

# Luis Kevin Huaman Bolanger

Flat 11, Clarence House, 152 North Row  
Milton Keynes  
MK9 1AY

Mobile: +44 07931 986 641

Email: [lk.huaman@gmail.com](mailto:lk.huaman@gmail.com)

## **EDUCATION**

### **September 2009 – Degree in Computer Engineering, Polytechnic Univ. of Madrid.**

- In 2014, I achieved a scholarship in distributed systems research laboratory for implemented a paper published in 2010 [Pepper: An Elastic Web Server Farm for Cloud Based on Hadoop](#).
- Received training in Zookeeper Development, HDFS High Availability configuration in Cluster. (60 Hours)
- Received training in Configuration and installation Hadoop on the Cluster UPM (10 hours).
- Receive training in Pfsense Firewall Configuration (UPM, 4 hours).

### **October 2016 – November 2016. Scholarship in Everis Spain.**

- Workshop Qlikview (Everis, Onsite, 4 hours)
- Workshop Scala (Everis, Onsite, 4 hours)
- Hadoop Development Training (Everis, Onsite, 16 hours)
- AWS Development Training (Everis, Onsite 40 hours).

### **December 2017 – March 2017. Master in Pragis Technology.**

- Hadoop Essentials, Pig, Hive, Hbase, Flume, Kafka, Ozzie, Hue, Impala, Zookeeper, Spark, Cloudera Manager, Kerberos, Kibana, Logstash, Kibana.

### **December 2018 (4 days) - Training ADM-221 HDP Operations: Administration Foundations**

- My company paid a formation related with design, installation, configuration, and management of the hortonworks Data Platform ([HDP](#)).

### **July 2019 (3 days) - Training Kafka**

- Completed the Confluent Developer Training.

## **PROFESSIONAL EXPERIENCE**

### **December 2020 – February 2021. BigData developer in Santander UK Technology**

**Money Manager:** It was an internal initiative for Santander Innovation with the collaboration of Personetics IA company (Third-party), the objective was to generate an amplified spectrum of insights about the spending behaviour of the user and different movements from one day to another over their banking accounts.

Responsible for the following tasks in the project:

- Analysis of sources of data to process. Dockerization of IA software solution of Personetics & deployment in Openshift.
- Develop deployment scripts.
- Dockerization of SQL Server infrastructure in Develop and deployment in Openshift.
- Optimisation and audit in transformations queries using Hive SQL.
- Implantation of the model of data in MS-SQL Server.
- Adaptation of Sqoop scripts to bulk data into MS-SQL-Server.

### **April 2018 – November 2020. BigData developer in Santander UK Technology**

**Industrial Aggregator:** The initiative for The Santander Innovation team was to build an app over Open Banking API, that allows the user a global vision and unified about all personal and business accounts in different banks. The user could access and check out in one unique point, balances, transactions for their banking accounts in real-time, it is integrated into the app.

Responsible for the following tasks in the project:

- Develop the functionality of token renovation in spring boot 2.X.X & use of the Kafka-stream library in java.

- Implementation of Redis cache using spring boot 2.x.x. Implementation of acceptance test in groovy and spring boot 2.X.X to verify case of uses with Kafka.
- Develop a service of monitorization of consumer lag for all services.
- Support to the team in docker tasks Support and modification functionalities in services made in Scala with spring boot 1.X.X.
- Support to the team in docker tasks Support and modification functionalities in services made in Scala with spring boot 1.X.X.
- Implementation of delete Service in Hbase made in Scala and spring boot 1.X.X

#### **December 2017 – February 2018, BigData developer in Strategybigdata**

**RSI Smart-logs Phase 2:** Create an infrastructure in the cloud, installation and configuration of Hortonworks with Ansible, Yarn queue configuration, views of ambari, user management, improvement of cluster configuration and implemented best practices.

Responsible for the following tasks in the project:

- Installation of a cluster hortonworks in AWS with Ansible.
- Defined a system of Yarn queue in the cluster to according to the jobs executed in the cluster.
- Inspection of cluster configuration and make a plan of growth for add new nodes.
- Complete analysis of cluster production and creation of recipes to improve the cluster, in different layers, like system operative, network, hdfs, yarn.

#### **September 2017 – December 2017, BigData developer in Strategybigdata**

**RSI Smart-logs:** The objective of the project was to establish a login platform that allows to find the traceability of the life cycle of a transaction through the different systems of the client's architecture. The project covered two types of applications, one application in real-time and another in batch mode. Responsible for the following tasks in the project:

- Re-designed the data lake in the cluster.
- Defined the elastic mapping according to the needs of the business.
- Built the architecture in docker-compose to get agility in the development.
- Installation and configuration of Logstash in the cluster, optimising the component and agent.
- Installation and configuration of agent flume in the cluster to accept stream data from other source.
- Developed a process of parsing xml files in spark to apply a logic defined in a business dictionary.
- Implemented a process of filtering and masking confidential data in spark streaming.
- Defined a DDL for staging data processed of spark job, for quick access from hive.

#### **June 2017 – September 2017, BigData developer in Strategybigdata**

**RSI SentimentAnalysis:** Implementation of a Real-Time system of sentiment analysis for digital media and social networks.

