LAMANSYS
REALISE YOUR IDEAS

lamansys.com.ar
Tandil, Buenos Aires - Argentina

JUAN CAMILO DIAZ ORTEGA

[MEDELLÍN, COLOMBIA]

BIG DATA ARCHITECT

System Engineer with +22 years of experience in Software Development, 15 years as Software Architect, 12 years in Big Data and streaming architecture (Architect and Developer), 9 years leading & coordinating development teams (growing and mentoring). Big Data Architect, On-Premise and on Cloud. Multi-language developer. Graph databases implementations and Query Languages on them (CypherQL, AQL, SparkQL, GraphQL), Big Data, and Machine Learning passionate. Love to work with Scrum and Kanban Methodology. Strong management skills to handle multiple interdisciplinary teams. Great flexibility to adapt to cross-cultural environments. Eager to learn and develop into different types of platforms. Team player with strong communication skills.

</> TECHNICAL SKILLS

Programming Languages	Java, Python, .Net, JavaScript, Scala, PHP, HTML5, CSS3
Libraries and Frameworks	Spring, Hibernate, Ibatis, Django, Grovy, Grails, JQuery, AngularJS, ReactJS, D3, NodeJS, Bootstrap
Big Data	Hadoop, HDFS, MapReduce, Hive, Sqoop, Pig, Oozie, Falcon, Drill, Eagle, Impala, Phoenix

lamansys.com.ar
Tandil, Buenos Aires - Argentina

</> TECHNICAL SKILLS

Realtime / Streaming	Spark, Scala Akka, Apache Flink, Apache Kafka, Kafka Streams, Apache Flume, Apache Beam, Apache Storm, Apache Samza
NOSQL DataBases	Hbase, Cassandra, MongoDB, Redis, DynamoDB, Apache Ignite, Scylla, VoltDB, Neo4j, OrientDB, TigerGraph, ArangoDB, AWS Neptune, Cypher QL, AQL, GraphQL, SparkQL
Distributions	Cloudera, Hortonworks, Apache
Cloud Platform	AWS Big Data Tech, GCP, Azure Big Data Tech
Machine Learning	Spark MLlib, Tensorflow, Torch/PyTorch, Keras, Theano, Scikit-learn
Relational Databases	SQL Server, Oracle, MySQL Postgre
Integration tools	Mule, WSO2, SAP PI
Methodologies	TDD / ATDD/ BDD - Agile (Scrum, Kanban)
Others	Virtualization and Containers, CI/CD, Jenkins,DevOps, BPM, BPMN, SOAP, Rest Services, Apache Tinkerpop 7 Gremlin, Microsoft Cognitive Services, PowerBI, GCP Composer/Apache Airflow, Zeppelin, NiFi, Ignite, Kylin, Druid, Tableu,dbt

lamansys.com.ar
Tandil, Buenos Aires - Argentina

PROFESSIONAL EXPERIENCE

FREELANCER - SOFTWARE ARCHITECT, BIG DATA ARCHITECT, TECH ADVISOR
SINCE DEC 2020

- **Business Area:** Various
- **Role and Responsibility:**
  - Cloud Big Data Architecture (AWS, Google Cloud Platform, Azure)
  - Lambda and kappa architecture definitions.
  - Recommender System (Machine Learning)
  - Health care system and information analytics using Hadoop, Spark, Kafka, Cassandra platforms.
  - Building real time event processing system based on spark streaming to handle real time retail information and predicting trends.
  - Building twitter sentiment analysis on Spark (Python and Scala).
  - Consultant on Streaming Architecture, Real Time Streaming, Real-time Data Analysis using, Kafka, Apache Pulsar, Spark Flink, Samza, Storm
  - Consultant on Data Lakes implementation and Real-Time Data
  - Graph Databases Consulting Services
  - Build a clickstream distributed streaming System using Play Framework, Akka, Apache Kafka and Python or Scala
  - Fast Data call center project using Play Framework, Akka, Scala Kafka, Spark and NLP
  - ETL implementation systems using streaming architecture Kafka and Spark
  - Data Lake Consulting Services
  - Ingestion (streaming and batch) process Consulting Services
  - Big Data Visualization Consulting Services
  - Face and Image Recognition Consulting Services
  - NLP Consulting Service
  - Machine Learning Implementation tools Consulting Service (Not Data Science)
  - Talent development and Talent selection. Coaching teams. Training for better development practices
  - Project: Data Lake migration - Falabella Holding - Role: Data Architect Subject Manager Expert Data Warehouse migration to Data Lake architecture using Google Cloud Platform. Cloud Storage, BigQuery, Pub/Sub, Kubernetes, BigTable, Cloud Spaner. Support to Machine Learning process with GCP Stack (AI Platform)
  - Project: Data Lake migration - Belcorp - Role: Data Architect Subject Manager Expert Data Warehouse migration to Data Lake architecture using AWS.
  - Project: Data Lake migration - DirectTV - Role: Data Architect Subject Manager Expert Data Warehouse migration to Data Lake architecture using AWS and AWS Lake Formation.

