

# Integration of tools for decision making in vehicular congestion

Nelson Iván-Herrera-Herrera<sup>a</sup>, Sergio Luján-Mora<sup>b</sup> & Estevan Ricardo Gómez-Torres<sup>a</sup>

<sup>a</sup> *Facultad de Ciencias de la Ingeniería e Industrias, Universidad Tecnológica Equinoccial, Quito, Ecuador. nelson.herrera@ute.edu.ec, etevan.gomez@ute.edu.ec*

<sup>b</sup> *Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Alicante, España. sergio.lujan@ua.es*

Received: September 15<sup>th</sup>, 2017. Received in revised form: March 15<sup>th</sup>, 2018. Accepted: March 21<sup>th</sup>, 2018.

## Abstract

The purpose of this study is to present an analysis of the use and integration of technological tools that help decision making in situations of vehicular congestion. The city of Quito-Ecuador is considered as a case study for the done work. The research is presented according to the development of an application, using Big Data tools (Apache Flume, Apache Hadoop, Apache Pig), favoring the processing of a lot of information that is required to collect, store and process. One of the innovative aspects of the application is the use of Twitter social network as source of origin. For this, it used its application programming interface (API), which allows to take data from this social network and identify probable points of congestion. This study presents results of tests carried out with the application, in a period of 9 months.

*Keywords:* congestion; traffic; Twitter; big data; application; Quito.

# Integración de herramientas para la toma de decisiones en la congestión vehicular

## Resumen

Este estudio tiene como finalidad presentar un análisis de la utilización e integración de herramientas tecnológicas que ayudan a tomar decisiones en situaciones de congestión vehicular. La ciudad de Quito-Ecuador es considerada como un caso de estudio para el trabajo realizado. La investigación se presenta en función del desarrollo de una aplicación, haciendo uso de herramientas *Big Data* (Apache Flume, Apache Hadoop, Apache Pig), que permiten el procesamiento de gran cantidad de información que se requiere recolectar, almacenar y procesar. Uno de los aspectos innovadores de la aplicación es el uso de la red social Twitter como fuente de origen de datos. Para esto se utilizó su interfaz de programación de aplicaciones (*Application Programming Interface*, API), la cual permite tomar datos de esta red social en tiempo real e identificar puntos probables de congestión. Este estudio presenta resultados de pruebas realizadas con la aplicación, durante un período de 9 meses.

*Palabras clave:* congestión; tráfico; Twitter; big data; aplicación; Quito.

## 1. Introducción

La dinámica de las grandes ciudades genera situaciones complejas y difíciles de abordar. Entre estas situaciones se tiene la alta densidad poblacional en espacios geográficos muy estrechos y sin posibilidades de crecimiento urbano.

La ciudad de Quito es la capital del Ecuador, siendo esta el centro económico del país, donde se desarrollan actividades comerciales, administrativas y turísticas. Debido a su gran expansión poblacional, la ciudad ha experimentado

grandes cambios, que ha tenido como consecuencia del problema de congestión vehicular [1].

Entre las principales causas de la congestión vehicular<sup>1</sup> están las características topográficas de la ciudad, ya que cuenta con grandes montañas al lado occidental y desniveles y valles al lado oriental. Esta situación ha generado, que la ciudad se expanda en dirección norte – sur, es decir de forma alargada y con muy poca organización y planificación [2], sumadas a un explosivo crecimiento del parque automotor, han agudizado aún más el problema [2].

**How to cite:** Herrera-Herrera, N.I., Luján-Mora, S. and Gómez-Torres, E.R., Integración de herramientas para la toma de decisiones en la congestión vehicular DYNA, 85(205), pp. 363-370, June, 2018.

<sup>1</sup> Relación entre la capacidad de la Infraestructura Vial (Oferta) y el tráfico que circula por ella (Demanda).

A través del tiempo han existido importantes iniciativas de movilidad en la ciudad, incluyéndose sistemas de control de tránsito, monitoreo, construcción de vías perimetrales, e inclusive la medida “pico y placa” para restringir el uso vehicular en horas pico, de acuerdo a la placa del vehículo [3]. Cabe mencionar que esta alternativa ha sido una iniciativa que se ha adoptado en varios países de América Latina. Sin embargo, todas estas iniciativas no han sido suficientes y la congestión aumenta cada vez más, generando malestar en la ciudadanía.

Con base a esta problemática, se necesita una aplicación que permita: recolectar, almacenar, procesar y visualizar información no estructurada en tiempo real; lo cual sirve de soporte a la toma de decisiones, en situaciones de congestión vehicular, tanto a nivel de entidades oficiales como de personas.

Esta aplicación hace uso de dos grupos principales de datos, los datos sensoriales y sociales. Para el primer tipo de datos, se considera los recopilados a través de sensores, que miden el flujo vehicular en sectores estratégicos, colocados a lo largo de la ciudad. Para lo cual un agente de tránsito por experiencia, determina que puntos son propensos a generar congestión y por lo tanto donde se los deben colocar. El otro grupo de datos hace uso de información pública, existente a través de las redes sociales, en este caso, Twitter.

La información disponible de sensores del volumen de tráfico, así como de las redes sociales, se procesará mediante una aplicación creada para ello. Posteriormente se presentarán los resultados de este análisis, que servirá de apoyo a la toma de decisiones, en situaciones de congestión vehicular.