The objective of the project was to analyse, identify and build the real image of the client in the digital domain, based on the content shared on the network. The base architecture built serves as a preamble to potential behavioural analysis, an identification of the degree of trust among its users, analysis of the strength of the relationships among the members of its community, interests, propagation patterns of messages, etc, allowing to know the latest trends in their environment and to create marketing campaigns aimed at more specific targets or increase the potential of visualisation of campaigns.

The development of the project was based on the extraction of the digital context (digital media, twitter) of the RSI client environment (Rural Computer Services), using technologies such as Apache Flume and Scrapy as integers agents, Apache Kafka for event storage, Spark Streaming for processing, elastic-search for the indexing of information on demand and kibana as a graphical visualisation tool for data.

Responsible for the following tasks in the project:

- Designed the Crawler ingest architecture.
- Built a specific architecture in docker-compose with all components and improve the development and integration test.
- My work consisted in develop a job in spark streaming (real-time) that get all information stored in Kafka topic, making request to API Meaning cloud to determines the sentiment of news, API MOZ to get the position of media similar to Alexa, until finally writing in Elasticsearch.
- Creating a Web Cache in Hbase for media storage processed of crawler.
- Installation and configuration of elastic-search in cluster mode.
- Defined the elastic mapping according to the needs of the business.
- Configuration of agent flume, for ingest tweets in Kafka with a little modification in twitter source for accept some special character.

### **December 2016 – June 2017, BigData developer in Pragsis Technologies**

**IFRS9MI:** Implementation of an information exploitation system according to current IFRS9 regulations applied to local and corporate countries in ISBAN Santander.

The aim of the project was supply the analysis the data to the users, providing an advanced reporting system with open universe characteristics, granting the possibility to build detailed reports on demand, management of aggregated and historical information with multidimensional analysis, taking into account the nature of the data, traceability of the operations carried out on it and homogenisation of the golden sources, ensuring a 360° vision that allows self-provision of adjustments and risks according to the current regulations.

Responsible for the following tasks in the project:

- Design the architecture ETL, including integrating of technical validations and functional validations.
- Extended the technical validation module from SUPRA, made in java and scala to adjust them to the requirements of the project.
- Designed DDL definition according to business requirements In Hive.
- My tasks included management of the work team.

### **October 2016 – December 2016, Solution Assistant BigData in Everis, Madrid, Spain.**

**ESIR:** Project ESIR (Intelligent System of electrical network) in Electric Company Gas Natural Fenosa, XML Parsing and validation tag files according to established business rules.

Responsible for the following tasks in the project:

- Development of a XML parsing process and validation tag files according to established business rules in python, using AWS Ec2 and AWS SQS to get information (AWS SQS similar to Kafka).

### **July 2016 – September 2016, Solution Assistant BigData in Everis, Madrid, Spain.**

“Project inception” in Electric Company Gas Natural Fenosa, migration of information in the Amazon Cloud, as their primary objectives, allowing the availability of information in AWS Redshift, data validation tasks, performance testing and generate the automation of loading process with Luigi. Preprocessing and load files AWS Redshift, repeated data filtering and creation of Aggregate tables. Responsible for the following tasks in the project:

- Development of process of preprocessing and loading of files in AWS S3 and load in Redshift, AWS files in Redshift, performing the repeated filtering of data and the creation of aggregate tables.

## **November 2015 – June 2016. Solution Assistant BigData in Everis. Madrid. Spain.**

**Big Data MKT CEPSA:** Analyst and developer in the implementation of the new analytical platform “Big Data Commercial MKT” for the Spanish Gas Company CEPSA, That new system had two main objectives, unify all the previous system of the company and store the daily transactional information allowing to do crossed analysis.

Responsible for the following tasks in the project:

- Developing automation scripts in Bash for Cloud services like AWS s3 for deleting multiples objects in parallel over buckets (S3).
- Developing automation scripts in Bash load files from S3 to Redshift.
- Configuration in AWS Console a alarm for start and stop architecture of processing Amazon SNS.
- Development of automated flow orchestration of the different services and processes of ETL in the cloud with time restrictions and fault tolerance (Datapipeline).
- Development process concatenating files made in Spark (Scala).
- Spools development process to the various sources of data in Oracle, Installation and configuration of a SQL-Client Oracle in Ec2.

## **TECHNICAL SKILLS**

- Git, bash, Java, maven, Spring boot 2.X.X, Spring boot 1.X.X., Jenkins, Openshift, groovy, Ms-SQL-Server.
- linux (bash), AWS Ec2, AWS EMR, AWS Datapipeline, AWS S3, HDFS, Hive, maven, java (SUPRA), scala, python, AWS Redshift, docker, docker-compose, scrapy, elasticsearch, logstash, kafka, spark (pyspark), spark-streaming (pyspark), hortonworks, ambari, yarn.

## **CERTIFICATIONS**

- Development Distributed Applications using Zookeeper (Big Data University).
- Hive Advance (Big Data University).
- Hortonworks Administration

## **TRAINER**

### **February 2018 – BigData developer in Strategybigdata**

Imparted education in Hadoop Essentials, Hive, Hortonworks, Ambari, Cloudera, Cloudera Manager to applicant in my company.