lamansys.com.ar
Tandil, Buenos Aires - Argentina

PROFESSIONAL EXPERIENCE

GLOBANT - DATA ARCHITECT SUBJECT MANAGER EXPERT - JAN 2016 - NOV 2020

Project: Ford Digital Transformation - **Business Area:** Automotive

- **Project Description:** Creation of Data Lake on premise architecture (Cloudera/Hortonworks) with the aim of unifying data to provide a solution to the implementation of initiatives via Machine Learning sales forecast analytics, pricing analytics for vehicles, pricing analytics for parts, employee data management system.
- **Role:** Data Architect Subject Manager Expert

Project: Sodimac One House Audit - **Business Area:** Retail

- **Project Description:** Audit data ingest processes in order to certify the processes that go from the sources to the final transformation zone, guaranteeing the suitability of the data, allowing to establish whether it is possible to leave the certified use of Cloudera. Audit performed on Google Cloud Platform, in Composer (Airflow), BigQuery, GKE tools.
- **Role:** Data Architect Subject Manager Expert

Project: Oportun PoC GCP BigQuery migration from AWS - **Business Area:** Retail

- **Project Description:** Running PoC with the goal of showing the processing capabilities of BigQuery vs Redshift, that allows Oportun to consider a migration to BigQuery as an analytical platform.
- **Role:** Data Architect GCP Subject Manager Expert

Project: CMPC Cloud Implementation - **Business Area:** Forestry

- **Project Description:** Advice and Accompaniment to the CMPC company (Chile) about valuation of cloud providers (AWS, Google Cloud, Azure) with the aim of implementing Data Lake. Detail of the options that each Provider has and how they match what CMPC requires.
- **Role:** Data Architect Subject Manager Expert

Project: Direct TV - **Business Area:** Television

- **Project Description:** Data Lake Implementation using AWS. Consolidate all sources into S3 as scalable storage and create ETL / ELT process using AWS Glue (Crawlers, Jobs), Lake Formation Implementation.
- **Role:** Technical Leader / Data Architect Subject Manager Expert

lamansys.com.ar
Tandil, Buenos Aires - Argentina

PROFESSIONAL EXPERIENCE

GLOBANT - DATA ARCHITECT SUBJECT MANAGER EXPERT - JAN 2016 - NOV 2020

Project: Metadata Integration (PoC) - Falabella Holding - **Business Area:** eCommerce

- **Project Description:** Implement Metadata processes in Google Cloud Platform apply Data Governance, Data Dictionary, Data Lineage and Data Security principles using Graph Databases and Gremlin (Apache Tinkerpop) as query Language. JanusGraph + BigTable, JanusGraph + Cassandra + Elasticsearch.
- **Role:** Data Architect Subject Manager Expert

Project: Data Lake Migration - Falabella Holding - **Business Area:** eCommerce

- **Project Description:** Data Warehouse migration to Data Lake architecture using Google Cloud Platform. Cloud Storage, BigQuery, Pub/Sub, Kubernetes, BigTable, Cloud Spaner. Support to Machine Learning process with GCP Stack (AI Platform).
- **Role:** Data Architect Subject Manager Expert

Project: Adaptive K12 - **Business Area:** Education

- **Project Description:** Create Big Data and Streaming architecture with AWS stack in order to replace Intellify platform. Technologies used: Kinesis, Athenea, Redshift, Redshift Spectrum, S3, Lambdas with Python, SQS, SNS, RDS, DynamoDB, Neo4j & Cypher QL, AWS Neptune, Elasticsearch, EMR, Glue, EMR Notebooks. Capture and processing all type of data through streaming for e-learning platform, Intellilfy platform integration. Architecture Definition, Developer (Java / Scala / Python), PoC developer and architect. Developed multiple MapReduce jobs in java for data cleaning and preprocessing. Moving data from HDFS to RDBMS and vice-versa using Sqoop. Analyzing/Transforming data with Hive and Pig.
- **Role:** Data Architect, Software Designer - Technical Leader

Project: Falabella - **Business Area:** eCommerce

- **Project Description:** GCP Cloud Dataproc, BigQuery, Dataflow, Pub/Sub streaming processes.
- **Role:** Data Architect, Software Designer - Technical Leader

PROFESSIONAL EXPERIENCE

GLOBANT - DATA ARCHITECT SUBJECT MANAGER EXPERT - JAN 2016 - NOV 2020

Project: POC for IoT streaming process - **Business Area:** Television

- **Project Description:** GCP, Pub / Sub, Coud Functions, Cloud Storage, BigTable, DataStore, Dataflow, Dataproc, BigQuery.
- **Role:** Data Architect, Software Designer - Technical Leader

Other functions:

- Consultant on Hadoop / Yarn / Map Reduce Data Architect projects.
- Consultant Streaming architectures Spark, Flink, Apache beam, Apache Kafka, Apache Pulsar.
- Consultant on Fast Data projects using Scala and Akka with Play Framework.
- Consultant on SMACK Stack (Spark, Mesos, Akka, Cassandra, Kafka)
- Consultant on architectural patterns for data-intensive solutions.
- Consultant on Data Lakes Implementation and Real-Time Data.
- ETL implementation systems using streaming architecture Apache Kafka and Spark.
- Team Technical Leader IA Labs Projects Face Recognition using TensorFlow, Pythorch, Python.
- Galaxies Detection IA Labs Projects using TensorFlow, Pythorch, Python
- CATS IA Labs Projects, NLP Recruiting Projects, Using TensorFlow, NLTK, Python, Neo4J.

