

1^{ra} edición

Estudios sobre Economía Circular e Industria 4.0

I Congreso Internacional de Ingeniería Industrial Aplicada
23, 24 y 25 de mayo del 2023, Guayaquil, Ecuador.

Tema

8

Big Data como apoyo para la administración de servicios públicos de la ciudad de Guayaquil, Ecuador.

Ernesto Max Loján Granda

Universidad de Guayaquil; Guayaquil, Ecuador
ernesto.lojang@ug.edu.ec
<https://orcid.org/0000-0002-1194-1522>

María José Trujillo Coloma

Universidad de Guayaquil; Guayaquil, Ecuador
maria.trujilloc@ug.edu.ec
<https://orcid.org/0000-0001-8619-224X>



Resumen

La motivación que ha inspirado este artículo de investigación es orientar la creación de una herramienta simple que sirva para mejorar la calidad de los servicios públicos abastecidos por la Municipalidad de Guayaquil. La red Twitter sirve como productora de denuncias acerca de carencia de servicios públicos, pero los datos que ofrece carecen de estructura definida, por lo que su tratamiento se vuelve complejo. Por otra parte, la tecnología Hadoop MapReduce, es capaz de procesarlos con fiabilidad y velocidad. En el presente artículo, se revisan y analizan tecnologías habilitadoras de la Industria 4.0 tal como el ecosistema Big Data asociado con aplicaciones de geolocalización que muestran mapas web de seguimiento y control, de manera que tanto autoridades como ciudadanos puedan monitorear gráficamente los problemas de la ciudad. Este estudio sugiere que es posible construir una arquitectura tecnológica con herramientas de código abierto para abordar temáticas de gestión de calidad de servicios comunitarios y al mismo tiempo coadyuvar al logro de metas de sostenibilidad ambiental.

Palabras clave: Industria 4.0, Big Data, Twitter, Hadoop, Geolocalización, Economía Circular.

Abstract

The motivation behind this research article is to guide the creation of a simple tool that can improve the quality of public services provided by the Municipality of Guayaquil. Twitter serves as a source of complaints about the lack of public services, but the data it provides lacks a defined structure, making it difficult to process. On the other hand, Hadoop MapReduce technology can process such data reliably and quickly. In this article, Industry 4.0 enabling technologies are reviewed and analyzed, such as the Big Data ecosystem associated with geolocation applications that display web maps for monitoring and control, allowing both authorities and citizens to graphically monitor the city's problems. This study suggests that it is possible to build a technological architecture with open-source tools to address community service quality management issues while also contributing to achieving environmental sustainability goals.

Keywords: Industry 4.0, Big Data, Twitter, Hadoop, Geolocation, Circular Economy.

Introducción

Internet se ha convertido en la principal fuente de datos acerca de temática social de diversa índole, un efecto de ello es la creación de sitios web que ofrecen mapas interactivos que toman información de redes sociales.

En concreto, mapas de seguridad ofrecen información sobre hechos delictivos, por ejemplo, el sitio www.mapadelincuencial.org.mx permite seleccionar ciudades de México y observar geográficamente lugares de peligrosidad delictuencial. En Ecuador, en el año 2009 se lanzó un sitio web de similares características, www.delitosecuador.com, el cual registraba experiencias de personas que sufrieron actos delictivos, los datos se obtienen a través de redes sociales como Twitter y Facebook. Por algún motivo dicho sitio web ya no funciona.

Otros ejemplos interesantes están en el sector turístico, que ofrece mapas en línea de hoteles y planos guía de la ciudad visitada, o el sector gastronómico, que muestra mapas interactivos para elegir restaurantes de acuerdo con costos y sus características. Y es que la tendencia actual, respecto de los sistemas de información, apunta a ofrecer a usuarios de todo tipo de servicio, visión interactiva geográfica en tiempo real, como apoyo para tomar decisiones o simplemente para estar bien informados.

En cuanto al sector del servicio público, específicamente en la ciudad de Guayaquil, no se encontraron sitios web que ofrezcan información interactiva, ni de otro tipo, sobre la problemática del servicio público. Es obvio que los gobiernos municipales se hallan muy interesados en mejorar la calidad de su gestión, consecuentemente, necesitan darle seguimiento gran extensión como Guayaquil, necesitan contar con mecanismos de atención más ágiles.

De hecho, la retroalimentación típica de calidad de servicio usada en las instituciones públicas, son los oficios con peticiones o reclamos que reciben, sin embargo, municipios de ciudades de gran extensión como Guayaquil, necesitan contar con mecanismos de atención más ágiles. La M. I. Municipalidad de Guayaquil no registra estadísticas públicas sobre quejas acerca de servicios municipales tales como: alcantarillado, agua potable, recolección de basura, alumbrado público, ordenamiento urbano, contaminación, ruido, seguridad, entre otros. En la página web del INEC tampoco se observan publicaciones al respecto. Sin embargo, en redes sociales como Twitter o Facebook es habitual observar quejas por problemas en los servicios municipales; en estos sitios los ciudadanos expresan malestar y típicamente adjuntan fotografías como evidencia de dichos problemas.

