

VINCULACION PROFESIONAL PARA EL MONITOREO DE SERVICIOS DE TICS UTILIZANDO BIG DATA

PROFESSIONAL LINKAGE FOR THE MONITORING OF ICT SERVICES USING BIG DATA

Marcelo Dante CAIAFA¹

Ariel AURELIO²

Adrián Marcelo BUSTO³

¹Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de la Matanza, Argentina
mcaiafa@unlam.edu.ar

²Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de la Matanza, Argentina
aaurelio@unlam.edu.ar

³Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de la Matanza, Argentina
abusto@unlam.edu.ar

Resumen:

La articulación entre la formación universitaria y la profesional es uno de los objetivos de nuestra labor educativa. El interés del trabajo es poner de manifiesto las competencias relevantes del estudiante en un proceso concreto de vinculación entre el ambiente académico y el ambiente productivo.

La investigación se basa en el desarrollo de herramientas para el monitoreo de servicios de los sistemas TIC (tecnología de la información y la comunicación). En esta articulación universidad-empresa se utiliza tecnología Big Data para el procesamiento de un gran volumen de datos contenido en los archivos generados por los sistemas de comunicaciones de una organización de servicios masivos que atiende a más de dos millones de clientes.

Se pretende construir un tablero de control basado en tecnología de código abierto ELK (Elasticsearch-Logstash-Kibana), que se abastecerá con los registros de detalle de llamada CDRs (Call Detail Records) generados por la plataforma de comunicaciones.

El objetivo es validar si el dominio de competencias técnicas es condición suficiente o si las habilidades blandas también son necesarias para lograr un eficiente desempeño en un proyecto de estas características.

Abstract:

The articulation between university and professional training is one of many targets from our educational work. The interest of this work is to show the relevant skills of the students in a concrete process of linking between the academic and the productive environment.

The research is based on the development of tools for the monitoring of services based on ICT systems (information and communication technology). In this university-company articulation, Big Data technology is used to process a large volume of data contained in the files generated by the communication systems of a mass service organization that serves more than two million clients.

It is intended to build a control panel based on an open source ELK (Elasticsearch-Logstash-Kibana) technology, which will be supplied with the call detail records CDRs (Call Detail Records) generated by the communications platform.

The objective is to validate if the domain of technical skills is enough condition or if soft skills are also necessary to achieve efficient performance in a project of these characteristics.

Palabras Clave: *Big Data, Tecnología de la Información, Competencias Profesionales del ingeniero del sector TIC, ELK*

Key Words: *Big Data, Information Technology, Professional Competences of the ICT engineer segment, ELK*

Colaboradores: *José Antonio Krajnik, Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de la Matanza, Argentina, jkrajnik@unlam.edu.ar; María de los Milagros Garófalo, Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de la Matanza, Argentina, mgarofalo@alumno.unlam.edu.ar; Christian Ariel Gomez, Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de la Matanza, Argentina, chgomez@alumno.unlam.edu.ar*

I. CONTEXTO

El área de conocimiento pertenece a Ingeniería de las Comunicaciones, siendo el campo de aplicación las Comunicaciones/Telecomunicaciones. Temática: Empleo de herramientas específicas de Big Data en la vinculación profesional para el monitoreo de servicios de las TICs [1].

II. INTRODUCCIÓN

El presente trabajo se basa en la utilización de herramientas de Big Data, específicamente el ELK en el monitoreo de servicios de los sistemas de TICs (tecnologías de la información y la comunicación). En el objetivo que es la articulación universidad-empresa la utilización del modelo de Big Data [7], permite el procesamiento de un gran volumen de datos contenido en los archivos generados por los sistemas de comunicaciones (CDRs o Call Detail Records) de una

organización de servicios masivos que atiende a más de dos millones de clientes. Por otro lado, Big Data analítica es un conjunto de tecnologías y técnicas que requieren nuevas formas de integración para revelar grandes valores ocultos en grandes conjuntos de datos [2]. La articulación universidad-empresa [6], es de vital importancia como vínculo entre la institución formadora de profesionales y la industria, en la cual se aplican los conocimientos adquiridos en la producción de bienes y servicios.

III. MÉTODOS

Como se ha indicado en el ítem II INTRODUCCIÓN, se utilizará el modelo de Big Data y las herramientas para su tratamiento. En éste trabajo el volumen de datos a ser analizado consta de archivos de CDRs (Call Detail Records), siendo la herramienta a ser utilizada el ELK. Para realizar éste análisis se utiliza un servidor en el cual se descargan éstos archivos (CDRs) y la correspondiente herramienta (ELK).

