

## Big Data Technology for monitoring ICT service data

Marcelo Dante Caiafa<sup>1</sup> , Ariel Aurelio<sup>1</sup> , Adrian Marcelo Busto<sup>1</sup> 

<sup>1</sup> Universidad Nacional de La Matanza, Buenos Aires, Florencio Varela 1903, 1754 Buenos Aires, Argentina

{macaiafa,aaurelio,abusto}@unlam.edu.ar

**Abstract.** Data analysis has become an important source of knowledge for organizations. An adequate treatment allows to obtain valuable information. Its massive processing is possible from Big Data technologies.

The work is based on the use of an open source platform for the processing of files generated by the communication systems of a mass service institution with three hundred branches that serves more than two million customers.

The research addresses the need to consolidate results that add value to decision-making and improve the operational efficiency of information and communication technology (ICT) services.

The objective is the development of a control panel based on measurement of key indicators. It will allow the monitoring of its operating costs and the level of quality of customer care. For this, the ELK (Elasticsearch-Logstash-Kibana) set is used, fed with the call detail records known as CDR (Call Detail Records).

**Keywords:** Big Data, ICT, CDR, ELK.

### 1 Introduction

Big data offers ICT engineers a real opportunity to capture a more comprehensive view of their operations and services [1]. Big data analytics is a set of technologies and techniques that require new forms of integration to disclose large hidden values from large datasets [2]. As an example of different use cases based on CDR analysis can be mentioned, operational efficiency and improvement of the customer experience [3].

This work aims to respond to the need to analyze the operating cost of a telecommunications infrastructure of a large organization and the level of quality of the care services it provides. The result is a dashboard with consolidated information built with ELK technology, from the processing of CDRs generated by its telephony servers.

The research work focuses on the process of developing a dashboard that consolidates main indicators according to the following objectives:

1. Analysis of operating costs based on network traffic data flows according to service monitoring, to detect fraudulent behavior when it occurs.
2. Monitoring of the quality of care services, resulting from the representation of operation time for each of the interactions.

## 2 Work Development

The development of the work is structured in 5 fundamental stages. In the first stage, the detail of the communications system were carried out to know the data source. In the second stage, the CDR records are obtained and a data dictionary is prepared. Stage 3 deals with indexing the database. Each CDRs field with relevant information is assigned a specific type of parameter from text file to JSON document. Stage 4 runs the ETL process. A text file is build from which the Logstash module is configured for data ingestion. In the last stage, searches are carried out according to the specific objectives to be achieved, so Kibana module allows them to be consolidated into a control panel.

### 2.1 Data source's context and relevance

The data was provided by an organization with more than two million clients dedicated to mass consumption services. Its products are heavily regulated, that is why the competitive strategy focuses on differentiation based on the quality of customer service. It enhances the value of this work. The linking process to productive environment was necessary in order to know the details of service model and the infrastructure that supports the business services. These activities were key to understanding the technical architecture of the platform, the different models of care it supports, and the interpretation. The CDR data is used for collection, settlement, billing, network efficiency, fraud detection, value-added services, business intelligence, etc [6].

### 2.2 Database indexing

To build the database, you must create an index in Kibana. To do this, it is necessary to define the field type when formatting the data structure that is expected to be received. This is done through the DevTools section with PUT command.

```
PUT /cdr2020DBv2 {
  "mappings": {
    "properties": {
      "cdrRecordType": {"type": "integer"},
      "globalCallID_callId": {"type": "integer"},
      "origLegCallIdentifier": {"type": "integer"},
      "dateTimeOrigination" : {"type": "integer"},
      "dateTimeOrigination_formatted" : {"type": "date"},
      "dateTimeConnect_formatted" : {"type": "date"},
      "origIpv4v6Addr" : {"type": "ip"},
    }
  }
}
```