En efecto, las redes sociales son una importante fuente de información en tiempo real, sobre problemas y calidad de los servicios provistos por el cabildo guayaquileño. Por lo tanto, sería valioso acceder a las denuncias y comentarios que la ciudadanía emita, con relación a dichos servicios, ya que permitiría tener la retroalimentación necesaria para planificar acciones correctivas y atacar los problemas eficazmente, así como mantener registros estadísticos que permitan el seguimiento y planeación de políticas que contribuyan al bienestar de la ciudad.

Ahora bien, hace relativamente poco tiempo, recolectar información de redes sociales analizarla y convertirla en información útil, usando medios de procesamiento convencionales, hubiera sido demasiado complicado como para considerar seriamente una implementación de este tipo. Las bases relacionales tradicionales fueron diseñadas para asimilar datos estructurados, compatibles, y de volumen limitado. Datos originados en sitios como Facebook, Twitter o LinkedIn por citar algunos, no pueden ser llevados ni procesados en una base de datos relacional.

Sin embargo, con el advenimiento de los habilitadores tecnológicos de la Industria 4.0, como la computación en la nube y el desarrollo de tecnologías de análisis de datos masivos no estructurados, se hace posible, a bajo costo, escarbar dentro de estructuras de información no convencionales para obtener soporte estratégico de negocios (Juhasz, 2018), (Alami & ElMaraghy, 2021).

Por otra parte, el presente artículo sugiere que el diseño del sistema de información planteado contribuye a abordar aspectos concernientes a sostenibilidad, una buena parte de denuncias de la comunidad tiene que ver con el desperdicio de recursos vitales, tales como agua o energía eléctrica, incluso temas relacionados con contaminación ambiental. Así, el presente artículo hace una revisión de tecnologías de código abierto para el análisis masivo de datos no estructurados, con el objetivo de orientar el diseño de un sistema de información, usando la red Twitter como fuente de datos, que ofrezca un mapa web de calidad de servicios suministrados por la M.I. Municipalidad de Guayaquil.

Así, esta investigación se plantea las siguientes interrogantes:

1. ¿Es posible diseñar una arquitectura que capture información de redes sociales, de manera simple y rápida, para mejorar la calidad de atención de reclamos de la comunidad?
2. ¿Tal arquitectura, contribuye a la sostenibilidad ambiental?

A continuación, se revisa e identifica en la literatura disponible, definiciones clave que puedan llevar a la construcción una arquitectura Big Data capaz de procesar datos provenientes de la red social Twitter, y desplegarlos visualmente.

Big Data

El concepto Big Data trata el manejo masivo de datos, en el orden de Petabytes o más, semi o no estructurados, distribuidos en la nube de cómputo, a los cuales no es factible aplicar métodos usuales de procesamiento (Bhoola, Kruger, Peick, Pio, & Tshabalala, 2014), (Puyol Moreno, 2014). La figura 29 muestra sus características esenciales.

Figura 29.

Características Big Data.



Nota. Adaptado de *Big Data* de Enrique G., 2014, Gravatar. Biz (https://gravatar.biz/bi/big-data/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+gravatar+%28Business+Intelligence%2C+Data+Warehouse+%3A+Gravatar%29)

Ahora bien, Big Data en sí mismo no produce valor alguno. El análisis de Big Data combina técnicas y tecnologías que permiten transformarlo en información útil (Maltby, 2011). En Maltby (2011), se describen técnicas y tecnologías utilizadas para manejar grandes volúmenes de datos como, por ejemplo, la Minería de Datos, técnica que combina métodos estadísticos con el objetivo de hallar patrones de comportamiento en un conjunto de datos. También se ha desarrollado software para facilitar el análisis de datos a gran escala: Hadoop MapReduce es una tecnología de procesamientos masivo de datos (Almeer, 2012).

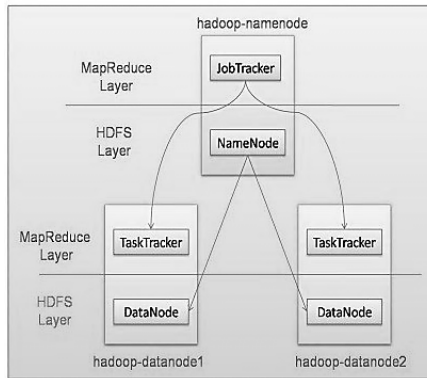
Apache Hadoop MapReduce

Hadoop MapReduce es un proyecto de código abierto, desarrollado por Doug Cutting (Cutting era un empleado de Yahoo!, donde dirigió el proyecto Hadoop a tiempo completo) que se especializa en analizar datos masivos semi o no estructurados, velozmente y en tiempo real.

Hadoop usa un sistema de archivos HDFS (Hadoop distributed file system) similar al usado por Google File System, para el almacenamiento de datos distribuidos, de manera que puedan ser procesados paralelamente por servidores independientes denominados NODOS. El procesamiento en los nodos es administrado por la tecnología MapReduce, la cual tiene la habilidad de dividir el trabajo en tareas concurrentes y juntar posteriormente los resultados y mostrarlos (Loganathan, Sinha, Muthuramakrishnan, & Natarajan, 2014).

Figura 30.

