



---

---

# Diseño conceptual del modelo de *big data* para el IMT

Alfredo Bueno Solano  
Carlos Mario Pérez González  
Jared Piña Barcenas  
José Elías Jiménez Sánchez  
Miguel Gastón Cedillo Campos  
Marisol Barrón Bastida

**Publicación Técnica No. 537**  
**Sanfandila, Qro, 2018**



---

**SECRETARÍA DE COMUNICACIONES Y TRANSPORTES**  
**INSTITUTO MEXICANO DEL TRANSPORTE**

**Diseño conceptual del modelo de *big data* para el  
IMT**

**Publicación Técnica No. 537**  
**Sanfandila, Qro, 2018**

---



Esta investigación fue realizada en la Coordinación de Integración del Transporte del Instituto Mexicano del Transporte, por los investigadores: Dr. Alfredo Bueno Solano, el Dr. J. Elías Jiménez Sánchez, el Dr. M. Gastón Cedillo Campos, el Mtro. Jared Piña Barcenás y la Mtra. Marisol Barrón Bastida.

Se destaca y agradece la colaboración del Mtro. Carlos Mario Pérez González, por su contribución al desarrollo del caso de estudio, y la dirección, revisión y aportaciones del Dr. Carlos D. Martner Peyrelongue, coordinador de Integración del Transporte en el Instituto Mexicano del Transporte, cuyas observaciones enriquecieron y dieron certeza a los análisis realizados.

Así mismo, se agradece la participación de José Alfonso Fernández López, Maribel Leyva Gaxiola y Cristina Uribe Pineda, estudiantes del programa de doctorado en Gestión Tecnológica de la Universidad Autónoma de Querétaro (UAQ), quienes participaron activamente en la elaboración de este proyecto.

Esta investigación es el producto final del proyecto de investigación interna *TI-16/17 Diseño conceptual del modelo de big data para el IMT*.

Finalmente, los investigadores agradecen el apoyo brindado por el Laboratorio Nacional de Sistemas de Transporte y Logística (SiT-LOG Lab), cuyos instrumentos permitieron la elaboración del presente documento, así como la difusión de sus resultados en el Congreso Internacional de Logística y Cadena de Suministro en su edición 2017.

# Contenido

---

Índice de figuras.....	v
Índice de tablas .....	vii
Sinopsis .....	ix
Abstract.....	xi
Resumen ejecutivo.....	xiii
Introducción.....	1
<b>2 Antecedentes de la gestión del conocimiento, la tecnología y <i>big data</i> .....</b>	<b>5</b>
2.1 Entorno de la organizacional actual.....	5
2.1.1 Gestión tecnológica.....	5
2.1.2 Gestión del conocimiento .....	7
2.1.3 Madurez tecnológica .....	9
2.2. <i>Big data</i> .....	11
2.2.1 Clasificación de los datos y datos abiertos.....	12
2.3 <i>Big data</i> en el sector de transporte y logística .....	14
<b>3 Modelos de referencia para el análisis de <i>big data</i> .....</b>	<b>17</b>
3.1 Análisis y ciencia del dato ( <i>big data</i> ).....	17
3.2 Tipos de análisis .....	19
3.3 Métodos de análisis .....	20
<b>4 Estudio y resultados .....</b>	<b>23</b>
4.1 Introducción .....	23
4.1.1 Objetivo general .....	23
4.1.2 Alcance .....	23

