

UNIVERSIDAD DE MURCIA

MÁSTER UNIVERSITARIO EN TECNOLOGÍAS DE ANÁLISIS DE DATOS  
MASIVOS

TRABAJO FIN DE MÁSTER

---

# Una arquitectura Big Data para el procesamiento en tiempo real de datos en aplicaciones de gestión de flotas

---



*Alumno:*  
*Rubén Garrido Montesinos*

*Directores:*  
*Jesús J. García Molina*  
*Diego Sevilla Ruiz*

Septiembre de 2018

## Resumen

Extended abstract

# Índice

<b>1. Introducción</b>	<b>5</b>
1.1. Motivación . . . . .	5
1.2. Objetivos . . . . .	5
1.2.1. Objetivo principal . . . . .	5
1.2.2. Objetivos secundarios . . . . .	5
1.3. Metodología . . . . .	6
1.4. Organización del documento . . . . .	6
<b>2. Estado del arte</b>	<b>7</b>
<b>3. Fundamentos y herramientas</b>	<b>8</b>
3.1. Qué es el Big data . . . . .	8
3.2. Docker . . . . .	8
3.3. Apache Hadoop . . . . .	8
3.4. Apache Spark . . . . .	8
3.5. Apache Kafka . . . . .	8
3.6. Elastic . . . . .	8
3.7. MongoDB . . . . .	8
<b>4. Requisitos</b>	<b>9</b>
<b>5. Arquitectura propuesta</b>	<b>10</b>
5.1. Modelo de arquitectura . . . . .	10
5.2. Selección de las herramientas elegidas . . . . .	10

<b>6. Aplicación desarrollada</b>	<b>11</b>
6.1. Hardware utilizado . . . . .	11
6.2. Datos obtenidos y preprocesado . . . . .	11
6.3. Detalles de implementación . . . . .	11
6.3.1. Diseño de la arquitectura con las herramientas seleccionadas . . . . .	11
6.3.2. Montaje de la imagen de Ubuntu . . . . .	11
6.3.3. Pasos a seguir para añadir aplicaciones sobre la imagen de Ubuntu . . . . .	11
6.3.4. Montaje de las imágenes de Hadoop . . . . .	11
6.3.5. Montaje de las imágenes de Zookeeper y Kafka . . . . .	11
6.3.6. Montaje de las imágenes de Spark . . . . .	11
6.3.7. Montaje de las imágenes Elastic . . . . .	11
6.3.8. Montaje de las imágenes de MongoDB en integración de OSM . . . . .	11
6.3.9. Integración de los datos y pruebas . . . . .	11
6.4. Uso de la arquitectura . . . . .	11
<b>7. Validación</b>	<b>12</b>
<b>8. Conclusiones y trabajo futuro</b>	<b>13</b>

# 1 Introducción

## 1.1 Motivación

Internet-of-Things (IoT) y las tecnologías Big Data han producido avances significativos en el dominio de los sistemas de gestión de flotas de vehículos. El paradigma IoT ha permitido mejorar el proceso de seguimiento y monitorización de vehículos y las técnicas Big Data son muy apropiadas para el análisis en tiempo real de la gran cantidad de datos obtenidos en este proceso (Loukides and Bruner, 2015). De este modo, han surgido nuevas aplicaciones de gestión de flotas que gestionan mejor los recursos de la empresa y ofrecen un procesamiento más sofisticado y escalable en sus diferentes escenarios.

Las aplicaciones de gestión de flotas son, por tanto, un dominio apropiado para la aplicación de arquitecturas Big Data. El seguimiento de los vehículos genera un gran volumen de datos que, a través de diferentes técnicas de análisis de datos, podemos extraer información de gran interés para las empresas dedicadas a este sector. Dado que las aplicaciones de gestión de flotas, deben procesar esta gran cantidad de datos y proporcionar diferentes beneficios, las tecnologías Big Data son muy apropiadas para tratarlos. Algunos de los beneficios que podemos obtener con este tipo de aplicaciones es reducir el coste de gestionar la flota, ser más responsable con el medio ambiente y poder controlar cualquier tipo de robo o mal uso de los vehículos de la empresa propietaria de la flota (Gómez, 2018) (de Aledo, 2018).