Arquitectura MapReduce.



Nota. Adaptado de *La “Gran División”* de W. Inmon y otros, 2019, Arquitectura de datos (Segunda Edición); Prensa Académica (<https://www.sciencedirect.com/topics/computer-science/hadoop-environment>)

En la figura 30 se puede apreciar que MapReduce usa una arquitectura Maestro-Esclavo, donde Job Tracker (maestro) distribuye Task-trackers (esclavos) a lo largo de nodos habilitados, los cuales almacenan datos usando el sistema HDFS.

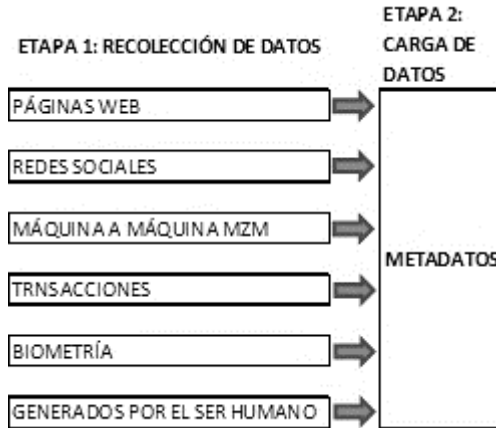
Arquitecturas Big Data

En Camargo Vega, Camargo Ortega, & Joyanes Aguilar (2014) se analizan las arquitecturas Big Data propuestas por Krishnan Krish , Microsoft y Bob Marcus ,obteniendo como resultado una arquitectura que reúne lo mejor de los tres modelos. En la figura 31 se puede apreciar las etapas 1 y 2.

En la etapa 1 la fuente de datos es variada y no estructurada, en la etapa 2 se estructuran los datos en grupos, mediante el uso de bibliotecas semánticas construidas de acuerdo con el problema a resolver.

Figura 31.

Etapas 1 y 2 de arquitectura Big Data.

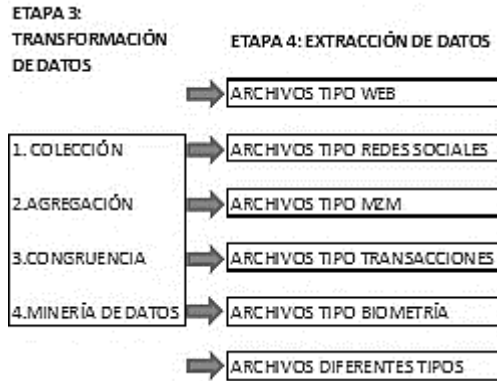


La etapa 3 corresponde a la transformación de los datos. Primeramente, se hace la *Colección* de datos que consiste en la búsqueda dentro de los Metadatos, luego la *Agregación*, que reúne los datos de similar estructura, posteriormente la fase *Congruencia* reúne datos sin importar si son de igual estructura y en último lugar, la fase *Minería De Datos* halla relaciones entre los datos colectados por los pasos anteriores.

Finalmente, La etapa 4, extracción de datos, es la etapa donde los datos se transforman en información útil y visual. La figura 32, muestra las etapas 3 y 4 en conjunto.

Figura 32.

Etapas 3 y 4 de arquitectura Big Data.



Twitter Data Collection

Tal como se ha explicado, Hadoop MapReduce proporciona almacenamiento y procesamiento de datos desde fuentes variadas, a gran escala y procesamiento en paralelo a fin de optimizar los tiempos de respuesta. Sin embargo, dependiendo del propósito que se quiera dar al sistema informático es necesario el uso de componentes adicionales. El ecosistema Hadoop proporciona herramientas que funcionan en la capa más alta de Hadoop, de manera que el usuario no tenga que lidiar con complejos algoritmos de extracción, análisis y visualización de resultados.

En Galvéz-Peréz, y otros (2015), se hace un análisis de los componentes necesarios para producir el “Twitter Data Collection” (TDC), la creación de una gran base de datos de Tweets. El autor (op. cit.) evalúa los distintos APIs de Twitter disponibles para manejo de datos masivos, eligiendo Apache Flume como la mejor opción.

Apache Flume

Apache Flume se define como servicio distribuido, seguro y con alta disponibilidad para una eficiente recolección, unión y movimiento de grandes cantidades de información, no solo de log data, sino también de cantidades masivas de eventos de datos incluyendo tráfico de datos de red y datos generados por social media procedente de fuentes de datos no relacionales (Galvéz-Peréz, y otros, 2015).

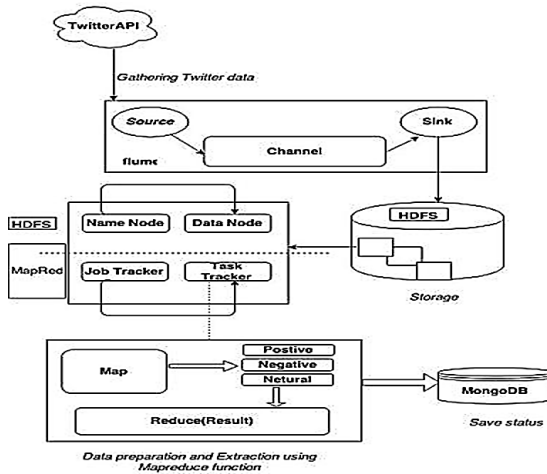
La razón por la que Apache Flume resulta el mejor API para crear el TDC radica en que además de ofrecer seguridad y eficiencia, HDFS retorna datos en formato JSON (JavaScript Object Notation.), el cual es el formato de intercambio de datos que usa Hadoop.