---

4.2 Almacenamiento y procesamiento de los datos.....	25
4.3 Análisis y ciencia del dato.....	27
4.3.1 Resumen estadístico de las medidas de confiabilidad.....	33
4.3.1.1 Coeficiente de variación.....	34
4.3.1.2 Índice de amortiguación ( <i>Buffer Index</i> ).....	34
4.3.1.3 Índice de tiempo planificado ( <i>Planning Time Index</i> ).....	35
4.3.2 Estadísticos de amplitud y sesgo.....	36
4.3.2.1 Estadísticos de sesgo ( $\lambda$ skew).....	36
4.3.2.2 Estadístico de amplitud ( $\lambda$ var).....	38
4.3.3 Análisis de confiabilidad.....	39
4.3.3.1 Índice de fluidez: A2069_2068.....	42
4.3.3.2 Índice de fluidez: A2053_2054.....	44
4.3.3.3. Índice de fluidez: A2060_2063.....	45
4.3 Conclusiones del caso práctico.....	47
<b>5. Conclusiones.....</b>	<b>51</b>
<b>Bibliografía.....</b>	<b>55</b>



# Índice de figuras

---

Figura 2.1 Fases del modelo de gestión de la información .....	11
Figura 2.2 Clasificación de tipo de datos.....	13
Figura 2.3 Factores para la implementación de <i>big data</i> .....	14
Figura 3.1 Diagrama de capas de la plataforma <i>big data</i> .....	18
Figura 3.2 Análisis y ciencia del dato .....	19



## Índice de tablas

---

Tabla 4.1 Puntos propuestos para la medición .....	24
Tabla 4.2 Información requerida para los puntos de medición .....	25
Tabla 4.3 Datos de casetas analizadas.....	27
Tabla 4.4 Información de los ajustes de los datos.....	29
Tabla 4.5 Resumen estadístico de los datos.....	32
Tabla 4.6 Percentiles de probabilidad .....	33
Tabla 4.7 Percentiles de los datos. ....	37
Tabla 4.8 Resumen estadístico de las medidas de confiabilidad (p.m.).....	40
Tabla 4.9 Resumen estadístico de las medidas de confiabilidad (p.m.) (continuación) .....	40
Tabla 4.10 Índices más altos del índice de fluidez .....	41
Tabla 4.11 Comparación de valores del índice de fluidez .....	48



## Sinopsis

---

En el presente estudio se elaboró una propuesta conceptual del modelo de *big data* para el desarrollo de investigaciones en materia de transporte desde el punto de vista del sector gubernamental. A través del análisis exhaustivo de la literatura en la materia y de consultas a expertos de los grupos de trabajo del Foro de Intercambio de Estadísticas de Transporte de América del Norte (ETAN) y el Foro Internacional del Transporte (FIT) de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), se desarrolló un esquema que pueda servir como base para la elaboración de herramientas que permitan a investigadores y responsables de la toma de decisiones a disminuir los tiempos de respuesta, haciendo uso eficiente de los recursos presupuestarios.



# **Abstract**

---

In the present paper, a conceptual proposal of the Big Data model was laid out for the development of transportation research from the point of view of the governmental sector. From an exhaustive analysis of literature and outcomes of the North American Transport Statistics Exchange (NATS) and the OECD International Transport Forum (ITF) work groups, this paper can be integrated to serve as a basis for the development of tools to support researchers and those responsible for response times, making efficient use of budgetary resources.



# Resumen ejecutivo

---

El presente estudio de investigación es resultado de una búsqueda bibliográfica exhaustiva de propuestas metodológicas para identificar modelos de gestión de estrategias de *big data*, que permitan derivar en un modelo conceptual para su operación en el Instituto Mexicano del Transporte. El objetivo se centra en la construcción de un modelo conceptual generalizable apropiado para el desarrollo de investigaciones futuras en materia de *big data* en el sector transporte.

El capítulo 2 presenta los antecedentes de la gestión del conocimiento, la gestión tecnológica y el *big data*. Partiendo de una serie de definiciones que sirven como base para solventar la propuesta de desarrollo de este proyecto.

En el capítulo 3 se describen los tipos de análisis que se pueden utilizar en un modelo de *big data*. Así mismo, se definen los métodos estadísticos, métodos de aprendizaje automático y análisis de gráficos y procesamiento de lenguaje natural.