En un trabajo previo [4], se realizó un análisis de los factores que influyen en la congestión vehicular en Quito; además del estudio de la cuestión acerca de herramientas tecnológicas, que han ido utilizadas alrededor del mundo, como una alternativa de solución a este problema.

El presente artículo está orientado a la presentación técnica de la aplicación y sus componentes; sin embargo el estudio realizado incluye todos los elementos formales de una investigación. En este sentido, partiendo desde el planteamiento del problema, hasta las conclusiones y recomendaciones, se efectuó la fundamentación teórica respaldada en el análisis del uso de técnicas y estrategias utilizadas mundialmente relacionadas al tráfico vehicular.

Los aspectos metodológicos del presente trabajo se orientaron a una investigación de campo para conocer, a través de la aplicación de instrumentos, la percepción de la población sobre la congestión y el uso de redes sociales.

Los hallazgos de la investigación de campo proporcionaron información mediante la cual se logró fundamentar de mejor manera la presente propuesta que a continuación se detalla.

El presente trabajo tiene el contexto de una investigación Big Data [5], pues los elementos de esta tecnología se conjugan perfectamente:

**Volumen:** miles de datos serán generados mediante la red social Twitter y la proporcionada por los sensores instalados en la ciudad.

**Variedad:** datos generados por sensores y la red social Twitter.

**Velocidad:** diariamente se generarán datos mediante la red social Twitter a una importante velocidad. Esto requiere que su procesamiento y posterior análisis, normalmente, ha de hacerse en tiempo real, para mejorar la toma de decisiones [6].

**Veracidad:** la información de la red social Twitter, debe ser filtrada para seleccionar, únicamente, la que sirva para el estudio.

## 2. Trabajos relacionados

Se han realizado varias investigaciones relacionadas con herramientas tecnológicas utilizadas en situaciones de congestión vehicular.

- Utilización de modelos macroscópicos [7], equipando a los vehículos con dispositivos de detección cada vez más sofisticados, como cámaras. Las personas y los vehículos están compartiendo datos de detección para mejorar la experiencia de conducción [8].
- Sistemas que monitorean, procesan y almacenan grandes cantidades de datos, lo que permite detectar la congestión del tráfico de manera precisa. Para esto se utiliza una serie de algoritmos que reducen la emisión localizada de vehículos, mediante el re-encaminamiento de los automóviles [9].
- Utilización de sistemas de detección participativa, como Foursquare e Instagram, que se están volviendo muy populares. Los datos compartidos en estos sistemas tienen la participación activa de los usuarios mediante la utilización de dispositivos portátiles. Estos sistemas pueden ser vistos como una especie de sensores que, junto a información de condiciones de tráfico, constituyen en eficientes predictores de congestión vehicular [10].

En este ámbito, otra de las redes sociales potentes que permiten realizar análisis en tiempo real, de los mensajes permitidos y poder determinar puntos de congestión vehicular, es la red social Twitter [11]. Otro insumo importante al momento de analizar y detectar situaciones de tráfico es la utilización de sensores inalámbricos. Estos dispositivos están ganando más atención en la detección de tráfico [12], por lo cual, adicional a Twitter, se utilizará también información de sensores de conteo vehicular. Esto permitirá tener dos criterios previos que, finalmente mediante contraste, ayudará a entregar información más precisa de puntos de congestión vehicular en la ciudad de Quito.

## 3. Herramientas utilizadas

### 3.1. Información de sensores

Para lograr la medición del tráfico en la ciudad, el Municipio de Quito, a través de la Secretaría de Movilidad, dispone de sensores a lo largo de la localidad.

La Fig. 1 muestra un ejemplo de sensores de tráfico que están ubicados en lugares estratégicos donde existe más afluencia vehicular, de acuerdo a los criterios del Municipio de Quito. Estos sensores recolectan información de la cantidad de vehículos que circulan por cada punto.



Figura 1. Sensores en la ciudad.  
Fuente: Tomado de [2].

Adicionalmente, el Municipio de Quito dispone de parámetros que corresponden a valores preestablecidos de acuerdo al sector, donde se establece una cantidad límite máxima. Si el número de vehículos sobrepasa este valor, esto implica que el lugar se encuentra congestionado.

La aplicación que se ha desarrollado y se presenta, genera información útil que establece posibles puntos de congestión vehicular; mediante el análisis de datos generados por sensores y; la correlación con la información presentada en base a tuits recolectados en un determinado momento. La información se presenta en un mapa elaborado para la investigación, que utiliza el API de Google Maps; los puntos de congestión que puedan ser analizados, con lo cual se establecen sectores de mayor tráfico y, a su vez, horas pico. Finalmente, mediante estos resultados se pueda tomar decisiones de cambio de ruta y comunicarlas a la ciudadanía, a efectos de aliviar la congestión vehicular existente.

### 3.2. Datos de Twitter

La segunda fuente de datos considerada para el análisis corresponde a los datos que proporcionan las personas al hacer uso de la red social Twitter. Se requiere tener una cuenta en Twitter, además de ser necesaria la activación del servicio de GPS en el teléfono inteligente o tableta. Esta red social genera información variada y, por tanto, es utilizada con diversos propósitos, uno de ellos es estimar el tráfico existente en la ciudad. Por lo que los tuits que se generan a todas horas tienen una geolocalización, que permite identificar el origen del mismo. Para lo cual se hace uso del API de Twitter que se detalla a continuación.

