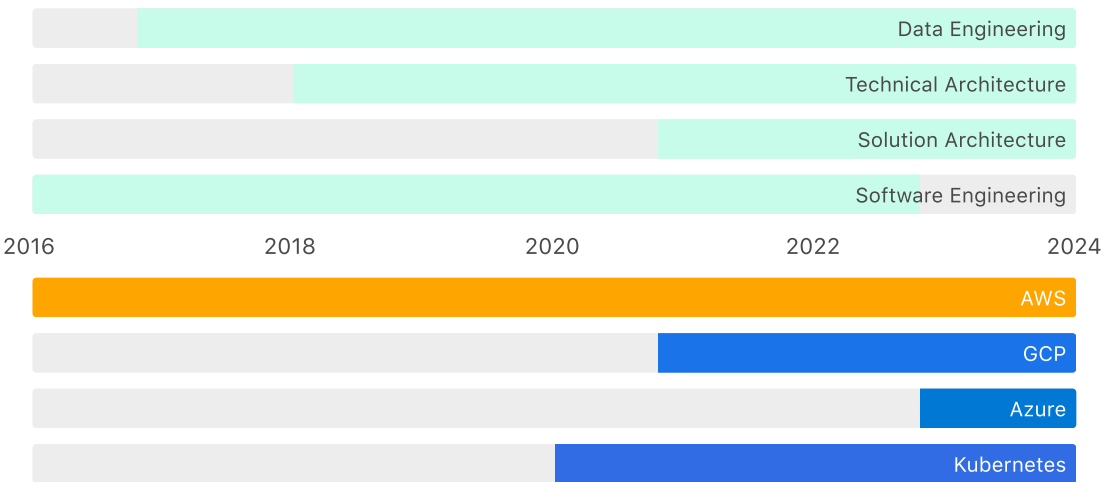




Karlos Álvarez
Software | Data Engineer | ML |
Solutions Architect
karlos.alvan@proton.me

Profile | Experience | Education | Languages

Software Engineer with experience in a wide range of IT businesses along many countries. Specialized in Data Engineering, Information and System Architecture, my key value is to provide meaningful clues for the wise design and elastic architectural process of high performance/critical software solutions.



Profile | Experience | Education | Languages

Data Engineer: Archipelo (b2b) (Palo Alto, United States)
October, 2023 - December, 2024

Archipelo collects SDLC data. My work as Data Engineer was to provide an alternative to a complex Streaming Graph where data products are constructed in real time to provide actionable data to several application users. The Streaming Graph is based on Dataflow, running Apache Beam as a Streaming system. We process different sources of information and combine them together to create new actionable data products necessary to a set of business use cases. I managed the slicing of this complex streaming graph into small sized, simplified `horizontal` of processing` different data to different dimensions of the information collected and managed by the product. A key part of my work is to allow our database to be really fast on writes, as we had data consistency issues while using a Postgres Sink to write produced data products. This forced us to use every capability of the Postgres technology available to cope with our needs.

Skills: Event Streaming, Cloud Engineering, Data Engineering, Software Engineering, SAAS, Data modelling, Data Pipelines, Process Optimization, Graph Architecture, System Redesign, Production Systems, Database First

- GCP
- Postgres
- Apache Beam
- Pubsub
- GCS
- Cloud Scheduling
- Advanced SQL
- High performance database oriented design
- Dataflow
- Oauth2
- Auth0
- Sentry
- Jira
- Terraform

Data Engineer / Technical Architect: Twist.eco (b2b) (London, United Kingdom)
September, 2022 - September, 2023

Technical Architecture, Cloud Engineering and Data Engineering for the Twist platform.

From one side, my work is to focus on the data engineering processes to integrate variegate customer data into our proprietary data model. We’ve created the whole system from scratch, it is versionable, multi-cloud datalake ecosystem where the major provider is GoogleCloud. Data transformations are run through the execution of workflows to carry out inter-dependent operations, provide consistent logging of actions and able of providing status updates at every step within the workflow. The orchestration runs complex sql queries (T), chained ETL/ELTs, daily imports, etc. into scalable units of processing, as the lowest cost per Gb solution available. The system is already hosting customers in production, is highly maintainable and allows the data science team to update the queries in production under certain I/O configuration constraints.

From the other, an important part of my work is to talk to our customers to design custom AdHoc solutions in order to integrate their data with our system. Along the process, we design the best solution according to the customer needs to fit time and technological constraints into a particular case. From there, our customer is able of leveraging our data model with his own business data.