En el capítulo 4 se pretende conocer y desarrollar el análisis de confiabilidad de los tiempos de tránsito en locaciones específicas del recorrido del autotransporte de carga, con la finalidad de analizar la variabilidad estacional en los tiempos de traslado en el recorrido de las principales rutas de carga del país.

Finalmente, se presenta una línea de trabajo futuro con la intención de mejorar el método a través de la inclusión de un mayor número de variables, que permita brindar mayor certeza a los administradores al momento de tomar decisiones.



# Introducción

---

La gestión y el análisis de datos confiables y suficientes pueden traducirse en beneficios y ventajas competitivas para las organizaciones de todos los sectores. Sin embargo, la obtención y el tratamiento de los datos representan uno de los retos más grandes para las organizaciones de todos los tamaños sin que este reto sea exclusivo del sector empresarial: otras áreas, como las organizaciones dedicadas a la investigación y desarrollo (I+D), requieren del acopio de datos útiles y fiables para poder construir modelos complejos y sofisticados que sirvan como herramientas para la toma de decisiones (Long y Talbot, 2015).

En este sentido, entender los elementos que permitan el diseño y la operación de un transporte eficiente, seguro y fluido, se vuelve cada vez más en un aspecto crítico para mejorar el desempeño de las cadenas de suministro nacionales y de exportación (Bueno -Solano y Cedillo- Campos, 2013), y por otro lado, la creciente complejidad que enfrentan las organizaciones para el acopio y posterior análisis de los datos dificulta el desarrollo de herramientas para la toma de decisiones efectiva.

Esto se debe a que las fuentes de información son muy diversas: en algunos casos encontramos datos estructurados que se encuentran almacenados en una base de datos tradicional, mientras que en otros casos, los datos incluyen documentos en diferentes formatos, como registros de servicio al cliente, hasta fotos o videos generados por sensores autónomos.

Si a esto se suman las nuevas fuentes de información -como las llamadas fuentes humanas que generan datos a partir de su interacción con los medios de comunicación social, por ejemplo, WAZE, así como los datos producto de la interacción con sitios web-, es posible tener una idea de la complejidad de su tratamiento.

Y si además se agregan otros factores como son la disponibilidad y la adopción de dispositivos móviles nuevos, potentes y con un acceso inmediato a la información, sin duda estrategias como *big data* fomentarán e impulsarán la creación de nuevas fuentes de datos capaces de contribuir a mejorar la posición competitiva de las regiones (Long y Talbot, 2015).

Sin duda, las nuevas tecnologías impulsan y contribuyen a su integración en foros. Por ejemplo, en 2016 el Foro Internacional del Transporte (FIT) de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), integró el primer grupo de trabajo en temas de *big data* y datos abiertos. Dentro de los hallazgos principales se identificó que en el sector gubernamental:

- Existe una alta presión por los tomadores de decisión por acceder de manera rápida y precisa a la información, pero como producto de estrategias derivadas de un manejo eficiente del presupuesto.
- La necesidad latente de generar reportes eficientes y rápidos de las condiciones del transporte y, además de ser capaces de manejar diferentes fuentes de información en tiempo real.
- La demanda creciente de analizar datos generados por fuentes humanas en grandes cantidades y que demandan repuestas o servicios que satisfagan sus necesidades de forma inmediata, exige incrementar la capacidad de estudiar los datos en tiempo real.

En dicha reunión participaron representantes de diversas nacionalidades como Francia, Canadá y Noruega, entre otros, quienes expusieron sus perspectivas y líneas de acción en materia de *big data*. Si bien los miembros coinciden en que la cantidad y la diversidad de datos generados a partir de la digitalización de servicios públicos y de transporte, entre otros, además del uso intenso de dispositivos móviles y el aumento del fácil acceso a las redes sociales, ofrecen la posibilidad de un cambio de paradigma en el cómo y con qué propósito, los datos pueden ser de gran utilidad, aunque identifican que actualmente no se tiene claro qué rumbo o modelo se debe seguir para explorar esta herramienta (Hu, 2016).

