Andres | Data Engineer

SUMMARY

Andrés is a Data Engineer specialized in data integration, transformation, and analytics.
He demonstrates strong experience with data integration tools such as Apache Airflow and Apache Spark, including building data processing pipelines and auditing data files on AWS MWAA. He also has hands-on expertise auditing data quality across multiple databases.

He has practical experience working with XML, XSL, FTP, and API-based processes, and is familiar with Salesforce for basic data extraction and loading, with a self-taught understanding. Andrés is proficient in managing both cloud and on-premises data workflows and brings additional experience in modifying machine learning models to run on AWS SageMaker, as well as deploying Python functions on AWS Lambda.

On the backend side, he is skilled in Python and has contributed to data processing tasks using PHP and C#. He also works with AWS cloud services such as Glue, Redshift, S3, Lambda, and RDS. He is proficient with databases including PostgreSQL, MongoDB, and MySQL, and is adept at implementing CI/CD pipelines using GitHub Actions and Azure DevOps.

He brings complementary skills in Microsoft Office and Google Suite, along with foundational knowledge of backend technologies like PHP and TypeScript. With experience in the banking and credit sectors, he is well-versed in managing sensitive data and delivering robust, scalable data solutions.

He is a responsible individual with excellent interpersonal and teamwork skills. His English level is B2/B2+.

CORE SKILLS

- MySQL
- MongoDB
- Github Actions
- Azure DevOps
- XML
- XSL
- FTP processes
- FTP

- API processes
- Airflow
- Spark
- Pyspark
- AWS Services: Glue, redshift, RDS, S3, Lambda, EC2, Cloudformation, Sagemaker, MWAA.
- ETL

- Microsoft Office
- Google Suite
- Python
- PHP
- Typescript

WORK HISTORY

Exomindset
Data Analyst
January 2025 - Present
- Audit data quality across multiple databases.
- Build and maintain data processing pipelines using Apache Airflow and Apache Spark, including auditing data files on AWS MWAA.
- Support backend development tasks for data processing using PHP and C#.
- Adapt and optimize machine learning models for execution on AWS SageMaker.
- Deploy Python-based serverless functions on AWS Lambda.

Univers*idad Simón Bolívar*
Data Engineer
September 2024 - Present
- Data analysis and predictive modeling project focused on predicting the behavior of Internet access in Colombia, using a dataset enriched with historical and projected data. The project involved from search and data cleaning to the implementation of machine learning models, achieving a significant improvement in predictive performance thanks to feature engineering.

Nuwebs
*Laravel Backend De*veloper
January 2024 - April 2024
- Implement User flow functionalities in the backend and frontend
- Create an intuitive, attractive, and functional User Private Zone.
- Implement a Typeform-style test system in the interface.
- Develop optimized access tests to be intuitive and easy to navigate.

- Development of a Back-office Web Platform:
- Develop optimized access tests to be intuitive and easy to navigate.
- Implement a User Management tool.
- Provide Data Analysis and Visualization capabilities, including user data visualization and filtering, lead tracking, and comparative analysis.

- Design a private area for post-paid users:
- Provide access to the Discord Club.
- Allow Advanced User Account Management.
- Offer Payment/Membership History and Management.
- Provide a Resource Library.
- Satisfaction Surveys.

- Development of some general modules:

- Implement an Affiliate Management System in PHP-Laravel 10.
- Create Event Calendars with membership level-based visualization.
- Develop a Provider and Certified Expert Management Module with advanced filtering and detailed profiles.

● Follow an agile work methodology with regular 1:1 Updates.

Accenture
*Clou*d Intern
July 2023 - January 2024

● Create Terraform scripts to provision AWS and Azure resources and for the backend, such as instances, databases, S3 buckets, and serverless functions.
● Define networking configurations using Terraform, including VPCs, subnets, security groups, and route tables.
● Implement infrastructure automation using Terraform modules for reusability and maintainability.
● Microservices deployment of the designated programming language for the microservice (Node.js, Python, Java) along with AWS SDKs or AWS Lambda for serverless functions.
● Containerized microservices using Docker and manage them using Amazon ECS (Elastic Container Service) or Amazon EKS (Elastic Kubernetes Service).
● Deploy a static frontend application (Angular, Vue.js) to an S3 bucket for static web hosting
● Set up CDN (Content Delivery Network) using AWS CloudFront for faster content delivery to end-users.
● Implement security best practices such as IAM roles, policies, and permissions to ensure secure access to AWS resources.
● Configure auto-scaling policies for backend services to handle varying traffic loads efficiently.
● Implement fault tolerance and high availability using AWS services like AWS Elastic Load Balancing (ELB) and AWS Auto Scaling.
● Set up CI/CD pipelines using AWS CodePipeline and AWS CodeBuild for automated testing, building, and deploying applications.

Impulsadata
*Data En*gineer
March 2020 - February 2022

● 5 ETL pipelines with AWS Glue and Pyspark for AWS Redshift Data Warehouse as data destination
● ELT pipelines for Data Lakes CSV and Parquet for HDFS
● SQL Tuning for slow BI queries and procedures
● Worked with No-SQL databases like MongoDB, Redis, and DynamoDB

- Worked with some relational databases like Postgresql, and Oracle databases to export data to Data Warehouse for OLAP purposes
- Query tuning
- Third-party API integration to extract historical data and stats

EDUCATION

Universidad Autónoma *de Bucaramanga*
System Engineer
July 2017 - April 2024

CERTIFICATIONS

- Python Course: Comprehensions, Functions and Error Handling.
- Git and GitHub Professional Course
- Database Fundamentals Course
- Missing Data Management Course: Imputation
- PostgreSQL Applied to Data Science Course

LANGUAGES

Spanish: Native
English: B2/B2+