Recolección de datos con Flume

De acuerdo con Mahalakshmi & Suseela (2015), una arquitectura para análisis de comentarios usando Twitter, con el API Flume, se plantea en la figura 33.

Figura 33.

Arquitectura Flume – Hadoop Map Reduce.



Nota. Adaptado de *Big-SoSA: Social Sentiment Analysis and Data*, de Mahalakshmi & Suseela, 201, International Journal of Advanced Research in Computer and Communication Engineering pág. 305

De acuerdo con la figura 5, **Source** es la entidad por donde entran los datos a Flume, hay diversos tipos de **Source** que encajan en determinado tipo de datos. Por ejemplo, FlumeTwitterSource usa la librería Twitter4J para conectarse a **Streaming API** y así adquirir tweets de manera continua. **Channel** es un lugar de espera que conecta el dato con **Sink** de manera que **Source** y **Sink** puedan trabajar coordinadamente. **Sink** transmite cada dato recibido por un único **Channel** hacia el almacenamiento **HDFS**.

HDFS se encarga de segmentar y almacenar bloques de datos. El tamaño de un bloque en *HDFS* es de 64 MB por defecto. *Namenode* administra el árbol del sistema de archivos y los metadatos. *Namenode* tiene la función de Maestro, como ya se vio en la figura 33, MapReduce usa este esquema de trabajo para paralelizar el procesamiento. Los *Datanode* cumplen la función de *Esclavos* almacenando y recuperando bloques de datos e informando a *HDFS*

El proceso MAP crea pares Clave-Valor, de acuerdo con el diccionario de comentarios que se defina, con el objetivo de computarlos, luego el proceso *Reduce* se encarga de organizar los resultados antes de mostrarlos. En cuanto a la base de datos NoSQL, donde finalmente se almacenarán los resultados obtenidos, se pueden tener varias opciones: HBase, Cassandra, MongoDB, Riak, son algunas de ellas. Se sugiere usar HBase ya que se ajusta mejor con los TDC (Gil Pérez, 2014).

Análisis de Datos

Una vez organizada la información, se hace necesario un análisis rápido y eficaz. Entre las opciones para desarrollar consultas están: Hive, Pig o Jaql, componentes del ecosistema Hadoop, de las tres opciones cualquiera es recomendable ya que hacen análisis en corto tiempo, sin embargo, Hive destaca por ser algo más rápida (Rathee, 2013). Tal como se explica más adelante, también se pueden crear scripts sencillos como herramienta de conteo de ocurrencias de palabras, en lugar de crear código NoSQL.

Métodos de clasificación de textos.

En Kaushik & Mishra (2014), se evalúan métodos de clasificación de textos más utilizados: “la máquina de aprendizaje”, emplea algoritmos basados en el método Naive Bayes. Los algoritmos de máquina de aprendizaje entrenan a la computadora acerca de cómo reconocer frases a través de la vinculación probabilística de entradas y salidas textuales.

Otro método, denominado “Basado en Léxico”, asume que la polaridad colectiva de un documento es igual a la suma de polaridades de cada palabra o frase de ese documento. Ambas categorías de clasificación de textos son buenas y exactas, pero la Máquina de aprendizaje demanda mayor tiempo de ejecución de análisis ya que su estructura operativa no está pensada para el manejo de datos masivos (Kaushik & Mishra, 2014).

Por el contrario, el método basado en Léxico construye un diccionario de términos lexicográficos, cuyo procesamiento es más simple, lo cual le permite a este modelo ser más liviano y por ende tener una buena performance, acorde al ambiente Big Data (Galvéz-Peréz, *et al.*, 2015).

Diccionarios Lexicográficos

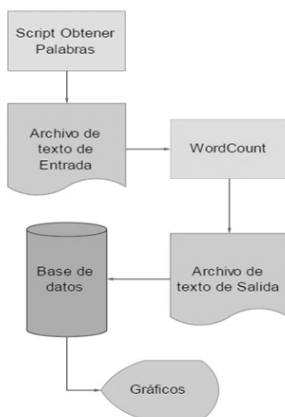
La técnica de diccionarios lexicográficos es una poderosa herramienta usada en la minería de texto, como ejemplo, el diccionario de análisis de sentimientos asigna un puntaje determinado de acuerdo con el tipo de palabra que puede estar entre tres categorías: positivo, negativo o neutral (Ramesh, Divya, Divya, Kurian, & Vishnuprabha, 2015), (Nurfadhlina, 2014). Para el caso de la problemática propuesta en el presente artículo, no se aborda el análisis de sentimientos u opinión, sino la denuncia. Por lo tanto, el diccionario sugerido discrimina frases que tengan relación a servicios municipales. Palabras tales como: alcantarilla, basura, ruido, aguas servidas, agua potable, luminarias, maleza, solar, etc., forman categorías que puedan graficarse en un mapa, como referencia de los problemas de la ciudad.