Uno de los objetivos específicos de este grupo es proponer y desarrollar proyectos de investigación de los cuales se puedan generar experiencias que evidencien las necesidades y problemáticas a las que se podría enfrentar la metodología a seguir, además de compartir experiencias con la finalidad de construir un modelo robusto que permita trabajar con múltiples fuentes de datos de forma segura, tanto para el generador de la información como para el responsable de su tratamiento.

Así mismo, como parte de los trabajos del XXXI Foro de Intercambio de Estadísticas de Transporte de América del Norte (ETAN), se ha propuesto abordar el tema desde una perspectiva de innovaciones estadísticas en el transporte, con la intención de definir las perspectivas y retos nacionales de Canadá, Estados Unidos y México en materia de *big data*.

En este contexto, la presente investigación permitirá contribuir a la consecución del objetivo 6 “Desarrollar integralmente y a largo plazo al sector con la creación y adaptación de tecnología y la generación de capacidades nacionales”, del Programa Sectorial de Comunicaciones y Transportes, mismo que se encuentra alineado con los objetivos 3.5 “Hacer del desarrollo científico, tecnológico y la innovación pilares para el progreso económico y social sostenible y, 4.9 “Contar con una infraestructura de transporte que se refleje en menores costos para la realización de la actividad económica” del Plan Nacional de Desarrollo 2013-2018.

Por ello, con el compromiso de estudiar las posibilidades de la herramienta de *big data*, el Instituto Mexicano del Transporte (IMT) se encuentra en la etapa de plantear

los siguientes proyectos que sirvan como base para explorar la técnica de macrodatos.

Caso 1. Análisis de la confiabilidad del tiempo de tránsito del transporte de carga por carretera.

Caso 2. Relaciones entre las características topológicas de la infraestructura de transporte de carga y la robustez de la red carretera nacional.

Caso 3. Análisis de las interacciones de cadenas de suministro.

Caso 4. Análisis del impacto de los accidentes carreteros en la cadena de suministros.

Si bien existe un interés en el desarrollo de todos los casos de estudio, es necesario, en esta primera etapa, definir un enfoque en la construcción de un modelo viable que, para efectos de una demostración práctica, se elegirá el caso 1, con base en que se ha identificado la importancia de estudiar datos a gran escala, para poder calcular y proponer acciones de mejora que permitan mejorar la confiabilidad del tiempo de viaje.

Sin duda, la confiabilidad es trascendente, ya que interviene en la decisión de dónde, cuándo y cómo realizar el viaje, además de influir sobre el tiempo extra utilizado para evitar retrasos o esperas, cuyo costo supera los tiempos promedio utilizados, y se traduce en mayores gastos (Margiotta, 2010). La necesidad de contar con indicadores para los sistemas de transporte surge de dos tendencias: Primero, la importancia del sistema de transporte de carga se ha incrementado en las últimas décadas tanto a nivel nacional como internacional, y segundo, la urgencia para la medición de la información generada al final de la cadena de suministro se ha intensificado (TRB, 2011).

Otro punto a favor en la elección del caso es que, en México el modo de transporte de mercancías más utilizado es el autotransporte por carretera, al mover éste el 81% de la carga terrestre, de la cual, el 56% corresponde a carga nacional. Esta diferencia se acentúa en el movimiento de pasajeros, donde se alcanza el 97% del total nacional. Así mismo, en 2016 el autotransporte tuvo una participación del 56.2% del total de toneladas transportadas, lo que equivale a 536 millones de toneladas (de acuerdo con las estadísticas básicas de autotransporte de carga 2016 publicadas por la Secretaría de Comunicaciones y Transportes de México).

Sin embargo, comprender el contexto del autotransporte y la integración con otros sectores de la sociedad no es una tarea trivial; por el contrario, representa un reto serio y complejo, sobre todo debido a la carencia de datos confiables y actualizados. Aquí es precisamente donde se plantea el concepto de confiabilidad y fluidez con la finalidad de evaluar el desempeño.