### 3.3. Uso del API de Twitter

Para poder utilizar el API de Twitter es necesario tener una cuenta en esta red social y crear otra adicional de desarrollador, la versión utilizada es la 1.0 [13]. La aplicación

```
"geo"; "coordinates" y
"place"; "name"; "full_name"; "country_code"; "country"
```

Figura 2. Campos de geolocalización de un Tuit  
Fuente. Los autores

se ha habilitado para la lectura de tuits; sin embargo, la misma permite la publicación de los mismos en caso de ser requerido, para lo cual se habilita la escritura en las opciones de la misma.

Se debe considerar que Twitter debe estar configurado en el teléfono inteligente o tableta, además del GPS. El cual en la mayor parte de casos está habilitado, por defecto.

### 3.4. Creación del hashtag

Con la finalidad de recolectar y filtrar información relacionada a la congestión generada por el tráfico vehicular en el Distrito Metropolitano de Quito, se creó la etiqueta<sup>2</sup> #TraficoUIO. La utilización de esta etiqueta fue socializada a toda la comunidad de la Universidad Tecnológica Equinoccial (UTE). Inicialmente se recolectaron todos los

```
"geo": { "type": "Point",
"coordinates": [38.9069,
-6.33876]
},
"coordinates": {
"type": "Point",
"coordinates": [-6.33876,
38.9069]
},
"place": {
"id": "6d564e12cf533f7a",
"url":
"https://api.twitter.com/1.1/geo/id/6d564e12cf533f7a.json",
"place_type": "city",
"name": "Quito",
"full_name": "Quito, Pichincha",
"country_code": "EC",
"country": "Ecuador",
"contained_within": [],
"bounding_box": {
"type": "Polygon",
"coordinates": [[[-6.6568728,
38.7268887],
[-6.1404139,
38.7268887],
[-6.1404139,
39.1814719],
[-6.6568728,
39.1814719]]]
}
}
```

Figura 3. Valores asociados a un tuit  
Fuente. Los autores

<sup>2</sup> La etiqueta o hashtag tiene como objetivo agrupar tuits que hablen de un mismo tema y facilitar la búsqueda de información

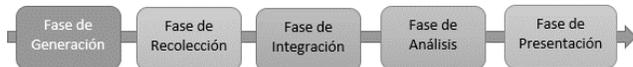


Figura 4. Flujo de información del sistema.  
Fuente: Los autores.

tuits asociados al hashtag antes mencionado. La Fig. 2 muestra los campos que permiten a la aplicación acceder a los datos de geolocalización del tuit.

La configuración del tuit considera la ubicación desde la que se genera el mismo, con lo cual es posible obtener la localización del tráfico para los tuits que denuncien esta situación.

La Fig. 3 detalla algunos valores asociados a un tuit, que permiten acceder a datos que muestran, por ejemplo: la ciudad, provincia, país, desde donde se generó el tuit.

La sección “coordinates” indica la geolocalización del tuit, que sirve para filtrar los tuits generados considerando solamente los originados en las vías, donde se ha identificado existe mayor congestión vehicular. La sección “place type” indica el tipo de lugar en el cual se están generando los tuits; el tipo indicado es una ciudad. Otra sección importante, que ayuda a delimitar los tuits generados es “country”, que indica la ciudad en la que se generan los tuits; para el presente estudio la ciudad descrita es Quito. Finalmente, la sección “name” permite identificar el lugar desde el cual se genera el tuit; el sitio de interés en el presente caso es la ciudad de Quito.

## 4. Metodología

La revisión documental de casos de ciudades con la misma problemática permitió analizar y evaluar las aplicaciones utilizadas a nivel mundial en relación a la movilidad inteligente.

En este contexto, los casos de estudio analizados y detallados en este estudio sirvieron para respaldar tanto en contenido o fundamento teórico como metodológico la presente propuesta.

En la Fig. 4, se muestra la arquitectura e integración de las fases que utilizará la aplicación de apoyo a la toma de decisiones en la congestión vehicular.

### 4.1. Fase de generación

Esta fase es contemplada como una entrada externa importante para la aplicación, pues aquí se generan los datos que posteriormente servirán de entrada, que alimentarán el proceso principal de la aplicación. Las fuentes generadoras de información son: red social Twitter y sensores de volumen de tráfico ubicados en ciertas intersecciones consideradas conflictivas para la congestión vehicular. La información generada por los sensores corresponde a la cantidad autos que están circulando a una distancia de tres metros a partir de la ubicación del sensor instalado, hacia adelante. La distancia establecida, obedece a la longitud promedio de un auto, que es la requerida, para que el sensor pueda detectar el movimiento vehicular.

### 4.2. Fase de recolección

En esta fase existen dos mecanismos de recolección de datos que serán utilizados por la aplicación propuesta, los cuales constituyen una importante fuente de datos primarios:

- *API Twitter*, mediante el cual se obtienen mensajes que reflejan la opinión de la población, en relación a situaciones de congestión vehicular detectadas.
- *Sensores de volumen de tráfico*, el segundo componente importante está relacionado con la información recolectada por los sensores que alimentan el Sistema de Registro Vehicular, que es manejado por el Centro de Gestión de la Movilidad (CGM) de la ciudad de Quito; en el mismo se registra la carga vehicular existente en determinadas intersecciones de la ciudad.