Skills: Solutions Architecture, Data Engineering, Cloud Engineering, CI/CD, SAAS, Data Pipelines, Business Workflow, Fault tolerance

Aws s3

Aws Athena

Aws Postgress

Aws Lamda

Aws lam

Google Cloud

BigQuery

Workflows

Cloud Functions

Compute

Https Load Balancing

Rolling Updates

Cloud Storage

lam

Python

Numpy

Pandas

Oauth2

FastApi

Threading

Asynchroous Python

Python-Google-Cloud

Terraform

Pulumi

Freelancing / Various

Kubernetes

AKS

Flux

Flux2

Prometheus

Data Engineer: Namecheap Inc (b2b) (Phoenix, United States) September, 2020 - June, 2022

Data architecture and engineering on top of a Cloud Native environment.

Cloud Native systems make the data engineering solution design challenging, as resource consuming processes, which most computing processes are, must coexist with production grade services, therefore they must be optimized and well planned.

The Solution stands as the main data ecosystem architecture, design of workflows and process definitions. Operating thorough a set of Kubernetes clusters where product and infrastructure management solutions are hosted.

In order to enable new information flows, I was working on some components on top of K8s to monitor admin actions from 100K+ websites, basically to orchestrate data transformations. Later, we created the Cdn product, we started having more critical perfomance and the existing solution proven deficient sizing for loads.

We designed a new data and compute cloud where to place all data related workflows and processes. Among the tools we used there are Elasticsearch for short term access log storage, Confluent Kafka Streaming, transformations with Spark streaming, s3 object storage (in-house), several scheduler stacks, PrestoSQL for distributed processing among others.

Provided design of our first end-to-end Machine Learning learning workflows based on Pytorch framework and Trails server.

Skills: Solutions Architecture, Event Streaming, Data Engineering, Cloud Engineering, Agile Software Engineering, Continuous Delivery, Devops, SAAS, Code Shipping, Machine Learning, Non-Linear Regression

Presto

Apache Hive

Confluent Kafka

Airflow

S3

Elasticsearch

Filebeat

Logstash

Kibana

Metabase

Docker

Kubernetes

Prometheus

Helm

Pytorch

Trains

Python

Golang

Microservices

Gitops

Flux

Flux2

Data / Software Engineer: Auvik Networks (b2b) (Barcelona, Spain) April, 2019 - April, 2020

Providing Design and Engineering for a big Elasticsearch Datastore that will host more than 2000 indices. Also, the design of a dynamic system for optimal Elasticsearch maintenance. The project is for a Network Analysis company centered on solutions for Network service providers. Involved in a high performance syslog monitoring project, in a High Quality Software environment. The project is a microservice oriented architecture with Golang, Scala and Python as core technologies. I've designed and developed solutions for Golang microservices, worked with Apache Flink for message processing and created development environments based on Skaffold for fast Kubernetes deployment. My main achievement is to provide an architecture on Elasticsearch to store unforeseen amounts of data in a large scale cluster, while keeping Elasticsearch performance at an optimal level in every stage of the project evolution. Prepare an ecosystem for transparent Elasticsearch management in a production system, with seamless operations with no interruption for the ingestion and presentation pipelines while guiding the development to have robust insights of data, shaping the final product and triggering the start of few projects inside the company.

On the microservice architecture, I made the design of the microservice components directly interacting with the Elasticsearch cluster, both write and read services, and overlook their development from other colleagues.

Skills: Event Streaming, Solutions Architecture, Data Engineering, Scrum, Software Engineering, Test Driven Development, Engineering Best practices, SAAS, Code quality

- Elasticsearch
- Kibana
- Apache Flink
- Docker
- Docker-compose
- Kafka
- Kubernetes
- Microservices
- Skaffold
- Helm
- Gittlab pipelines
- Golang
- Scala
- Python
- Datadog
- Grafana

Data Engineer: PeakAdx by Aedgency (full time) (Barcelona, Spain) June, 2018 - February, 2019

Focused in the analysis of the company ingestion data pipeline and data engineering BI reporting processes. High load of advertiser and publisher traffic using Pop, CPA, CPI business models, up to 2B of requests per day. Providing workflow solutions to reduce the pipeline processing time, leverage the reporting process to a stable delivery of traffic activity, reduce infrastructure costs linked to mid-long term storage, reduce data processing costs and improve general software quality trough the improvement of data engineering team-internal processes.

Some solutions applied: YARN clusters for self management of Spark processing pipelines, updating Jenkins pipeline with a distributed architecture sharing the master home among slaves, Isolation of a Logstash environment out from the core server farm (up to 2 billion requests/day) , improve the design, do the project engineering of a DSP system for the company, S3 management to optimize operations for the 3TB stored per day.

skills: Distributed processing, High Throughput Pipelines, Data Engineering, Data Architecture, Big Data, Software Engineering, Devops, Scrum

- Hadoop
- MapReduce
- HDFS
- YARN
- Spark
- Spark Streaming
- Elasticsearch
- Logstash
- EC2
- SQS
- SNS
- S3
- Glacier
- VPC
- IAM
- Redshift
- Mongodb
- Redis
- Jenkins
- Zeppelin
- Metabase
- Kibana
- Hue
- python

Software Engineer: Shoesize.me (full time) (Barcelona, Spain) January, 2017 - January, 2018

Managing architecture, development and key aspects of the technological/social product scope, of a SAAS Customer Integration and Services platform which operates as a unified portal for Shoesize.me clients.