El objetivo de la investigación es formular una propuesta conceptual del modelo de *big data* apropiada para el desarrollo de investigaciones en materia de transporte desde el sector gubernamental, a partir de un análisis documental exhaustivo y consultas a expertos de los grupos de trabajo del ETAN y el FIT, para construir un esquema de referencia capaz de servir como base en la elaboración de herramientas que brinden a los tomadores de decisión alternativas de solución viables y competitivas para el sector.

Es importante mencionar que el documento presenta una propuesta conceptual para el desarrollo de un modelo de análisis con la herramienta de *big data* en las investigaciones donde se requiera el estudio masivo de datos con diferente estructura de recolección de datos, sin pretender por ahora convertirse en un estándar de aplicación en las investigaciones de transporte, sino simplemente ser una aportación que auxilie a los investigadores a lograr desarrollos en la materia.

## **2 Antecedentes de la gestión del conocimiento, la tecnología y *big data***

---

Se define como tomador de decisiones a la persona responsable de interpretar la información tácita y explícita de su entorno, con el objeto de conducir el destino de una organización. Para Dogson, Gann y Slater (2008), entre las funciones del tomador de decisiones se encuentra el uso de información confiable, con la cual diseña estrategias para fomentar la eficiencia y la capacidad de innovar, y al mismo tiempo, busca incrementar su competitividad dentro de un entorno sustentable.

### **2.1 Entorno de la organizacional actual**

La organización moderna es descrita por Castro (2003) como un sistema técnico y social, en donde los sistemas de información interactúan con las personas. Para llegar a esta definición, Castro (2003) analizó el concepto de organización desde diversos puntos de vista y autores: la organización, como un sistema de actividades humanas orientadas a la concreción de metas, es un enfoque sociológico proporcionado por Max Weber (1947), citado por Pugh y Hickson (1989); la organización como un entorno de relaciones de poder (Kahn, 1964; Jackson y Carter, 1991), y la organización como un lugar de “violencia controlada donde el poder involucra a todos quienes participan en ella” (Foucault 1997).

De acuerdo con estos enfoques, la organización actual es un entorno en donde sistemas de información procesan datos para incorporarlos a un proceso social. Este fenómeno se conoce como teoría de la estructuración de Giddens (1989), en donde existe un fenómeno de “recursividad” que permite la institucionalización de los procesos y sistemas de la organización, a fin de transformarlos en hábitos. Esta rutina busca regirse a partir de una serie de reglas, sistematizada y medida mediante el uso de la tecnología. Con lo anterior, es posible hablar del conocimiento desde dos dimensiones: el comportamiento técnico estructural y del conocimiento del comportamiento organizacional y de la intersección (Castañón, *et al.*, 2011).

#### **2.1.1 Gestión tecnológica**

El conocimiento es el elemento más importante en cualquier organización innovadora, por lo que es importante identificar cómo adquirirlo, representarlo, retenerlo y administrarlo (Zorrilla, 1997). El conocimiento consiste en verdades y creencias, perspectivas y conceptos, juicios y expectativas, metodologías y “know-how”, de acuerdo con (Macintosh 1997). La tecnología es un actor fundamental en la tarea de gestionar el conocimiento, al proporcionar componentes (hardware y software), procedimientos (metodologías y estándares) y medios (redes de comunicación) para preservarlo de forma digital, y con ello garantizar su prevalencia a lo largo del tiempo.

A menudo se piensa que la sola adquisición e implantación de elementos tecnológicos conlleva un proceso de gestión tecnológica. Sin embargo, cuando los tomadores de decisiones son incapaces de crear las condiciones para que el personal comprenda y utilice dichos elementos se genera desconcierto, incertidumbre y escepticismo. A este respecto, Zorrilla (1997) propone que el proceso de gestión tecnológica implique las siguientes actividades: i) prospección, ii) selección, iii) negociación, iv) adquisición, v) adaptación y vi) modificación e innovación.