IV. RESULTADOS Y OBJETIVOS

A. Objetivos

La finalidad del proyecto es aportar solución innovadora al sector de las de las tecnologías de la información que representan las comunicaciones unificadas de telefonía, video y colaboración. A partir de trabajo se pretende desarrollar de un tablero de comando que brinde información basado en la explotación de los datos provistos por los registros CDRs.

General: Diseñar e implementar un tablero de comando a partir de indicadores específicos desarrollados sobre infraestructura de Big Data de código abierto que utiliza los registros CDR como fuente de datos. Dicha herramienta pretende aportar valor al análisis de la utilización que los clientes realizan en los sistemas de comunicaciones unificadas sobre los diferentes tipos de servicios.

Específicos: Identificar las tecnologías, herramientas de software y requerimientos de hardware necesarios para la implementación de una plataforma de Big Data de código abierto a partir de cada uno de sus componentes. Realizar la implementación, configuración e integración de las herramientas constitutivas del ambiente seleccionado y su interacción con las fuentes de datos a partir de los CDR para lograr su explotación desde ElasticSearch. Diseñar los diferentes indicadores y componer un tablero de comando visualizable desde Kibana que permita consolidar los indicadores de factores de ocupación, calidad de servicio y niveles de adopción.

B. Resultados

1. Resultados en cuanto a la producción de conocimiento:

Uno de los resultados del presente trabajo será la identificación de las mejores prácticas que validen el valor y la oportunidad que representa la plataforma de Big Data de código abierto conforme a su capacidad para generar información importante para los gestores de los sistemas de telecomunicaciones.

Se estima que lo anterior sea escalable a otras plataformas tecnológicas de forma tal que el servicio pueda ser ofrecido por los alumnos de la universidad en el mercado local.

2. Resultados en cuanto a la formación de recursos humanos:

Dentro de los integrantes del equipo está la participación activa de un estudiante que tendrá su primera incursión en los trabajos de investigación.

Todas las tareas estarán supervisadas por el director del proyecto. El estudiante compartirá su horario con algún miembro del equipo de investigación, quién lo guiará y asesorará para que pueda concretar las tareas encomendadas a lo largo del proyecto.

3. Resultados en cuanto a la difusión de resultados:

Está planificada la asistencia a congresos como CONAISI y CACIC que además de permitir la difusión de los resultados posibilita el intercambio con pares sobre diversos tópicos relacionados con las plataformas tecnológicas utilizadas, los procedimientos seleccionados y las conclusiones alcanzadas.

4. Resultados en cuanto a transferencia hacia las actividades de docencia y extensión:

Actualmente en la currícula de la Universidad existen diversas asignaturas que dan tratamiento a las tecnologías de la información y las comunicaciones. Por ejemplo, en las asignaturas de ingeniería electrónica, Servicios avanzados de comunicaciones, Redes de computadoras y Desarrollo de Mercados Tecnológicos, entre otras. La particularidad que comparten entre ellas es el abordaje sobre distintas temáticas de la gestión de los sistemas de comunicaciones. Tanto desde un enfoque estrictamente técnico como de gestión de recursos para lo cual esta propuesta espera representar un agregado de valor.

5. Resultados en cuanto a la transferencia de resultados a organismos externos a la UNLaM:

Está prevista la transferencia de conocimientos a organismos externos a la UNLAM como el IEEE mediante la publicación de los resultados en su librería digital IEEE Xplore.

6. Vinculación del proyecto con otros grupos de investigación del país y del exterior:

Está prevista la vinculación con un grupo de investigación del exterior compuesto por Sara B. Elagib, Aisha-Hassan A. Hashim ; R. F. Olanrewaju, Department of Electrical & Computer Engineering, Faculty of

Engineering, International Islamic University Malaysia, Kuala Lumpur Malaysia debido a su experiencia en un trabajo referido del año 2016 publicado en el IEEE.

