

JORGE CARHUARICA

Sr. Data Engineer | BSc Computer Science

@ jcarhuaricaa@uni.pe github.com/jc-arhuarica
📍 Lima, Peru in linkedin/jorge-carhuarica 🌐 https://jc-arhuarica.com/

I am a data engineer and I have development experience on big data technologies such as Hadoop, Snowflake, NoSQL(MongoDB, Hbase), Kafka, Spark. I followed a high level training in computer scientist which allows me to acquire advanced reasoning methods and work rigour.

🎓 EDUCATION

Jul 2012-Apr 2016 Bachelor of science, **Computer Science**, National University of Engineering.

📁 EXPERIENCE

January 2023
Current

Data Architect, QUALIFACTS.

Improve of the Analytics Advance Data platform for Carelogic

- > Optimized data loading incremental process using a pruning strategy in Snowflake tables with xml content, reducing time by 80%.
- > Optimized data loading incremental process using Snowflake Tasks instead of Snowpipe and S3 connector instead of Snowflake connector, reducing cost by 30%.
- > Experience in creating objects Views, Schemas and Databases, Warehouse, Task, SP's and managed users based on RBAC.

Snowflake Task Snowpark Snowflake Stage Snowpipe DBT MSK Kafka Connectors

April 2020
January 2023

Data Engineer, QUALIFACTS.

CAC Project (Cloudera - PAAS) - The Analytics Data platform for CareLogic (ERP platform for mental health clinics)

- > Involved in designed and implemented a pipeline for the migration of the Operational Data Store(ODS) tables from Oracle to MariaDB by a streaming data processing under Lambda Big Data Architecture. Finally on hive there are external tables pointing to MariaDB(last 30 days) and orc historical tables.
- > Involved in designed and implemented a pipeline for the migration of the configurable forms(CF) tables from Oracle to Hbase by a streaming data processing under Lambda Big Data Architecture where semi-structured data is desnormalized. Finally on hive there are external tables pointing to MariaDB(last 30 days) and orc historical tables.
- > Involved in designed and implemented a pipeline for the migration of materialized view(MV) from Oracle to MariaDB by a micro batch data processing.
- > Responsible for implementing Python scripts automation for different initiatives (Looker Api Standup, DML and DDL hive scripts Standup, Incremental Spark Processing).

CAA Project (Snowflake - SAAS) - The Analytics Advance Data platform for CareLogic (ERP platform for mental health clinics)

- > Involved in designed and implemented a pipeline for the migration of the Operational Data Store(ODS) tables from Oracle to S3 by a Talend Job. Finally generate snowflake tasks to copy data from S3 into Snowflake tables.
- > Involved in designed and implemented a pipeline for the normalized transformation of the Operational Data Store(ODS) tables from RAW into Curated Layer by a DBT job.

Apache Hive Apache Oozie Apache Sqoop Apache Hbase Apache Kafka Apache Storm Apache Spark Talend
SQLines Data Oracle MariaDb MongoDB Bash Scripting Python Java Scala Groovy Jenkins Gitlab
Snowflake

<p>Oct 2019 March 2020</p>	<p>Data Engineer, INDRA. Personal Financial Management(PFM) Datio Project - Banco Bilbao Vizcaya Argentaria (BBVA)</p> <ul style="list-style-type: none"> > Responsible on the dictionary process of tables according the the standard rules on Sophia. > Responsible for implementing and deploy the ingest artifacts by Skynet. > Responsible for implementing the dependencies between the jobs by meshes on control-m. > Responsible for implementing quality of rules by Hammurabi in order to validate the quality of the data. > Developed Spark applications using Pyspark and Spark-SQL for data extraction, transformation and aggregation from multiple file formats for analyzing and transforming the data to uncover insights into the customer usage patterns. <p>HDFS Apache Spark Datio Scala Control m Bitbucket</p>
<p>Jun 2018 Sep 2019</p>	<p>Data Engineer, EVERIS. Proyecto Real Time : COE - Center of Expertise - Banco de Crédito del Perú (BCP)</p> <ul style="list-style-type: none"> > Involved in the implementation of a use case for failed transactions from customer without funds with good credit historical in order to offer cash in advance by a streaming architecture. > Creation of the schemas and tables required and perform data ingestion for required tables with IBM DataStage. > Creation of an OBt by processing of tables about POS Transactions, Demographic, BCP Cards, TC Billing, Channel Transactions, Campaigns. <p>Proyecto Matriz Variable : AEMA - Area Analisis Estrategico Y Modelos Avanzados - Banco de Crédito del Perú (BCP)</p> <ul style="list-style-type: none"> > Implementation of predictive statistical variables for propensity, churn, Segmentation, Fraud Detection, Analytical Marketing models. > Involved in the designing and implementation for the migration of business logic on stored procedures from Oracle to one big table (OBt) by implemented spark code transformations. <p>Proyecto Datalake : Banco de Crédito del Perú (BCP)</p> <ul style="list-style-type: none"> > Implemented new artifacts on DataStage in order to ingest tables from Oracle, semi-structure files (json, xml) to HDFS. > Creation of FDS, FPH, MIS formats for Certification and Production passes. > Creation of DDL and DML files for hive internal and external tables with partitions and Hql queries through HUE. <p>HDSF Apache Hive Apache Spark IBM DataStage Python Gitlab</p>

TECHNICAL SKILL SET

BIG DATA

Data Processing	Hadoop(Map Reduce, Hive), Spark.
Data Processing Cloud	Databricks
Data Warehouse Cloud	Snowflake
Data Ingestion	Sqoop, Kafka, Talend Studio, IBM DataStage, SQLines Data.
NoSQL Database	Hbase
Workflow Scheduler	Oozie, control-M
Hadoop Distribution	Apache, Hortonworks, EMR.
Visualization	Looker

CLOUD PLATFORM

AWS EC2, S3, RDS, EMR, MSK

PROGRAMMING LANGUAGES

Language Python, Scala, Java, Groovy, Bash.