El concepto de gestión tecnológica surge de la fusión de conceptos y métodos provenientes de la ingeniería y la administración, con el objetivo de organizar la producción y la aplicación industrial del conocimiento, particularmente en empresas de base tecnológica (EBT) (Solleiro, 1988).

La gestión tecnológica consiste, de acuerdo con el mismo autor, en el desarrollo científico de técnicas para entender y resolver una diversidad de problemas, tales como: la creación de escenarios, la prospección tecnológica; el uso eficiente de recursos públicos; el uso de los resultados provenientes de la investigación; la planeación y el control de proyectos; así como la legislación en la materia. Por su parte, Mejía F. (1998) plantea que la gestión tecnológica integra los diferentes procesos relacionados con el manejo de la tecnología en un marco administrativo eficaz.

En este sentido, Bernal y Landaverde (1995) proponen los siguientes procesos que hacen posible la gestión tecnológica al interior de una organización:

- **Gestión del conocimiento:** permite al personal de la empresa elaborar su visión de futuro a través del aprendizaje colectivo, la creación de modelos mentales y la disponibilidad de una visión internacional.
- **Administración de la información:** integra procesos sistemáticos, continuos, acumulativos y descentralizados, dentro de la empresa. Esto en cuanto a la búsqueda, selección y acopio de información, diseño de sistemas de información internos y para la toma de decisiones, así como a la generación interna y manejo participativo de la información.

- Administración de procesos productivos: se lleva a cabo a través de la creación de políticas de la empresa sobre copiado, asimilación, adaptación, mejoras, diseño y rediseño de tecnologías de materiales y materias primas, y tecnologías de equipo, operación, productos y procesos.
- Gerencia de procesos administrativos y directivos: es la responsable de crear procedimientos mediante los cuales la empresa conoce, negocia, maneja, evalúa y transfiere tecnologías. Además, es la encargada del monitoreo prospectivo tecnológico, la vigilancia de la propiedad industrial y los contratos de asistencia tecnológica y de servicios técnicos.
- Gestión tecnológica: abarca cualquier tipo de organización, aunque está mayormente asociada a centros de investigación, industria de la transformación y empresas de base tecnológica. Sin embargo, el sector educativo a nivel mundial ha experimentado un lento y gradual acercamiento a la adopción de prácticas que le permitan la gestión y la administración de la información. De acuerdo con Picciano (2012), las instituciones educativas son dirigidas conforme a un modelo de administración escolar, basado en la experiencia y el instinto de quienes las encabezan.

De acuerdo con Casassus (2000), el tema central de la teoría de la gestión es la comprensión e interpretación de los procesos de la acción humana en una organización. Históricamente se dice que la gestión consiste en una indagación acerca de las motivaciones de las personas en su lugar de trabajo y qué los puede impulsar a mejorar su desempeño.

## **2.1.2 Gestión del conocimiento**

Cuando una organización adopta procedimientos y prácticas para alcanzar algún grado de gestión del conocimiento, se espera que esta iniciativa repercuta en algún tipo de beneficio económico. Sin embargo, al respecto Prieto (2013) identifica dos tipos de visiones de parte de los investigadores del tema. Por un lado, Levitt y March (1988), Huber (1991) y Leonard (1992) manifiestan una posición neutral o negativa sobre la correlación entre la gestión del conocimiento y el beneficio económico. Y en contraparte, Fiol y Lyles (1985), Senge (1990) o Stewart (1997) afirman la existencia de una correlación positiva. Ambas posturas coinciden en reconocer al conocimiento y el aprendizaje como elementos generadores de mejores resultados organizacionales.

Cada organización, en función del contexto y necesidades respectivas, aborda la gestión del conocimiento desde un enfoque en particular. Prieto (2013) sugiere la siguiente clasificación al respecto:

- Gestión pasiva del conocimiento: está orientada a la obtención de mejores resultados financieros y muestra poco interés por la gestión del conocimiento.