Tendencias de Palabras

Una vez obtenido el diccionario lexicográfico, es necesario analizar el texto. WordCount es una aplicación sencilla hecha en java que se encarga de analizar texto de una lista y hallar las ocurrencias, similar a una matriz de adiciones (Gallardo, Bermeo, & Cedeño, 2012), el modelo a seguir se define en la figura 34.

Figura 34.

Estructura de análisis de texto.



Nota. Adaptado de *Sistema de reportes y análisis sobre tendencias en la Web de la ESPOL usando Hadoop para el procesamiento masivo de los datos* de Gallardo, Luis; Bermeo, Fabricio y Cedeño, Vanessa (2012, pág. 9)

El script “Obtener Palabras” en este modelo, representa el input provisto por Flume, luego convertido en diccionario lexicográfico (Archivo de texto de entrada) tomado por WordCount.

Creación de Mapas Web IP Geolocalización

Existen muchas herramientas de IP Geolocalización en el mercado, la mayoría de ellas de pago. MaxMind GeoLite, freegeoip.net, hostip.info, ipinfodb.com son algunas de ellas. En cuanto a exactitud, va a ser variable ya que estas tecnologías dependen de la cercanía del proveedor de internet (Martínez, 2013). Todas estas herramientas son de propósito general y no es recomendable usarlas para geolocalización de eventos debido al riesgo de inexactitud (Zander, 2012).

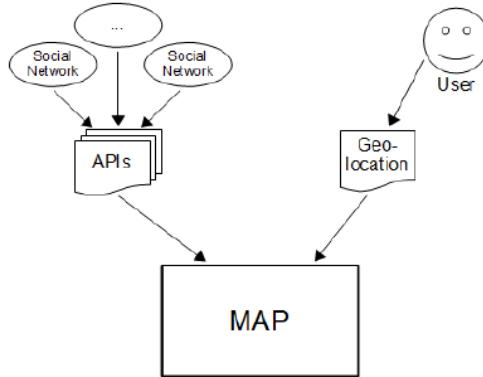
Estructura de despliegue de Tweets

La estructura básica para desplegar Tweets geolocalizados en un mapa muestra en la figura 35, se pueden identificar tres componentes principales: Redes Sociales, Algoritmo de geolocalización y el mapa donde se despliegan los datos. A fin de obtener la geolocalización se pueden seguir varias alternativas (Martínez, 2013):

- Vía GPS (sistema de posicionamiento global), lo que requiere tener un módulo GPS en cada móvil.
- Triangulación, lo cual demanda contar con dispositivos inalámbricos de acceso en los alrededores.
- Detección a través del protocolo IP, aunque este método no siempre garantice exactitud debido principalmente al uso frecuentes de VPN (Virtual Private Network).
- Ingreso, vía formulario, de las coordenadas exactas de ubicación de los usuarios.

Figura 35.

Estructura básica Geolocalización.



Nota. Adaptado de *LT map A web-based application for the display* de Martínez, Aron, (2013, pág. 4)

Las APIs obtienen geolocalización eligiendo la alternativa de acuerdo con el orden mostrado en el párrafo anterior, por ejemplo, si no se dispone de GPS, se intenta usar Triangulación.

Tweography

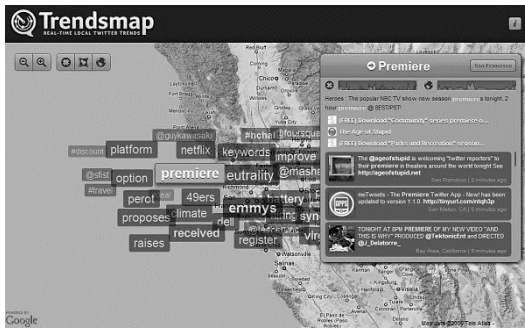
Tweography es una interesante herramienta de geolocalización de tweets en un mapa, que muestra la posición de una persona el instante de emitir el Tweet, usa autenticación como medio de validación de identidad. Sin embargo, sufre una alta tasa de interrupciones, lo cual no le hace fiable, además de no tener documentación suficiente de soporte (Martínez, 2013).

Trendsmap

Trendsmap es una herramienta que permite conocer los sucesos ocurridos en tiempo real, hashtag más populares y tendencias en Twitter, dispone de un módulo de Geo ubicación basado en Google, que hace posible conocer la ubicación exacta desde donde se emiten los tweets (Shavitt & Zilberman, 2010; Weidemann, 2014). Trendsmap es similar a Google Maps, de uso sencillo e intuitivo, en la figura 36 se muestra una captura de pantalla de una aplicación que usa Trendsmap.

Figura 36.

Captura de pantalla de TrendsMap en acción.



Nota. Adaptado de *Trendsmap: tendencias de twitter en tiempo real* en Studio-cassette, 2013 (<http://www.studiocassette.com/trendsmap-les-tendances-twitter-en-temps-reel/>)

Trendsmap hace posible la creación de modelos simples asociados a Twitter (Zhang & Gelernter, 2014), un ejemplo de la estructura de datos procesada por TrendsMap se muestra en la tabla 11. Tal como se puede apreciar, su estructura es simple e intuitiva

Tabla 11.