Para obtener la información relacionada con este componente, se realizaron varias entrevistas con autoridades del CGM de Quito, obteniéndose dos elementos importantes: conteos vehiculares y la georreferenciación de los sensores.

### 4.3. Fase de integración

Con la finalidad de agrupar los tuits generados, junto con los sensores distribuidos en la ciudad, se utilizó el método de agrupación K-means [14], concentrando los tuits en varios grupos dependiendo de la ubicación de los sensores.

Posteriormente, se filtró el grupo de tuits generados, mediante un análisis semántico y de sentimientos [15], recogiendo los tuits en los que se denota malestar y cuyo texto final mencione temas de congestión vehicular.

Finalmente, el grupo de tuits fue integrado con la información generada por el conteo de los sensores, para su posterior análisis y contraste de la información.

### 4.4. Fase de análisis

Con la información recolectada mediante redes sociales y sensores de congestión vehicular se obtuvieron índices de congestión. Los mismos fueron generados para cada uno de los sectores en los cuales están ubicados los sensores y alrededor de los cuales existen tuits reportando situaciones de congestión. Los índices son:

- RVD (retraso viaje días).
- FRv (nivel de congestión del tráfico).
- Índice de tiempo de viaje (TTI).

Posteriormente, se determinó la correlación de Pearson [16] entre los tres índices generados. En la Tabla 1 se puede determinar si existe tráfico en cada uno de los sectores de la ciudad, mediante la aplicación de reglas de decisión.

El nivel de correlación Alta, es la que se considera en la investigación, para mostrar gráficamente posibles puntos de congestión vehicular.

Tabla 1.  
Reglas de correlación y resultados

Regla	Correlación	Resultado
R=0	Nula	Tráfico nulo
0<R≤0.4	Débil	Tráfico leve
0.4<R≤0.7	Media	Tráfico medio
0.7<R≤1	Alta	Congestión

Fuente. Los autores

#### 4.5. Fase de presentación

Finalmente, con los datos almacenados y el procesamiento de los mismos, se ha creado una aplicación desarrollada en JSP para el *backend*, JQuery (Biblioteca de Java) y bootstrap para el *frontend*. Se utilizó Mysql como motor de base de datos y como servidor web Apache Tomcat. El entorno de desarrollo Netbean IDE. El sistema desarrollado es multinavegador y multiplataforma.

El software creado permite incluir la base de datos tanto de los lugares donde existe un sensor en Quito, como de la base de datos de tuits que indican la existencia de tráfico en la ciudad. Esta herramienta web permite apreciar cartográficamente los resultados del análisis y poder determinar posibles zonas de congestión vehicular.

El desarrollo del software contempló tres partes centrales que son:

1. Presentación de lugares de mayor tráfico en la ciudad, considerando los datos de los sensores que dispone el Municipio de Quito.
2. Presentación en tiempo real de los tuits generados en la ciudad, mediante de la red social Twitter, considerando para ello el parámetro ubicación del tuit. Sin embargo, la principal función es presentar en tiempo real el tráfico en la ciudad.
3. Análisis de resultados mediante reportes tanto gráficos como analíticos de los diferentes ámbitos de análisis de tráfico en la ciudad a partir de los tuits cargados en la base de datos.

#### 5. Elección de las herramientas a utilizar

Para determinar la plataforma a usar en la propuesta de desarrollo, en la Tabla 2 se muestra el análisis de las distintas opciones existentes.

Apache Hadoop<sup>3</sup> es actualmente la herramienta libre más utilizada para el análisis de Big Data. Hadoop proporciona dos elementos principales: un *framework MapReduce* y un sistema de archivos (*Hadoop Distributed File Systems*, HDFS). Por un lado, el HDFS provee una base de datos tolerante a fallos con una alta disponibilidad. Por otra parte, el paradigma de programación MapReduce, permite dividir y paralelizar los cálculos entre un número indefinido de ordenadores de bajo coste. En el uso de un clúster de computadores convencionales para el análisis de datos radica una de las mayores ventajas de Apache Hadoop [17].

Tabla 2.  
Indicadores de plataformas usadas para *Big Data*.

Plataforma	A	B	C	D
Apache Hadoop	No	Alta	Alta	Sí
Disco project	No	Alta	Baja	Sí
Apache Storm	No	Alta	Baja	Sí

A: Limitación de entrada de datos.  
B: Adaptabilidad con otros lenguajes.  
C: Facilidad de uso y soporte.  
D: Uso en temas relacionados.

Fuente. Los autores

Por tanto, Hadoop proporciona una doble funcionalidad que en algunos casos las plataformas alternativas no van a poder ofrecer. Es decir, algunas de las herramientas cumplen únicamente una de las funcionalidades de Hadoop, aunque de un modo muy eficiente [17].

La arquitectura de Hadoop permite asegurar la entrega de mecanismos de conmutación por error y recuperación; está basada en flujos de contenidos multimedia en la Web. La red social Twitter genera información que se puede recolectar mediante Apache Flume; una herramienta que toma esta información en línea y la envía a un almacén centralizado como HDFS o HBase [18].

#### 5.1. Herramientas para la obtención de datos

Existe gran variedad de bibliotecas para las API de Twitter en diferentes lenguajes. Recientemente, Twitter dispone de una librería hbc; un cliente Java HTTP para consumir de Streaming API, la cual tiene como requisito introducir las respuestas en formato JSON que devuelve Streaming API dentro de Hadoop Distributed File System. Para elegir la opción más adecuada se puede realizar una comparación entre los componentes de software más comunes para este propósito, Apache: Sqoop, Kafka, Scribe y Flume [19].