### 2.3 ETL Processing

The ETL (extract, transform and load) is the process of collecting data, adapting its fields and loading data to the base. This is done by configuring the logstash:

```

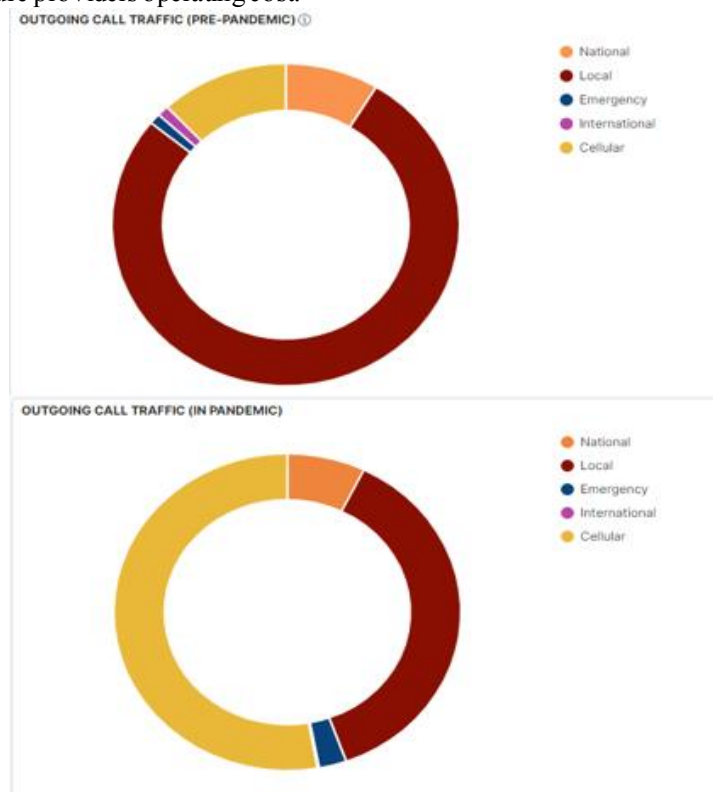
input {file {path => "C:/Users/unlam/CDR2020.txt"
  start_position => "beginning"} }
filter {csv {columns => ["cdrRecordType",...]}
  date { match => ["dateTimeOrigination","UNIX"]
  target => ["dateTimeOrigination_formatted"}
  date {match => ["dateTimeConnect", "UNIX"]
  target => ["dateTimeConnect_formatted"}
output {stdout {}
  elasticsearch {index => "cdr2020DBv2"} }

```

Three instances are identified. The input consists of indicating the file path to extract the data. The next step is the filter where all fields are listed in comma separated format. In the particular case of dates, the data received in UNIX format can be converted to a data type. Finally, the output indicates the indexname where the data will be loaded.

### 3 Results

The parameter "CalledPartyNumber" was used to classify voice traffic according to destination categories: Local, National, Emergency, International & Cellular. It allows to measure providers operating cost.



**Fig. 1.** Outgoing telephone traffic distribution

This filter was applied to different periods to show the changes caused by the restrictions caused by Covid 19 pandemic. The figure 1 to compare the data between the pre-pandemic period (Nov/Dec 2019) to the pandemic period (Nov/Dec 2020).