Formato de campos de metadatos Twitter.

Nombre de Campo	Fuente de información	Ejemplo
Texto Tweet	Tweet enviado por usuario el momento de publicarlo.	"Basura en 10 de Agosto y Quito!"
Coordenadas	Posición del dispositivo del usuario que envía el tweet	37.59628575, -82.24814532
Localización Usuario	Información de la cuenta del usuario	Lopez, Juan
Time Zone	Time Zone ingresada por el usuario durante la apertura de	Bogotá, Lima, Quito
Descripción	Descripción ingresada por el usuario	Ingeniero Espol, casado

Integración con ODBC

Hadoop no solamente se circunscribe dentro del ambiente Big Data. Es posible, con el uso de drivers especiales, una integración con Bases de Datos Relacionales (RDBMS). Dentro del ecosistema de Hadoop, el módulo Hive desarrollado para relacionarse con HDFS como un Data Warehouse, provee mecanismos para facilitar la consulta de datos. Hive está situado en la capa superior de la arquitectura Hadoop MapReduce lo cual simplifica la comunicación con HDFS. Hive usa un lenguaje de consulta denominado HQL de estructura sintáctica similar a SQL, lo cual es una ventaja para quienes ya están familiarizados con lenguaje de consulta RDBMS convencionales.

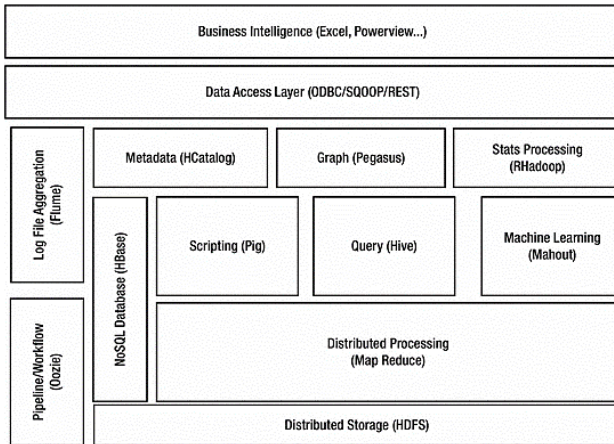
Asimismo, Hive proporciona un controlador ODBC configurable en el entorno Microsoft. En Sarkar (2014), se propone una arquitectura Hadoop MapReduce que interactúa con herramientas del tipo Office. En la figura 37 se describe la arquitectura propuesta, donde puede observarse que la capa de acceso contiene el Driver ODBC que conecta con la capa de aplicación.

De este modo, los datos de Twitter ya procesados pueden ser importados a Excel y desde ahí pueden construirse gráficos e incluso mapas, con mucha facilidad y versatilidad (PowerView, disponible en Excel versión 2013).

Hay que resaltar el hecho de que los gráficos y mapas producidos en Excel u otra herramienta tipo ofimática, son aprovechables internamente, si el requerimiento apunta a mapas interactivos publicados en la web, Trendsmap es la mejor opción.

Figura 37.

Arquitectura Hadoop – ODBC.



Nota. Adaptado de *Pro Microsoft HDInsight* de Sarkar, Debarchan , Apress (2014, pág. 128).

Materiales y métodos

Se realizó una revisión bibliográfica para obtener una visión general de los casos de uso de la tecnología Big Data y herramientas de código abierto en el ámbito de manejo masivo de datos obtenidos en redes sociales. Se consultaron varias bases de datos y se identificaron 30 artículos relevantes que se utilizaron para un análisis detallado siguiendo el método analítico-sintético, la muestra es no probabilística – intencional. Producto de este análisis se construyó una arquitectura tecnológica para el desarrollo de un sistema de información capaz de procesar datos masivos no estructurados y llevarlos a mapas interactivos y geolocalización.

Resultados y discusión

En este trabajo se ha presentado una arquitectura de tecnología Big Data que de manera simple permite interactuar con usuarios que hacen reclamos por defectos o mala calidad de servicios comunitarios provistos por el municipio de una ciudad. Así, a través de mapas digitales intuitivos se asocian comentarios realizados en la red Twitter y su geolocalización. Esta herramienta sería de gran utilidad para el mejoramiento de la gestión municipal, así, la respuesta a la primera cuestión planteada en este artículo queda resuelta.

Ahora bien, con respecto a la segunda cuestión planteada, de acuerdo con Korhonen, Nuur, Feldmann, & Eshetu Birkie (2017), definen la adopción de economía circular EC como contribuyente a las tres dimensiones del desarrollo sostenible: reutilizar, reciclar y reducir, y limita el flujo de materiales y energía a un nivel que la naturaleza tolera, extrayendo más valor económico de los flujos y infraestructuras existentes de la economía (Korhonen, Honkasaalo, & Seppälä, 2018). Fundamentalmente, los gobiernos locales están comprometido con tales objetivos, así, los residuos y emisiones son los causantes de la carga ambiental que deben soportar las ciudades, producto de su expansión y desarrollo inherente (Sarkis & Zhu, 2008), (Tong, Wang, Chen, & Wang, 2018). Consiguientemente, las tecnologías de la información deben apoyar el desarrollo sostenible, la innovación tecnológica desempeña un papel central en la transformación hacia la economía circular (De Jesus & Mendonça, 2018; Tura, y otros, 2019).