- Gestión orientada al comportamiento: reconoce el conocimiento como un elemento inseparable de la operación y los empleados, debido a que operan en entornos complejos y en constante cambio. El conocimiento se comparte y se reinterpreta.
- Gestión del conocimiento técnico-estructurada: se basa en la interpretación de grandes cantidades de información y, con ello, trata de predecir e interpretar su entorno complejo. Utiliza la tecnología para acercar información a quien lo requiera y, por ello, su capacidad de respuesta debe ser rápida.
- Gestión del conocimiento activa: fusiona los enfoques técnico-estructurales orientados al comportamiento para la construcción de marcos de aprendizaje, el cual se ve fortalecido por la interacción de los componentes social y tecnológico.

Con base en la escala anterior, sería de esperar que aquella empresa con una gestión del conocimiento activa sea más lucrativa que aquella cuyo enfoque es pasivo. Sin embargo, Prieto (2013) reconoce la satisfacción de las necesidades de mercado como el único proceso que garantiza el crecimiento de la organización. La gestión del conocimiento y los procesos de aprendizaje ayudarán a mejorar la relación con los clientes, lo cual le permite a la organización ser sensible a las necesidades del mercado, aumentar el tiempo de respuesta y, con ello, obtener un mayor rendimiento económico y financiero.

El aprendizaje de una organización, de acuerdo con Bontis (1999) y Vera y Crossan (2000), citados por Prieto (2013), se clasifica en dos dimensiones: la dimensión estática, construida por el cúmulo de conocimientos tácitos y explícitos inmersos en los individuos de la organización, y una dimensión dinámica, la cual requiere de procesos de aprendizaje para la evolución de los cúmulos de información, a través de un ciclo de absorción, difusión y uso del conocimiento. En ambos casos, el cúmulo de conocimiento es la entrada y la salida del proceso de aprendizaje que hace posible “la exploración y la explotación de los conocimientos” (March 1991, citado por Prieto, 2013).

La adopción de un modelo de gestión del conocimiento activa por parte de una organización implicará dar solución a los problemas derivados del entendimiento de los conocimientos tácitos y explícitos en la organización. Estas problemáticas son clasificadas por Pérez *et al.* (2007) en:

- Espaciales: identifican el lugar donde reside el conocimiento dentro y fuera de la organización. El reto consiste en transferirlo y hacerlo tangible para la organización.
- Temporales: aseguran su prevalencia en el tiempo y permiten su uso generalizado tantas veces como sea necesario.

- Jerárquico-social: proviene de la cultura organizacional y la rigidez jerárquica, las cuales obstaculizan la transmisión, generación y aplicación del conocimiento.

Las problemáticas espaciales y temporales pueden reducirse con la adopción y el uso de procedimientos, estándares y tecnología, a pesar de la complejidad para abordar los problemas jerárquico-sociales, debido a la dificultad de estandarizar elementos provenientes de componentes culturales y del comportamiento de las personas de la organización.

### **2.1.3 Madurez tecnológica**

La gestión de la información se ha convertido en un problema complejo para las organizaciones, dado que este proceso es considerado como necesario para garantizar la supervivencia en el mercado. Esta gestión deriva en la reducción de costos y una mayor eficiencia, además de crear una ventaja competitiva. Para llevar a cabo una correcta gestión tecnológica, han surgido diversas metodologías, como acciones aisladas y fragmentadas, ante las cuales aún prevalecen las actividades de captura manual y de recaptura de datos que afectan tanto el flujo como la calidad de los datos (CMMI Institute, 2014).