V. DISCUSIÓN

A. Desarrollo del trabajo

El desarrollo del trabajo se encuentra estructurado en cinco etapas. En la primera etapa, el detalle del sistema de comunicaciones se ha llevado a cabo para conocer la fuente de datos. Para la segunda etapa, se obtienen los registros de los CDR y, se prepara el diccionario de datos. La etapa tres negocia con la indexación de la base de datos. Cada campo de los CDRs con información relevante es asignado a un tipo específico de parámetro desde el archivo de texto a un documento JSON. La etapa cuatro corre los procesos ETL. El archivo de texto se construye a partir de que el módulo de Logstash se configure para la ingestión de datos dentro de la correspondiente base de datos. En la última etapa, las búsquedas se llevan a cabo acorde a objetivos específicos que tienen que alcanzarse, con lo cual, el módulo de Kibana, permite a los datos ser consolidados en un panel de control.

B. Contextualización de la fuente de datos.

Para alcanzar los objetivos propuestos se utilizó como fuente de datos registros CDRs, generados por los sistemas de comunicaciones de una organización con más de dos millones de clientes dedicada al servicio masivo [3]. Está compuesta por miles empleados con cientos de sucursales a lo largo del territorio nacional. Dado la extensa normativa que regula sus productos, su estrategia de competencia se focaliza en la diferenciación a partir de la calidad de atención al cliente, lo que potencia el valor del presente trabajo.

Dentro del proceso de vinculación fueron necesarias distintas sesiones de trabajo entre ambos equipos de trabajo, del lado de la universidad y del lado de la empresa. Las mismas se enfocaron en descubrir detalles de la infraestructura y aspectos funcionales de los servicios referidos entre los que se citan:

- la arquitectura funcional de los sistemas que son fuentes de datos,
 - el plan de direccionamiento IP de los segmentos de red LAN, MAN y WAN,
 - el plan de numeración del servicio telefónico y su interconexión con prestadores,
 - la estructura de los diferentes modelos de atención al cliente a estudiar.
- Estas actividades fueron clave para la adecuada contextualización de los datos y su correspondiente interpretación [5].

C. Indexación de la base de datos

Para generar la base de datos es preciso crear un índice en Kibana. Se definen los campos necesarios para dar formato a la estructura de datos que se espera recibir. Esto se ejecuta en el apartado DevTools con una petición PUT. La figura 4 muestra parte del archivo utilizado:

```
PUT /cdr2020Dbv2 {
  "mappings": {
    "properties": {
      "cdrRecordType":{"type":"integer"},
      "globalCallID_callManagerId":{"type":"integer"},
      "globalCallID_callId":{"type":"integer"},
      "origLegCallIdentifier":{"type":"integer"},
      "dateTimeOrigination" :{"type":"integer"},
      /se utiliza tipo entero porque de acuerdo al
      diccionario de datos la fecha y hora están en formato
      UNIX (cantidad de segundos transcurridos desde 01/01/1970)
      "dateTimeOrigination_formatted" :{"type":"date"},
      /se realiza la transformación del campo a partir
      del tipo de archivo .conf del proceso ETL.
      "dateTimeConnect_formatted" :{"type":"date"}
      "origNodeId" :{"type":"integer"},
      "origSpan" :{"type":"integer"},
      "origIdAddr" :{"type":"integer"},
      "origIpv4v6Addr" :{"type":"ip"},
      /se utiliza tipo IP, en lugar del tipo entero ya
      que Kibana permite discriminar rangos de IP.
      "destIpv4v6Addr":{"type":"ip"},
      "callingPartyNumber" :{"type":"integer"},
      "callingPartyNumberPartition" :{"type":"text"},
      /se utiliza tipo texto porque es el formato en que
      se espera recibir el nombre origen de llamada.
      "origNodeId":{"type":"integer"},
      "callingPartyUnicodeLoginUserID":{"type":"text"},
      "origCause_location":{"type":"integer"},
      "dateTimeImport":{"type":"date"},
      "origCause_value":{"type":"integer"},
      "origMediaPrecedenceLevel":{"type":"integer"},
      "origMediaTransportAddress_IP" :{"type":"integer"},
```