This project aims to make customer products available online, allowing users to integrate company services, manage their own product configurations, allow access to third party users, etc. I've designed the web service API for this project, this API has been created using Finagle, a great Scala technology build from Twitter, great for asynchronous communications. I've designed the platform as a software product, the client architecture has been created using React/Redux pair which offers powerful system structure, states and innovative interface/display objects. It combines several databases as the product platform makes use of a huge database to provide data reports, which are part of the commercial product of the company.

Skills: Full Stack Development, Solutions Architecture, Customer Facing, Software Engineering, Project Engineering, SAAS Product Engineering, System Integration

- Scala
- React
- Redux
- Finagle
- Rest Services
- Web Applications
- Mongodb
- Postgresql
- Janus Graph
- Redis
- AWS DynamoDB
- AWS EC2
- AWS ELB
- AWS CloudFront
- AWS S3

R+D Engineer: BBVA Data Analytics (b2b) (Barcelona, Spain) August, 2016 - December, 2016

Provided IT Architecture for the Information System in charge of performing large statistical analysis of big data managed using an Apache Spark architecture. Working together with the data science team, I managed to reduce the computational cost of nightly executions which take several hours of processing. Deeply study of model inference for a various set of modeling paradigms, deep study of neural networks design and neural networks learning processes, learning supervision or machine learning software models. As a result, many bottlenecks and risk areas present in the system execution where removed, the big process was sliced into smaller independent sub processes. This allowed us to parallelize the process into several pipes of data, to avoid system memory overflows, create sub process states which could be recovered and continued in case of crash, better monitor and supervise the learning process, and reduce the processing time.

Skills: Customer Facing, Data Engineering, Big Data, System Engineering, High Throughput Pipelines, Machine Learning, Supervised and non-supervised learning, Clusterization, Feature Extraction, Classification, Model Evaluation, Model Training, Model Re-training

- Python
- Multithreading
- Thread Safe Memory
- Celery
- Redis
- Mongodb
- Apache Spark
- AWS EC2
- AWS S3
- Scikit-learn
- SGD Classifiers
- Random Forest

Software Engineer: Ingravity Media (b2b) (Barcelona, Spain) September, 2013 - June, 2016

Working in several projects as Software Engineer, among which a remarkable one is to provide WOO Sports with robust IT Architecture. The project involved in planning, development, scalability of the Information architecture. Now our system is providing service to 15K mobile applications for kite surfers, continuously growing. Complete architecture, project engineering and development of the Http Rest Service for the mobile application, complete design of the code life-cycle across the vary system environments. Completely built using Php. Scalability and Security driven development, code maintainability and microservice oriented scalability.

Skills: Development Operations, Data Migrations, System Recovery, Distributed Systems, NoSQL, SQL, SOA, Software Engineering, Project Engineering, System Engineering, Distributed Logging, Logstash, Elasticsearch, High performance Systems, Fault tolerance, Autoscaling, CI/CD, Versioning Architecture, Web services, Rest services

- Gant
- Php
- Python
- Bash scripting
- Logstash
- Redis
- MongoDb
- Apache
- Nginx
- Oauth2
- Request throttle
- React
- jQuery
- Amazon Web Services
- AWS S3
- AWS EC2
- AWS Lambda
- AWS ELB
- VPC
- Rackspace

Software Engineer: Amadeus (full time) (Nice, France) September, 2011 - April, 2012

Software engineering in one of the most complex and big IT System in the world. It's offering robust service to critical operational needs from Airlines all over the world. I got involved in the impressive ticketing infrastructure, being aware of very complex software development processes, exigent quality checks, a huge organizational structure. The policies of software using really strict deploying policies, advanced unit and integration testing systems, as it runs a really hard-to-reproduce environment.

Worked in dense inter-team collaboration workflows, with long code integration processes. Amadeus is a huge Distributed System, compliant with several ISO Certifications, having wide complex node interactions, infinite use cases, always up and running.

skills: Software Engineering, Auditing Software Development Systems, Service Oriented Architecture, Systems Engineering, Automated deployment

C/C++ C++ STD Libraries Svn

Profile | Experience | Education | Languages

Politecnico di Milano (Milan, Italy)
Master's Degree Information and Communication Engineering
2007 - 2010

Universidad de Extremadura (Cáceres, Spain)
Master's Degree in Computer Science
2001 - 2010

Profile | Experience | Education | Languages

English	Fluent, Complete Professional Competence
Italian	Fluent, Complete Professional Competence
Catalan	Reasonable Personal competence
French	Reasonable Personal competence
Spanish	Mother tongue