TECH AND SOLVE - JAN 2015 - JAN 2016

Project: Cliente 360 - AFP Crecer El Salvador - **Business Area:** Finance

- **Project Description:** System Consolidation of the different sources of existing data. Architect and developer, Team leader, Continuous improvement team, achievement orientation, Continuous Integration and Continuous delivery process. Reduce attention times and consolidate client search in a single tool. .Net Technology, SQL Server, DB2, Oracle databases. Scrum and Kanban methodology.
- **Role:** Architect

PROFESSIONAL EXPERIENCE

TATA CONSULTANCY SERVICES - TECHNICAL LEADER - MAY 2014 - JAN 2015

Project: Alianza Gestión Documental - **Business Area:** Education

- **Project Description:** Application document management and report integration. Architect on Hadoop, integration with Sqoop and Pig tools. Streaming process with Apache Kafka, Apache Spark (Scala). Java (Spring Framework with Hibernate), Mule, Oracle, MS SQL Server, MongDB. Cascade methodology.
- **Role:** Tech Leader, Team Lead, Tech Coach

Project: Integracion - Proteccion - Role: Team Leader

- **Project Description:** Integration System through IBM WebSphere tools. Team Leader. Definition and Creation integration task. Scrum and Kanban methodology.

Project: ETL Implementation - Banistmo - **Business Area:** Banking

- **Project Description:** ETL implementation systems using streaming architecture Kafka and Spark with Python.
- **Role:** Data Architect

COMFENALCO ANTIOQUIA - DEVELOPER, ARCHITECT, TEAM LEADER - AUG 2003 - JUL 2013

Project: Historia Clinica / Siscomf IPS - Biosigno IPS Comfenalco

- **Project Description:** Development, architecture definition for Electronica Medical Records (EMR) and all IPS System.
- **Stack:** Java (Spring Framework with Hibernate), JEE, Scala, PHP, Mule, MongoDB.
- **Role and functions:** Developer, Architect, Team Leader, Technical Leader.
- Architect on Hadoop, integration Sqoop and own ETL (scala programing)
- Streaming process with apache kafka, apache Spark.
- Oracle, MS SQL Server, Informix, MongoDB. Cascade methodology.
- Building real time event processing system based on spark streaming to handle real time EMR information.
- Installed and configured Hadoop Cluster in Test and Production environments
- Building Spark and Spark SQL based platform for financial analytics.
- Analyzing/Transforming data with Hive and Pig.
- Building Log Streaming Analysis with Spark and Hadoop
- Fast Data call center project using Play Framework, Akka, Scala, Kafka and Spark.
- Implementation of requirements for the solutions applied to Caja de Compensacion. Java (Spring Framework with Hibernate), JEE, scala, PHP, Mule, Hadoop integration Sqoop and own ETL (scala programing) Streaming process with apache kafka, apache Spark Oracle, MS SQL Server, Informix, MongoDB Kanban methodology.

LAMANSYS
REALISE YOUR IDEAS

lamansys.com.ar
Tandil, Buenos Aires - Argentina

PROFESSIONAL EXPERIENCE

COMFENALCO ANTIOQUIA - DEVELOPER, ARCHITECT, TEAM LEADER - AUG 2003 - JUL 2013

Project: SAP Implementation

- **Role:** Integration DEveloper, Integration Architect, Integration Team Leader
- **Project Description:** Implementation of requirements for the solutions applied to Integration process with legacy applications (EPS, IPS, Caja de Compensacion). SAP, SAP Integration Process (SAP PI), ABÂP. Java (Spring Framework with Hibernate), JEE, scala, Streaming process with Apache Kafka, Apache Spark, Oracle, MS SQL Server, Informix. ASAP methodology.
- Development teams. Coaching teams. Talent development. Talent selection. Support for good development practices.

EDUCATION

POLITÉCNICO JAIME ISAZA CADAVID, Medellín, Colombia - Sept 2000

Technology in **Data Systematization**

UNIVERSIDAD INCCA DE COLOMBIA, Medellín, Colombia - Dec 2006

Bachelor's Degree in **System Engineer**

INSTITUTO TECNOLÓGICO METROPOLITANO, Medellín, Colombia - Apr 2013

Specialist in Project, Formulation and Evaluation

CERTIFICATES

PROFESSIONAL DATA ENGINEER, Google Cloud Certified (2021-2023)

Certification ID: m7yTnI (https://www.credential.net/fee7f4a7-cc51-4498-b439-89754a766ac8)

PROFESSIONAL CLOUD ARCHITECT, Google Cloud (2021-2023)

Certification ID: eXuSGW (https://www.credential.net/9ff46e62-9ee3-48db-9ed4-0909e3075046)

LANGUAGES

English: Upper-Intermediate | **Spanish:** Native