Movildata<sup>1</sup> es una empresa con sede en Murcia dedicada a ofrecer soluciones para la gestión de flotas. Esta empresa se ha integrado recientemente en Verizon Connect. Antes de llevarse a cabo esta integración, los tutores y alumno de este proyecto acordaron con Movildata desarrollar un proyecto piloto destinado a diseñar e implementar una arquitectura Big Data que se aplicase para ofrecer alternativas y nuevas funcionalidades a su solución de gestión de flotas. La empresa disponía de aplicaciones basadas en arquitecturas tradicionales con lo que el proyecto serviría como prueba de concepto de aplicación de tecnologías Big Data.

## 1.2 Objetivos

### 1.2.1 Objetivo principal

El objetivo principal de esta tesis de máster ha sido la elección de una arquitectura big data para aplicaciones de gestión de flotas y su evaluación en un caso de estudio definido a partir de la información proporcionada por Movildata. Se realizará una prueba de concepto de la arquitectura que ayude a la empresa a conocer las nuevas tecnologías Big Data y cómo se podría beneficiar de su aplicación.

---

<sup>1</sup><https://movildata.com/sobre-nosotros/>

### 1.2.2 Objetivos secundarios

Entre los objetivos secundarios encontramos los requerimientos típicos de las tecnologías Big Data. Por un lado, dicha arquitectura debe ser fácil de administrar y ampliar es decir, debe ser fácilmente escalable. En nuestro caso, buscamos reconocer la dificultad y capacidad de mantener estas tecnologías y la capacidad a tolerar fallos.

A pesar de ser una prueba de concepto, tendremos que enfrentarnos a las dificultades de implementar y mantener la arquitectura. Además, veremos la capacidad de realizar nuevos desarrollos sobre la misma, comprobar la facilidad de reemplazar cualquiera de las herramientas que la componen y comprobar cómo funcionan. Por último tendremos valorar la capacidad de reemplazar a las tecnologías propuestas con las que se usaban tradicionalmente.

Dado que el objetivo principal es Investigar las diferentes arquitecturas y tecnologías aplicables para el problema abordado, se deberán justificar las razones por las que hemos seleccionado determinadas herramientas. Por tanto, será necesario evaluar las distintas herramientas que nos ofrece el mercado.

Tras esto, decir que el hardware de desarrollo es limitado por lo que se debe encontrar la forma de exportar fácilmente las diferentes configuraciones. Por otro lado, nos ayudará a valorar si cumple los requisitos mínimos de rendimiento de las diferentes herramientas.

Por último, se debe comprobar que la solución es escalable horizontalmente es decir, se escalará añadiendo más máquinas y no añadiendo más hardware al servidor. Dado esto se deberá usar una tecnología de virtualización suficientemente potente y ligera para poder añadir y quitar máquinas que proporcionen capacidad a la estructura. Por otro lado, la escalabilidad debe darse tanto en el almacenamiento como en procesamiento.

## 1.3 Metodología

## 1.4 Organización del documento

- Carga y almacenamiento
- Aritmética
- Conversión de tipos
- Creación y manipulación de objetos
- Gestión de pilas (push y pop)
- Transferencias de control (branching)
- Invocación y retorno a métodos
- Excepciones

## 2 Estado del arte



## 3 Fundamentos y herramientas

3.1 Qué es el Big data

3.2 Docker

3.3 Apache Hadoop

3.4 Apache Spark

3.5 Apache Kafka

3.6 Elastic

3.7 MongoDB

## 4 Requisitos

## 5 Arquitectura propuesta

### 5.1 Modelo de arquitectura

### 5.2 Selección de las herramientas elegidas

## 6 Aplicación desarrollada

### 6.1 Hardware utilizado

### 6.2 Datos obtenidos y preprocesado

### 6.3 Detalles de implementación

#### 6.3.1 Diseño de la arquitectura con las herramientas seleccionadas

#### 6.3.2 Montaje de la imagen de Ubuntu

#### 6.3.3 Pasos a seguir para añadir aplicaciones sobre la imagen de Ubuntu

#### 6.3.4 Montaje de las imágenes de Hadoop

#### 6.3.5 Montaje de las imágenes de Zookeeper y Kafka

#### 6.3.6 Montaje de las imágenes de Spark

#### 6.3.7 Montaje de las imágenes Elastic

#### 6.3.8 Montaje de las imágenes de MongoDB en integración de OSM

#### 6.3.9 Integración de los datos y pruebas

### 6.4 Uso de la arquitectura

## 7 Validación

## 8 Conclusiones y trabajo futuro

# Referencias

Opengts. <http://www.opengts.org/>.

AWS. Presentación de la solución aws connected vehicle. <https://aws.amazon.com/es/about-aws/whats-new/2017/11/introducing-the-aws-connected-vehicle-solution/>.