Así también, para la EC es importante crear valor empresarial, además de la gestión ambiental, la optimización de costos tiene un rol determinante (Park, Sarkis, & Wu, 2010; Geng, Sarkis, & Ulgiati, 2016), por ejemplo, los municipios deben dotar de agua potable apta para el consumo humano o una red de ilu-

minación comunitaria de calidad, pero no solo eso, también debe cuidar que recursos como agua o luz eléctrica no se desperdicien. En tal sentido, es imposible para el municipio gestionar el control de recursos con tal granularidad, por el contrario, un sistema de información que reciba datos de ciudadanos responsables que denuncien fugas de agua, focos de contaminación, daños o falta de mantenimiento en infraestructuras, etc., facilita la administración de la ciudad ya que la dispersión escalar de flujo de recursos (Tong, Wang, Chen, & Wang, 2018; Park, Sarkis, & Wu, 2010), quedaría resuelta mediante su geolocalización.

De esta manera, se aborda la segunda interrogante del presente estudio investigativo, indudablemente, la arquitectura tecnológica propuesta, contribuye en la transición hacia la sostenibilidad ambiental, al mejor costo posible, ya que el uso herramientas de código abierto, combinado con un diseño simple y efectivo aporta nuevas ideas para el desarrollo de aplicaciones tecnológicas dirigidas a servicios comunitarios. Sin embargo, existen desafíos latentes, los cuales no dependen exclusivamente de la tecnología, aspectos regulatorios, de confianza, culturales, descentralización y otros, deben superarse para lograr la plena adopción de la EC (Geng, Sarkis, & Ulgiati, 2016).

Conclusión

En el presente artículo, se ha hecho una revisión de técnicas y tecnologías disponibles para llevar a cabo la implementación de un sistema de información, basado en el framework Hadoop Map Reduce, que permita el monitoreo de quejas sobre servicios ejecutados por la municipalidad de Guayaquil.

Por lo tanto, el correcto suministro de servicios públicos es un asunto de vital importancia para los gobiernos seccionales ya que la calificación de su gestión y popularidad depende de ello, de este modo, los municipios necesitan conocer las necesidades no resueltas de la ciudad. Por otro lado, es imperativo llevar estadísticas de problemas de manera que puedan adoptar estrategias de mejoramiento.

Es así como la red social Twitter, constituye una oportunidad para que los gobiernos municipales de grandes urbes puedan monitorear la calidad de servicios que entregan. Sin embargo, la tarea de revisar cientos de miles de mensajes no es sencilla, más aún, si dichos mensajes no tienen estructura fija, así, se sugiere para este efecto, la tecnología Hadoop MapReduce, de código abierto y desarrollado a partir del Big Data, asociada con otros componentes, permiten analizar y compendiar datos tomados de la red Twitter.

De la misma manera, se debe evaluar coherente las opciones disponibles antes de comenzar a trabajar en la implementación. Para esto, el criterio a seguir tiene dos pilares fundamentales: primeramente, el tiempo de respuesta debe ser al menos aceptable y; segundo, la implementación debe ser simple de realizar. De nada serviría una aplicación muy sofisticada que no entregue resultados, así como tampoco una aplicación costosa en términos de mantenimiento y actualización.

Finalmente, la implementación de proyectos tecnológicos innovadores tiene el potencial de contribuir significativamente a la sostenibilidad ambiental. En este sentido, la arquitectura propuesta en este trabajo investigativo puede ser una herramienta valiosa para lograr este objetivo. Además, este enfoque está en línea con las dimensiones de sostenibilidad propuestas por la Comisión Europea, que abogan por la promoción de soluciones tecnológicas innovadoras y sostenibles. Al aprovechar esta oportunidad y adoptar tecnologías innovadoras, se puede avanzar hacia una economía más circular y reducir el impacto ambiental. En resumen, la implementación de la arquitectura propuesta puede contribuir de manera significativa al logro de objetivos de sostenibilidad y al mismo tiempo permitir innovación tecnológica en línea con las políticas de EC.

Como trabajo futuro se propone integrar a las redes sociales Facebook y LinkedIn en este proyecto, incorporando un módulo que ilustre sentimientos sobre la gestión y administración municipal. El uso de mapas interactivos se alinea con la tendencia de mostrar gráficamente la realidad, lo cual es una poderosa herramienta de toma de decisiones, además, al proyecto actual habría que agregarle la vídeo y fotografía o realidad aumentada.

Referencias

- Alami, D., & ElMaraghy, W. (2021). A cost benefit analysis for industry 4.0 in a job shop environment using a mixed integer linear programming model. *Journal of Manufacturing Systems*, 81-97.
- Almeer, M. (abril de 2012). Cloud Hadoop Map Reduce For Remote Sensing Image Analysis. *Journal of Emerging Trends in Computing and Information Sciences*, 3(4), 637-644.
- Bhoola, K., Kruger, K., Peick, J., Pio, P., & Tshabalala, N. (2014). Big Data analytics. *Actuarial Society of South Africa's 2014 Convention* (págs. 107-143). Johannesburgo: Actuarial Society 2014 Convention.

