Some experience with NLTK and sklearn for comment classification.

Skillset: python, py.test, SQL, git, Web2py, Statistics, Selenium, HTML, CSS, JavaScript, jQuery, Bootstrap, R, Unix, AWS (S3, Lambda, DynamoDB, API Gateway, VPC, RDS, CloudWatch, CloudFront), Microsoft Power BI

Software Engineer (Intern) | UC3M | Madrid, Spain   **12.2013 – 06.2014 (7 mos)**

- I led the design and development of a native **Android** app in **Java** for *Bultaco Motors* (electric motorcycle manufacturer) through an R&D project with *Carlos III University of Madrid*.

- Main development: GPS navigation app (based on *OsmAnd*) with real-time collection & display of telemetry data (Bluetooth). Full development of a Machine Learning system (ANN/MLP) in Java aiming to predict the viability of unknown routes based on historic data. Information retrieval from SOAP (IIS Server).

Skillset: Java, git, SQL, Machine Learning

Education

Telecommunications Engineering (BSc + MSc) | UC3M | Madrid, Spain   **09.2008 – 09.2014**

Top 15%. Main Coursework: Computer Science, Digital Signal Processing, Electronics, Communication Systems, Networking, Statistics. Final Year Project Distinction with Honors: *Range Estimation System for Electric Vehicles with Artificial Neural Networks*, available here (in Spanish): http://dsaiztc.com/PFC.pdf

Other Technical Experience

Data Analyst Nanodegree | Udacity | Online   **03.2015 – 08.2015**

Built by industry and project-based curriculum for becoming a Data Analysis / Scientist.

- Exploratory Data Analysis with R. Explore new datasets to uncover meaningful patterns.

- Data Wrangling OpenStreetMaps Data. Use data munging techniques to clean OpenStreetMap data, converting it from XML to JSON format, importing the clean file into a MongoDB database and run some queries (python).

- Analyzing the NYC Subway Dataset. Looking for data trends with different analysis: statistical (regression analysis, hypothesis testing), raw data and visualizations. CSV data analyzed with different python data-analysis tools (pandas, stats, matplotlib, ggplot).

Secret Santa (Android APP)   **11.2014 – 12.2014**

Native Android application for making *Secret Santa* draws (http://dsaiztc.com/secretsanta). >15000 downloads.