- **Sqoop.** Biblioteca que permite importar datos desde una base de datos estructurada, organizada en entidades que tienen un formato definido, hacia Hadoop. Permitiendo además exportar datos en sentido contrario. No se va a utilizar en el proyecto ya que no se desea importar datos desde una base de datos relacional.
- **Kafka.** Su principal caso de uso es un sistema distribuido de paso de mensajes publicación-suscripción. Adecuado para sistemas altamente confiables y escalables de mensajería empresarial en los que se deben conectar múltiples sistemas, incluyendo Hadoop. Tampoco haremos uso de este software que no se necesita un sistema de paso de mensajes con varios tipos de sistemas informáticos.
- **Scribe.** Está diseñado para escalar a un número muy grande de nodos y resistir a fallos en la red y en el nodo. El soporte que recibe cada vez es menor.
- **Flume.** Servicio distribuido, seguro y con alta disponibilidad para una eficiente recolección, unión y movimiento de grandes cantidades no solo de log data, sino también de cantidades masivas de eventos de datos. Incluye tráfico de datos de red y datos generados por "social media" procedentes de fuentes de datos no relacionales.

Luego de este análisis se seleccionó Apache Flume ya que maneja datos semiestructurados como son los tuits en JSON. Una de las ventajas de esta herramienta es su estructura: simple, robusta y flexible, orientada a "Streams"; los flujos de datos se componen de agentes que pueden agregar o transformar los eventos.

La misma que es una solución Java que permite recolectar y mover grandes cantidades de datos, desde algunas fuentes, principalmente servidores de aplicaciones HDSF con mayor

<sup>3</sup> El tipo de instalación realizada es Single Node

```

TwitterAgent.sources.Twitter.keywords=      atasco,
embotellamiento, transito, circulación, congestión,
atascamiento, trafico, #tráficoUIO, #MOVILIDADuio,
#viasAlternasUIO, #RadaresUIO
TwitterAgent.sources.Twitter.locations=--92.68,-S.01,-
75.19--2.34
TwitterAgent.sinks.HDFS.hdfs.path=hdfs://localhost:
54310/user/hduser/flume/tweets
    
```

Figura 5. Configuración Apache Flume  
Fuentes. Los autores.

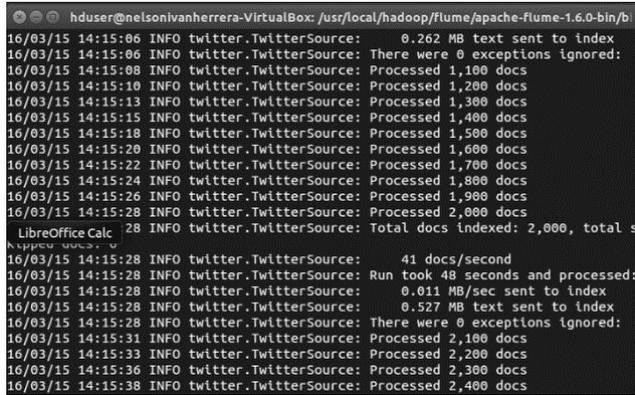


Figura 6. Pantalla de ejecución del agente Flume  
Fuente. Los autores

velocidad a un espacio de almacenamiento de datos, permitiendo subir datos de Hadoop [20].

La instalación fue realizada en el sistema operativo Ubuntu 14.04, la Fig. 5 muestra los parámetros de configuración del agente Apache Flume<sup>4</sup>.

Los campos utilizados para la configuración son:

**TwitterAgent.sources.Twitter.keywords:** esta configuración está dividida en dos grupos que son: keywords y hashtag. Los hashtags son los utilizados por el centro de gestión de la movilidad, por ello se los consideró.

**TwitterAgent.sources.Twitter.locations:** en esta configuración se colocó los puntos cardinales de la ciudad de Quito.

**TwitterAgent.sinks.hdfs.pathTwitter.locations:** dentro de esta configuración se colocó la ruta en donde se almacenarán los tuits recolectados.

En la Fig. 6, se muestra la ejecución del agente Flume, que tomará la información de tuits relacionados con situaciones de congestión vehicular, reportados por los ciudadanos.

### 5.2. Herramientas para el almacenamiento de datos

Para el almacenamiento de los datos y acorde con el análisis anterior, se ha seleccionado Apache Hadoop. Éste es un *framework* de software, que permite a las aplicaciones trabajar con miles de nodos. También es de licencia libre y se instaló la versión 2.6.0.

La Fig. 7 muestra la información de tuits almacenados en el sistema de archivos de Hadoop, se puede apreciar los permisos, el grupo, el tamaño y el nombre de cada uno de ellos.