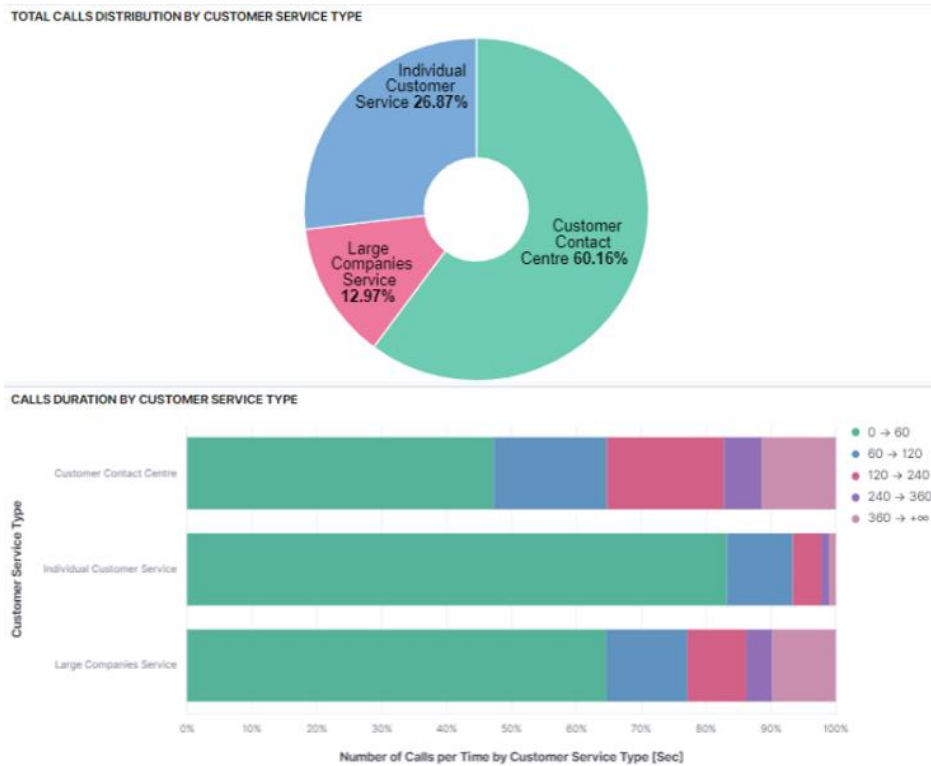


Fig. 2. Call details classified by customer service type

In the figure 2, both graphics were used to build the dashboard that measures the quality of customer service. The top graphic has the distribution of incoming calls classified by customer service type. The service is made up of three groups: customer contact center (CCC), service to individuals and large customers. The bottom graphic details the call duration time for the three customer service type.

## 4 Conclusions

The measurements reflected in the outgoing traffic dashboard reveal the operating costs of the telephone service. Comparing the same months of 2019 (preCovid-19) with 2020 (during Covid-19), an increase of 320% is observed for the Cellular destination category and 50% of reduction in the Local destination category. It can explain by the changes imposed by the restrictions of the pandemic that reduced the attendance of

personnel to branches in exchange for using cell phones affected. It reflects the increase in cost generated.

In the customer service dashboard, it is observed that the calls handled by the CCC represent 60% of the total calls to branches. The remaining traffic is two thirds for the individual sector and one third for large companies. At CCC, 45% of calls last less than a minute. Staff often adjust their behavior to requested productivity levels.

The calls answered by branch officials with a duration greater than five minutes, it is observed that large companies are 10%, in individuals it is only 1%. This is aligned with business objectives

The tasks of linking with the productive environment, although they require technical skills, highlight the need for soft skills for an adequate interaction and contextual interpretation.

### **Competing interests**

The authors have declared that no competing interests exist.

### **Authors' contribution**

"MC conceived the idea and conducted the experiments; MC, AA and AB analyzed the results and revised the manuscript. All authors read and approved the final manuscript."

## **5 References**

- [1] Chen, M.: Use cases and challenges in telecom big data analytics. In: APSIPA Transactions on Signal and Information Processing, vol. 5, p. e19, Cambridge University Press (2016)
- [2] Verma, J., Agrawal, S., Patel, B.: Big Data Analytics: challenges and applications for text, audio, video and social media data. In: International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.5, No.1 (2016)
- [3] Elagib, S., Hashim, A., Olanrewaju, R.: CDR Analysis using Big Data Technology. International Conference on Computing, Networking, Electronics and Embedded Systems Engineering (2015). <https://ieeexplore.ieee.org/document/7381414>, last accessed 2021/01/21.
- [4] Morato, J., Sanchez Cuadrado, S., Fernández, B.: Trends in the technological profile of information professionals. 25(2),169-178 (2016). <https://doi.org/10.3145/epi.2016.mar.03>, last accessed 2021/04/02.
- [5] Dobre, C., Xhafa, F.: Parallel programming paradigms and frameworks in Big Data Era. In: International Journal of Parallel Programming, 42(5), 710–738 (2014). <https://doi.org/10.1007/s10766-013-0272-7>, last accessed 2021/03/28.
- [6] Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U.: Challenges and Opportunities with Big Data. USA, Cyber Center Technical Reports (2011). <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1000&context=cctech>, last accessed 2021/03/17.