- Camargo Vega, J., Camargo Ortega, J., & Joyanes Aguilar, L. (2014). Arquitectura Tecnológica Para Big Data. *21*. Revista Científica.
- De Jesus, A., & Mendonça, S. (2018). Lost in Transition? Drivers and Barriers in the Eco-innovation Road to the. *Ecological Economics*, 75-89.
- Gallardo, L., Bermeo, F., & Cedeño, V. (2012). Sistema de reportes y análisis sobre tendencias en la Web de la ESPOL usando Hadoop para el procesamiento masivo de los datos. Guyaquil.
- Galvéz-Peréz, J., Gómez-Torrero, B., Ramírez-Chávez, R., Sánchez-Sandoval, K., Castellanos-Cerda, V., García-Madrid, R., Villatoro-Tello, E. (Enero de 2015). Sistema automático para la clasificación de la opinión pública generada en Twitter. *Research in Computing Science* *95*, 1(1), 23-36.
- Geng, Y., Sarkis, J., & Ulgiati, S. (2016). Sustainability, well-being, and the circular economy in China and worldwide. *ResearchGate*, 76-79.
- Gil Pérez, B. (2014). TDC (Twitter Data Collection): Creación de una gran base de datos de Tweets.
- Juhasz, L. (2018). Overview of industry 4.0 tools for. *ResearchGate*, 51-71.
- Kaushik, C., & Mishra, A. (septiembre de 2014). A Scalable, lexicon based technique for sentiment analysis. *International Journal in Foundations of Computer Science & Technology*, 4(5), 35-43.
- Korhonen, J., Honkasalo, A., & Seppälä, J. (2018). Circular Economy: The Concept and its Limitations. *Ecological Economics*, 37-46.
- Korhonen, J., Nuur, C., Feldmann, A., & Eshetu Birkie, S. (2017). Circular economy as an essentially contested concept. *Journal of Cleaner Production*, 544 - 551.
- Loganathan, A., Sinha, A., Muthuramakrishnan, V., & Natarajan, S. (julio de 2014). A Systematic Approach to Big Data. *International Journal of Information & Computation Technology*, 4(9), 869-878.
- Mahalakshmi, R., & Suseela, S. (abril de 2015). Big-SoSA: Social Sentiment Analysis and Data. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(4), 304-306.
- Maltby, D. (octubre de 2011). Big Data Analytics. *ASIST* *2011*, 9(13), 1-6.
- Martinez, A. (2013). LT map A web-based application for the display.

- Nurfadhlina , M. (agosto de 2014). A Review of Sentiment Analysis Approaches in Big Data Era. *University of Putra Malaysia*, 13(1), 7-12.
- Park, J., Sarkis, J., & Wu, Z. (2010). Creating Integrated Business and Environmental Value within the Context of China's Circular Economy and Ecological Modernization. *Journal of Cleaner Production*, 1494-1501.
- Puyol Moreno, J. (julio de 2014). Una aproximación a Big Data. *Revista de Derecho UNED*, 1(14), 471-505.
- Ramesh , R., Divya , G., Divya , D., Kurian, M., & Vishnuprabha , V. (abril de 2015). Big Data Sentiment Analysis using Hadoop. *International Journal for Innovative Research in Science & Technology*, 1(11), 92-96.
- Rathee, S. (Noviembre de 2013). Big Data and Hadoop with components like Flume, Pig, Hive and Jaql. *International Conference on Cloud, Big Data and Trust 2013*, 13(15), 78-82.
- Sarkar, D. (2014). *Pro Microsoft HDInsight*. Bangalore, India, Bangalore, India: Apress.
- Sarkis, J., & Zhu, H. (2008). Information technology and systems in China's circular economy: Implications for Sustainability. *Journal of Systems and Information Technology* , 1-41.
- Shavitt, Y., & Zilberman, N. (julio de 2010). A Study of Geolocation Databases. *IEEE Explore*, 29(10), 1-14.
- Tong, X., Wang, T., Chen, Y., & Wang, Y. (2018). Towards an inclusive circular economy: Quantifying the spatial flows of e-waste through the informal sector in China. *Resources, Conservation and Recycling*, 163-171.
- Tura, N., Hanski, J., Ahola, T., Ståhle, M., Piiparinen, S., & Valkokari, P. (2019). Unlocking circular business: A framework of barriers and drivers. *Journal of Cleaner Production*, 90-98.
- Weidemann, C. (2014). Geosocialfootprint(2013):social media location privacy web map. California.
- Zander, S. (mayo de 2012). How Accurate is IP Geolocation Based on IP. *Centre for Advanced Internet Architectures, Technical Report*, 2-5.
- Zhang , W., & Gelernter, J. (junio de 2014). Geocoding location expressions in Twitter messages: A preference learning method. *JOURNAL OF SPATIAL INFORMATION SCIENCE*, 9(9), 37-70.