### Browse Directory

Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	hduser	supergroup	206.58 MB	1	128 MB	FlumeData.1455033100115
-rw-r--r--	hduser	supergroup	74.25 MB	1	128 MB	FlumeData.1455033710095
-rw-r--r--	hduser	supergroup	128 MB	1	128 MB	FlumeData.1458069284046.tmp
-rw-r--r--	hduser	supergroup	0 B	1	128 MB	FlumeData.1458069284047
-rw-r--r--	hduser	supergroup	166.54 MB	1	128 MB	FlumeData.1458069284048
-rw-r--r--	hduser	supergroup	151.55 MB	1	128 MB	FlumeData.1458069284049

Figura 7. Recolección de tweets almacenados en Hadoop  
Fuente. Los autores

### 5.3. Herramientas para el procesamiento de datos

A continuación, se realizará una comparación entre los componentes de software más comunes para analizar el conjunto de datos JSON almacenados en bruto, que contiene HDFS. Las herramientas son: Hive (Apache Hive), Cascading (Cascading) y Pig (Apache Pig). Para investigación sobre datos en bruto son mejores los lenguajes procedimentales [19].

- **Cascading.** Framework de procesamiento de datos procedimental que permite a los usuarios construir flujos de datos en Java sobre Hadoop. Usa MapReduce para ejecutar todo su procesamiento de datos. Proporciona una librería de operadores, aunque también se pueden construir unos propios [19].
- **Pig.** Motor para ejecutar flujos de datos en paralelo sobre Hadoop. Usa MapReduce para ejecutar todo su procesamiento de datos. Para expresar estos flujos de datos utiliza el lenguaje procedimental Pig Latin. El cual incluye operadores tradicionales como join, sort, filter, etc. así como la posibilidad para los programadores de desarrollar sus propias funciones de lectura, procesamiento y escritura de datos [19].

Tanto Cascading como Pig son adecuados para Twitter Data Collection pero, finalmente, se ha escogido Pig ya que tiene más compatibilidades con todo el entorno Hadoop.

Pig Latin es un lenguaje de flujo de datos. Permite a los usuarios escribir cómo los datos de una o más entradas deben ser leídos, procesados y almacenados en una o más salidas en paralelo. Pig al igual que MapReduce, está orientado al procesamiento por lotes de grandes porciones del conjunto total de datos. Lee todos los registros de un archivo y escribe todas sus salidas secuencialmente. [19]

Se ha considerado el uso de Apache Pig para el procesamiento de datos, debido a que esta aplicación es un lenguaje de procedimientos que permite la consulta de grandes conjuntos de datos semiestructurados haciendo uso de Hadoop. La versión que se utilizó de Apache Pig es pig-0.16.0

## 6. Resultados

### 6.1. Información presentada por los sensores en Quito

Con el análisis de los valores presentados por los sensores y la relación con el parámetro de tráfico establecido por el

<sup>4</sup> La versión de utilizada es apache-flume-1.6.0



Figura 8. Mapa de tráfico de la ciudad de Quito (sensores)  
Fuente. Los autores

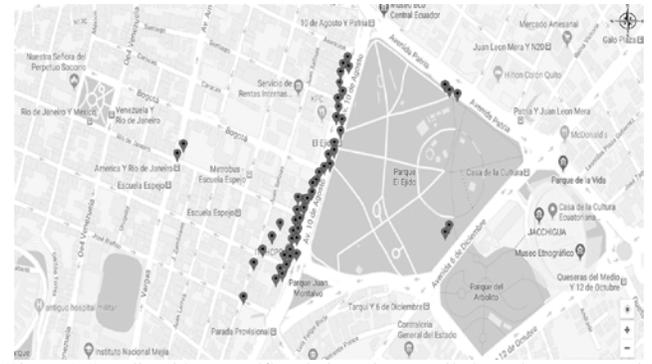


Figura 9. Localización geográfica de los tuits recopilados sobre un mapa de Quito  
Fuente. Los autores

CGM, la Fig. 8 presenta cartográficamente posibles puntos de congestión vehicular en la ciudad de Quito, que corresponden al tráfico existente en el sector norte de la ciudad en el horario de 6:00 am a 7:00 am.

Como se puede observar, existen dos arterias principales donde existe tráfico vehicular que son la Av. Prensa – Diego de Vásquez y la 10 de agosto. Además, se observa tráfico en la Panamericana Norte y en la Av. Eloy Alfaro, con ello es claramente identificable los tramos donde existe congestión vehicular. El gráfico presentado corresponde a solamente un segmento de la ciudad; sin embargo, se puede observar que existe una simetría en los datos. Esto se debe a que los sensores no se encuentran a lo largo de toda la ciudad; solamente se encuentran en vías específicas que han sido consideradas más transitadas. Pueden existir muchos lugares con tráfico y que no son detectados, pues la cobertura de los sensores, no abarca todos los posibles lugares de congestión vehicular.

Uno de los puntos centrales de la aplicación en la cual se utilizarán las herramientas estudiadas, es el poder presentar en tiempo real el tráfico en la ciudad. Para esto es necesario captar los tuits indicativos de tráfico, la recopilación en una base de datos y la presentación gráfica de los puntos donde existe congestión.

De la información obtenida por la municipalidad de Quito, actualmente los sensores generan información, que es utilizada bajo demanda para analizar situaciones de congestión vehicular. Hoy en día esta entidad no dispone de una aplicación que le permita observar en tiempo real el tráfico en la ciudad, además que la misma solo hace referencia a los sectores en los cuales están instalados los sensores, lo cual no permite analizar todos los sectores que registran tráfico vehicular, en la localidad.