Figura 1: Indexación de la base de datos
Fuente: Elaboración propia

D. Proceso ETL

El proceso ETL (Extract, Transform and Load) es el proceso de toma de datos, la adaptación de sus campos y tipos, y la carga de datos a la base. Esto se realiza configurando el logstash con el archivo que se ilustra en la figura 5. Se identifican claramente tres instancias. La primera es la entrada (input), que consiste en indicar la ruta del archivo de donde se extraen los datos. El siguiente paso es el filtro (filter) donde se listan todos los campos que contendrá en formato separado por comas. En el caso particular de las fechas, se pueden convertir los datos que se reciben en formato UNIX a un tipo dato. Por último, se tiene la salida (output) que indica el nombre del índice donde quedarán cargados los datos. El agregado del comando "stdout" habilita la opción para observar la carga en pantalla.

```
input {file {
  path => "C:/Users/unlam/CDR2020.txt"
  start_position => "beginning"
}
filter {csv {columns => ["cdrRecordType",
"globalCallID_callManagerId",
"origLegCallIdentifier",
"dateTimeOrigination", ".....",
"origVideoCap_resolution_Channel2"]
}
date {match => ["dateTimeOrigination", "UNIX"]
target => ["dateTimeOrigination_formatte"]
date {match => ["dateTimeConnect", "UNIX"]
target => ["dateTimeConnect_formatted"]
date {match => ["dateTimeOrigination", "UNIX"]
target => ["dateTimeOrigination_formatte"]
date {match => ["dateTimeConnect", "UNIX"]
target => ["dateTimeConnect_formatted"] }
output {stdout {
elasticsearch {index => "cdr2020DBv2"} }
```

Figura 2: Proceso ETL
Fuente: Elaboración propia

E. Resultados Obtenidos

El parámetro finalCalledPartyNumber se utilizó para clasificar el tráfico de voz según las categorías destino: Locales, Nacionales, Emergencia, Internacionales y Celulares. Esto permite registrar la distribución del tráfico telefónico y el costo operativo del servicio de los proveedores de la PSTN. El filtro se aplicó a diferentes períodos para evidenciar los cambios provocados por las restricciones generadas por la pandemia. La figura 6 compara los datos correspondientes a los períodos prepandemia (Nov/Dic 2019) y período durante la pandemia (Nov/Dic 2020)

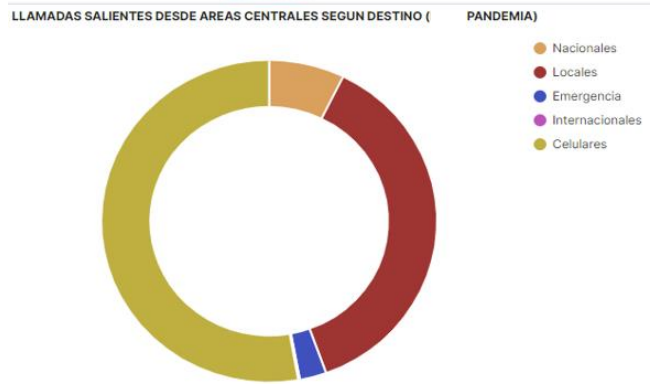
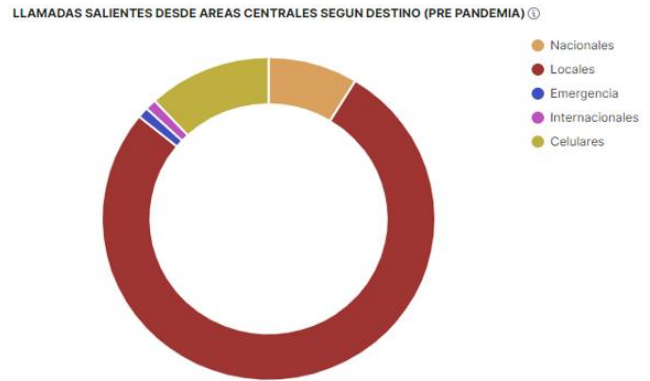


Figura 3: Distribución del tráfico saliente
Fuente: Elaboración propia

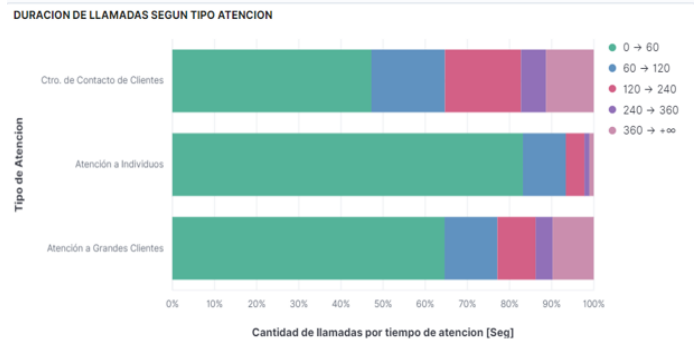
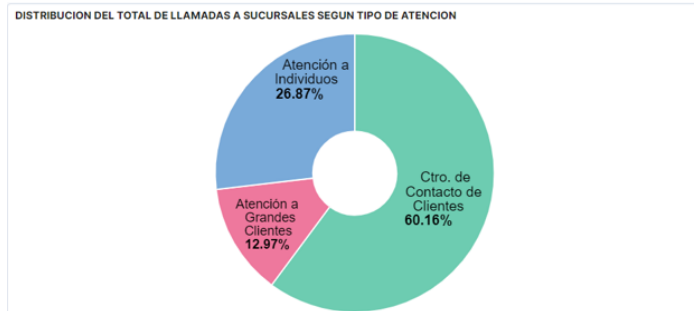