Azure. Lambda architecture for connected car fleet management. <https://azure.microsoft.com/en-in/resources/videos/build-2017-lambda-architecture-for-connected-car-fleet-management/>.

Baghel, A. (2016). Traffic data monitoring using iot, kafka and spark streaming. <https://www.infoq.com/articles/traffic-data-monitoring-iot-kafka-and-spark-streaming>.

de Aledo, A. G. (2018). El supremo avala el gps para vigilar al trabajador. [https://www.diariodecadiz.es/provincia/Supremo-avala-GPS-vigilar-trabajador\\_0\\_1276372522.html](https://www.diariodecadiz.es/provincia/Supremo-avala-GPS-vigilar-trabajador_0_1276372522.html).

DGT (2010). Cuadro para excesos de velocidad. [https://sede.dgt.gob.es/Galerias/tramites-y-multas/alguna-multa/consulta-de-sanciones-por-exceso-velocidad/cuadro\\_velocidad.pdf](https://sede.dgt.gob.es/Galerias/tramites-y-multas/alguna-multa/consulta-de-sanciones-por-exceso-velocidad/cuadro_velocidad.pdf).

Fleet, A. Fleet management software: Build or buy? [https://cdn2.hubspot.net/hubfs/437692/E-Guides,\\_e-books,\\_etc./Build\\_or\\_Buy\\_E-Book\\_20170629-1.pdf](https://cdn2.hubspot.net/hubfs/437692/E-Guides,_e-books,_etc./Build_or_Buy_E-Book_20170629-1.pdf).

Fomento, M. Tiempos de conducción y descanso. <https://www.fomento.gob.es/transporte-terrestre/inspeccion-y-seguridad-en-el-transporte/tempos-de-conduccion-y-descanso/conduccion/tempos-de-conduccion>.

Frampton, M. (2016). *Big Data Made Easy, A Working Guide to the Complete Hadoop Toolset*. APRESS.

Ghavam, P. K. (2016). *BIG DATA GOVERNANCE, Modern Data Management Principles for Hadoop, NoSQL and Big Data Analytics*. CreateSpace Independent Publishing Platform.

Gáspár, P., Szalay, Z., and Aradi, S. (2014). Highly automated vehicle systems. [http://www.mogi.bme.hu/TAMOP/jarmurendszerek\\_iranyitasa\\_angol/](http://www.mogi.bme.hu/TAMOP/jarmurendszerek_iranyitasa_angol/).

Gómez, J. M. F. (2018). *Introducción a la gestión de flotas de vehículos*. SafeCreative.

Hurwitz, J., Nugent, A., Halper, F., and Kaufman, M. (2013). *Big Data for Dummies*. John Wiley and Sons, Inc.

Karimi, A., Olsson, J., and Rydell, J. (2004). *A Software Architecture Approach to Remote Vehicle Diagnostics*. PhD thesis, IT UNIVERSITY OF GÖTEBORG.

Lohr, S. (2013). Business, innovation, technology, society the origins of ‘big data’: An etymological detective story. <https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/>.

- Loukides, M. and Bruner, J. (2015). *What Is the Internet of Things?* O'Reilly Media.
- Marz, N. and Warren, J. (2015). *Big Data, Principles and best practices of scalable real-time data systems*. MANNING.
- Oracle. Improving logistics and transportation performance with big data. <http://www.oracle.com/us/technologies/big-data/big-data-logistics-2398953.pdf>.
- Saghaei, H. (2016). Design and implementation of a fleet management system using novel gps/lonass tracker and web - based software. In *International Conference on New Research Achievements in Electrical and Computer Engineering*.
- Sateltrack (2009). The professional solution for web-enabled fleet management. [https://www.sateltrack.com/download/sateltrack\\_features\\_en.pdf](https://www.sateltrack.com/download/sateltrack_features_en.pdf).
- Schorpp, S. (2010). Dynamic fleet management for international truck transportation. Technical report.
- Syafrudin, M., Fitriyani, N. L., Rhee, J., Kang, Y.-S., Li, D., and Alfian, G. (2017). An open source-based real-time data processing architecture framework for manufacturing sustainability.
- Verizon (2015). United states securities and exchange commission. <https://www.verizon.com/about/sites/default/files/FLTX%2010K.PDF>.
- YugaByte. Iot fleet management - spark and kafka. <https://docs.yugabyte.com/latest/develop/realworld-apps/iot-spark-kafka/>.