## 6.2. Información recolectada de la red social Twitter

La aplicación creada permite, a través de los tuits; disponer de información inmediata del tráfico, presentando los lugares donde al momento las personas indica que existe tráfico vehicular, con una actualización cada 5 minutos. Sin embargo, la alerta de tráfico se mantiene por un período de 30 minutos, lo cual permite observar de mejor manera el tráfico en la ciudad.

Para observar y verificar los resultados, se realizó un estudio inicial por un período de 3 meses, incluyendo los meses de mayo,

abril y junio en los que se obtuvo información que sirve para validar la investigación realizada, contrastándola con la generada por los sensores existentes en la urbe.

En la figura presentada por el sistema puede observarse parte de la ciudad de Quito, donde se ha generado congestión vehicular. Se puede evidenciar también la existencia de lugares donde existe tráfico vehicular, de acuerdo a lo indicado por la población.

Comparando esta información con la presentada por los puntos informados a través de los sensores, se verifica por una parte la eficacia del sistema y por otra la falta de eficacia para determinar todos los puntos de tráfico con los sensores del municipio.

Al observar los datos generales de tráfico en la ciudad de Quito, señalados en la Fig. 9, establecidos a través de tuits, se puede observar que en la parte urbana existe a lo largo del día acumulación en diversos puntos, existiendo mayormente a nivel de sector centro norte puntos de congestión vehicular.

Con esta información se puede observar puntos de congestión, tanto con la información de los sensores, como de los tuits, mediante la interfaz gráfica de la aplicación. Para esto se utilizó la API de Google Mapas.

## 7. Conclusiones y trabajos futuros

La investigación ha permitido determinar, por una parte, la necesidad de la ciudad de Quito por contar con herramientas que permitan la toma de decisiones para mejorar la congestión vehicular.

Las herramientas de sustento y recopilación de la información analizada y considerada, han permitido la creación de una aplicación que permite medir la congestión vehicular. Para lo cual, se hace uso de los sensores existentes en la ciudad, pero principalmente mediante la data proporcionada por la red social Twitter, misma que ha sido recopilada para su procesamiento.

Los datos proporcionados por los tuits mostraron, en la mayoría de casos, mayor cobertura que los que presentan los sensores colocados a lo largo de la ciudad.

Se debe considerar como una opción viable y de menor costo el uso de la aplicación desarrollada, que toma los datos obtenidos de los tuits, para la toma de decisiones sobre congestión vehicular.

Se verifica la importancia y potencialidad de la aplicación, para que la misma sea utilizada como: ente emisor de alertas a los conductores, en la eficiencia y control del transporte público, entre otras, en beneficio de la ciudad de Quito.

Existe un sesgo en la investigación realizada, pues la misma se centra solamente en ciertos lugares de la ciudad en donde existen sensores, lo cual genera resultados parciales. Frente a esto, se recomienda considerar la implementación de sistemas de bajo costo (Arduino) [21], mediante la instalación de una constelación de sensores, en el resto de lugares de la ciudad.

La investigación realizada tiene una limitante importante, que está relacionada con la necesidad de que las cuentas de Twitter utilizadas deben tener activada la opción de geolocalización.

Una importante utilidad, que no forma parte del alcance de la presente investigación, hace referencia a poder retuitar los mensajes, informando rutas alternativas en caso de que las principales estén congestionadas.

Se limita la utilización del hashtag #TraficoUIO con la comunidad universitaria UTE.

Como trabajo futuro, se debe considerar mostrar gráficamente el tráfico desde el nivel Débil hasta el nivel Medio, y poder ejecutar acciones que prevengan situaciones de congestión vehicular.

Para comprobar que el sistema de sistema de información propuesto posibilita detectar situaciones de congestión vehicular en la ciudad de Quito. Se utiliza para el análisis estadístico el coeficiente de Pearson; que correlaciona el indicador de congestión vehicular de los sensores y el detectado por los mensajes de la red social Twitter.