Figura 4: Indicadores de atención al cliente
Fuente: Elaboración propia

Se utilizaron 2 gráficos para construir el tablero que mide el nivel de calidad de atención al cliente y se muestra en la figura 7. El gráfico superior tiene la distribución de llamadas entrantes clasificadas por tipo de atención. La atención se conforma en tres grupos: centro contacto con clientes, atención a individuos y grandes clientes. El gráfico inferior detalla la duración de las llamadas en grupos segmentados por minuto.

Respecto a las habilidades genéricas transversales puestas de manifiesto, fundamentalmente en las etapas 1, 2 y 5 del proyecto se relevaron la capacidad de indagación, pensamiento crítico, organización del tiempo y comunicación efectiva

VI. CONCLUSIONES

El trabajo permitió valorar la experiencia de enseñanza de la ingeniería a través de la participación de estudiantes en un proyecto concretos de articulación academia-empresa, donde la industria aportó contexto productivo y especialistas técnicos y de negocio [4].

Respecto a los datos de tráfico del servicio de telefonía, comparando los mismos meses de 2019 (pre Covid-19) frente al último 2020 (durante Covid-19) se observa un incremento del 320% para la categoría destino Celulares y una reducción del 50% de la categoría destino Locales. Esto refleja la forma en que afectaron los cambios impuestos por las restricciones de la pandemia, que redujo la concurrencia de personal a las sucursales a cambio de utilizar teléfonos celulares.

En el tablero de atención del cliente se observa que los llamados atendidos por el centro de contacto (CC) representan un 60% del total de llamadas a sucursales. El tráfico restante se distribuye en dos tercios para el sector individuos y uno a grandes empresas.

En el CC el 45% de las llamadas duran menos de 1 minuto. El personal del CC adapta su disponibilidad de recursos a niveles de productividad predefinidos.

Comparando las llamadas respondidas por oficiales en sucursal con duración mayor a 5 minutos, se observa que grandes empresas tienen un 10% del total, mientras que individuos es sólo 1%.

En este proceso de vinculación con el ambiente productivo se comprobó que, si bien requirió de competencias técnicas, puso de manifiesto la necesidad de competencias genéricas para las tareas de interpretación de los datos. Se comprobó capacidad de indagación para adecuada interpretación de la fuente de datos y de la arquitectura de servicios.

Desde el punto de vista de los estudiantes, el análisis de esta experiencia permitió identificar como ventaja, la

aplicación de los conocimientos técnicos en ambientes profesionales y el reconocimiento de las habilidades blandas como factores críticos de éxito dentro del proyecto.

Estas últimas resultaron clave particularmente en la etapa de contextualización de datos.

Como futuros trabajos se propone el estudio de metodologías de evaluación de competencias en proyectos de vinculación. Esto constituye un insumo importante para la valoración del desarrollo de habilidades blandas en entornos de integración de plataformas tecnológicas.

VII. REFERENCIAS Y BIBLIOGRAFÍA

- [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] 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.
- [6] S. Mutula. *Big Data Industry: Implication for the library and information sciences*. 2016. *African Journal of Library Archives and Information Science*, 26(2), 93–96. Disponible
<https://www.ajol.info/index.php/ajlais/article/view/167425>
- [7] M. Machado. *Applications of Big Data for Development*. 2016 Disponible en
<https://www.engineeringforchange.org/news/applications-of-big-data-fordevelopment/>

Recibido: 2021-06-06

Aprobado: 2021-07-26

Hervínculo Permanente: <https://reddi.unlam.edu.ar/index.php/ReDDi>

Datos de edición: Vol. 6 - Nro. 1 - Art. 3

Fecha de edición: Formato: 2021-07-27