## Referencias

- [1] INEC, Proyecciones de la población Ecuatoriana según Cantón, Cifras y estadísticas. [En línea]. [Consultado: 01-01-2015]. Disponible en: [http://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion\\_y\\_Demografia/Proyecciones\\_Poblacionales/proyeccion\\_cantonal\\_total\\_2010-2020.xlsx](http://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Proyecciones_Poblacionales/proyeccion_cantonal_total_2010-2020.xlsx). [Último acceso: 10 12 2017].
- [2] Alcaldía de Quito, Noticias, Distrito Metropolitano de Quito. [En línea]. [Consultado: 01-01-2011]. Disponible en: <http://www.diariocritico.com/noticia/155769/noticias/la-congestion-vehicular-sigue-sin-solucion-en-quito.html>. [Último acceso: 15 10 2015].
- [3] Pinto, G., Hacia una movilidad inteligente en la ciudad de Quito, pp. 20-25, 2013.
- [4] Herrera, N.H. and Santamaria, H.S., Analysis of the factors generating vehicular traffic, de ICEDEG Conference on eDemocracy & eGovernment, Quito, 2016.
- [5] Zikopoulos, C.E.P., Understanding Big Data: analytics for enterprise class hadoop and streaming data, Chicago: McGraw-Hill Osborne Media, 2011.
- [6] Aguilar, L.J., Big Data análisis de grandes volúmenes de datos en organizaciones, México: Alfaomega Grupo Editor, S.A, 2013.
- [7] Delgado, P.S.M.V.J., CBI Ciencias básicas e ingeniería. [En línea]. [Consultado: 7-12-2011]. Disponible en: <http://sgpwe.izt.uam.mx/pages/cbi/psb/vfinal.pdf>. [Último acceso: 08 03 2018].
- [8] Kwak, D., Liu, R., Kim, D., Nath, B. and Iftode, L., Seeing is believing: sharing real-time visual traffic information via vehicular clouds, IEEE Access, 4(12), pp. 3617-3631, 2016. DOI: 10.1109/ACCESS.2016.2569585
- [9] Cárdenas-Benítez, N., Aquino-Santos, R., Magaña-Espinoza, P., Aguilar-Velazco, J., Edwards-Block, A. and Medina-Cass, A., Traffic congestion detection system through connected vehicles and Big Data, Sensors, 16(5), p. 599, 2016. DOI: 10.3390/s16050599
- [10] Tostes-Ribeiro, A., Henrique-Silva, T., Duarte-Figueiredo F. and Loureiro, A.A.F., Studying traffic conditions by analyzing foursquare and instagram data, in: Symposium on performance evaluation of wireless ad hoc, sensor, & ubiquitous networks, Canada, 2014.
- [11] P.D.B.L. Eleonora D' Andrea, Real-Time Detection of Traffic From Twitter Stream Analysis, IEEE Transactions on Intelligent Transportation Systems, 16(4), pp. 2269-2283, 2015. DOI: 10.1109/TITS.2015.2404431
- [12] Nellore, K. and Hancke, G.P., A survey on urban traffic management system using wireless sensor networks, Sensors, 16(2), p. 157, 2016. DOI: 10.3390/s16020157
- [13] García, F.C., Umap, inteligencia colectiva extraída de las redes sociales, Profit, Barcelona-España, 2010.
- [14] Reyes-Zambrano, G. and Hidalgo-Veliz, R.N., Vehicular swept path analysis using K-means algorithm, International Journal of Applied Engineering Research, 11(18), pp. 9536-9538, 2016.
- [15] Villena-Román, J., Luna-Cobos, A. y González-Cristóbal, J.C., Análisis semántico de la opinión de los ciudadanos, SEPLEN, 53, pp. 59-162, 2014.
- [16] Universidad Distrital Francisco José de Caldas, [En línea]. [Consultado: 08-03-2018]. Disponible en: <http://repository.udistrital.edu.co/bitstream/11349/3815/3/ParraRincónDanielaAlejandraAnexo-3.pdf>.
- [17] Herramientas para Big Data: Entorno Hadoop., 2014.
- [18] Time Outsourcing Center, Introducción a Hadoop y su ecosistema, 2015. [En línea]. Disponible en: <http://www.ticout.com/blog/2013/04/02/introduccion-a-hadoop-y-su-ecosistema/>.
- [19] Pérez, B.G., TDC (Twitter Data Collection): creación de una gran base de datos de Tweets., Universidad Autónoma de Madrid, Madrid, España, 2014.
- [20] Schmarzo, B., Big data, el poder de los datos, Estados Unidos: Anaya Multimedia-Anaya Interactiva, 2014.
- [21] Faludi, R., Building Wireless sensor networks, United States of America: O' Reilly Media, 2011.

**N.I. Herrera-Herrera**, es profesor auxiliar de la Facultad de Ciencias de la Ingeniería e Industrias de la Universidad Tecnológica Equinoccial, Quito Ecuador. Obtuvo su grado de MSc. en Gerencia de Sistemas en la Universidad de las Fuerzas Armadas "ESPE", Quito-Ecuador. En la actualidad está finalizando su grado doctoral en la Universidad de Extremadura, España. Profesionallymente ha colaborado con importantes instituciones financieras en el manejo de proyectos tecnológicos. Su especialidad de investigación está centrada en ciudades inteligentes con Big Data. Desarrolla e implementa sistemas informáticos para Pymes. ORCID: 0000-0002-5781-6444

**S. Lujan-Mora**, es profesor titular de Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante. Obtuvo su grado de Doctor Ingeniero en Informática en la Universidad de Alicante, España en el año 2005 con una tesis sobre el modelado de almacenes de datos con UML. Ha realizado estancias posdoctorales en Estados Unidos y Ecuador. Desarrolla e implementa aplicaciones web desde 1997 y enseña a hacerlo en la universidad desde 1999. Sus temas principales de investigación incluyen el desarrollo de aplicaciones web, la usabilidad y la accesibilidad web, el e-learning, los cursos en línea abiertos y masivos (Massive Open Online Courses, MOOC) y los recursos educativos abiertos (REA). ORCID: 0000-0001-5000-864X

**E.R. Gómez-Torres**, se graduó de Ing en Sistemas en 1998, es MSc. en Gerencia de Sistemas en el 2001, todos ellos de la Universidad de las Fuerzas Armadas "ESPE" en Quito-Ecuador. Desde 2000 hasta la presente fecha labora como profesor en la ESPE, trabajando también como consultor en diferentes empresas del sector privado y público en el Ecuador y otros países; posteriormente se unió a trabajar en la Universidad Tecnológica Equinoccial, Quito, Ecuador, a partir del 2013 hasta la presente fecha. Su campo de investigación incluyen líneas como: desarrollo de aplicaciones móviles en contextos, Big data, inteligencia de negocios, analítica de datos, machine learning, IoT, modelos predictivos y minería de texto y datos. ORCID: 0000-0002-1171-7256