En respuesta a esta problemática, el CMMI Institute publicó la primera edición del modelo Data Management Maturity DMM en agosto del 2014, resultado de tres años y medio de investigaciones respaldadas por expertos en el área de gestión de la información. Esta edición refleja la experiencia del CMMI Institute en el diseño de modelos, como el de Integración de Modelos de Madurez de Capacidades (Capability Maturity Model Integration, CMMI), el cual ha contribuido a mejorar las prácticas de la ingeniería de software para más de 10,000 empresas. Como resultado, el DMM puede ser utilizado y adaptado por cualquier tipo de organización.

De acuerdo con el CMMI Institute (2014), las áreas prácticas del DMM se pueden definir como sigue:

- Estrategia de gestión de información
  - Estrategia de gestión de información
  - Comunicación
  - Funciones de la gestión de información
  - Casos de negocio
  - Financiamiento
- Gobernanza de la información
  - Gestión de la gobernanza
  - Glosario del negocio
  - Gestión de los metadatos
- Calidad de la información

- Estrategia para la calidad de la información
- Perfiles de datos
- Aseguramiento de la calidad de la información
- Limpieza de datos
  
- Operación de información
  - Definición de los requisitos de información
  - Gestión del ciclo de vida de la información
  - Gestión de la proveeduría de información
  
- Plataformas y arquitecturas
  - Enfoque con respecto a la arquitectura
  - Estándares para la arquitectura
  - Plataforma para la gestión de información
  - Integración de información
  - Información histórica, archivos y retención
  
- Procesos de soporte
  - Mediciones y análisis
  - Gestión de procesos
  - Procesos de aseguramiento de la calidad
  - Gestión de riesgos
  - Gestión de la configuración

Las empresas que desarrollan e implementan estrategias para únicamente dar respuesta a los requerimientos de información reglamentarios corren el riesgo de contar con procesos incompletos que impliquen una gran cantidad de tiempo y esfuerzo. Para evitar esto, es necesario contar con un modelo de referencia que clarifique los objetivos más complejos en materia de gestión del riesgo.



Fuente: Elaborado por tesistas de la UAQ.

Figura 2.1 Fases del modelo de gestión de la información

## 2.2. Big data

Laudon y Laudon (2014) señalan que las empresas buscan conocimiento en grandes volúmenes de datos provenientes del tráfico en la web como de correos electrónicos y redes sociales. El término para describir este nuevo contexto de abundancia es el *big data*, el cual marca la era de los grandes conjuntos de datos, provenientes de diversas fuentes y potenciado por el uso de teléfonos móviles y redes sociales (Reips, 2015).

El Instituto Global McKensey lo define como el conjunto de datos cuyo tamaño sobrepasa la capacidad de las herramientas de software tradicionales para capturar, almacenar, administrar y analizar las bases de datos. En respuesta a las limitadas técnicas de gestión de datos ha surgido una nueva generación de tecnologías (Long y Siemens, 2011), nuevas herramientas de administración de datos que permiten a las empresas la captura, el almacenamiento y el análisis de la información.

Se han identificado tres factores principales que han logrado impulsar el crecimiento del *big data* (McAfee, A. y Brynjolfsson, E., 2012), a saber:

- Volumen de datos: En 2012 se generaron 2.5 exabytes de datos al día, un número que se duplicó cada 40 meses, siendo una cantidad de información impresionante que hace 20 años no había en el internet.
- Velocidad: Permite obtener datos en tiempo real, lo cual agiliza la toma de decisiones.
- Variedad: El surgimiento de nuevas tecnologías y su abaratamiento permite que mayor cantidad de personas las puedan adquirir, por lo tanto, se tienen fuentes de información muy diversas, por ejemplo, uso de GPS, redes sociales, teléfonos inteligentes y, sitios web, entre otros.

McAfee y Brynjolfsson (2012) mencionan que con ayuda del *big data* los gestores pueden medir y conocer más a detalle el negocio, lo cual permitirá aplicar dicho conocimiento y mejorar el rendimiento de la empresa con una toma de decisiones basada en evidencias y no en la intuición (decisiones basadas en datos).

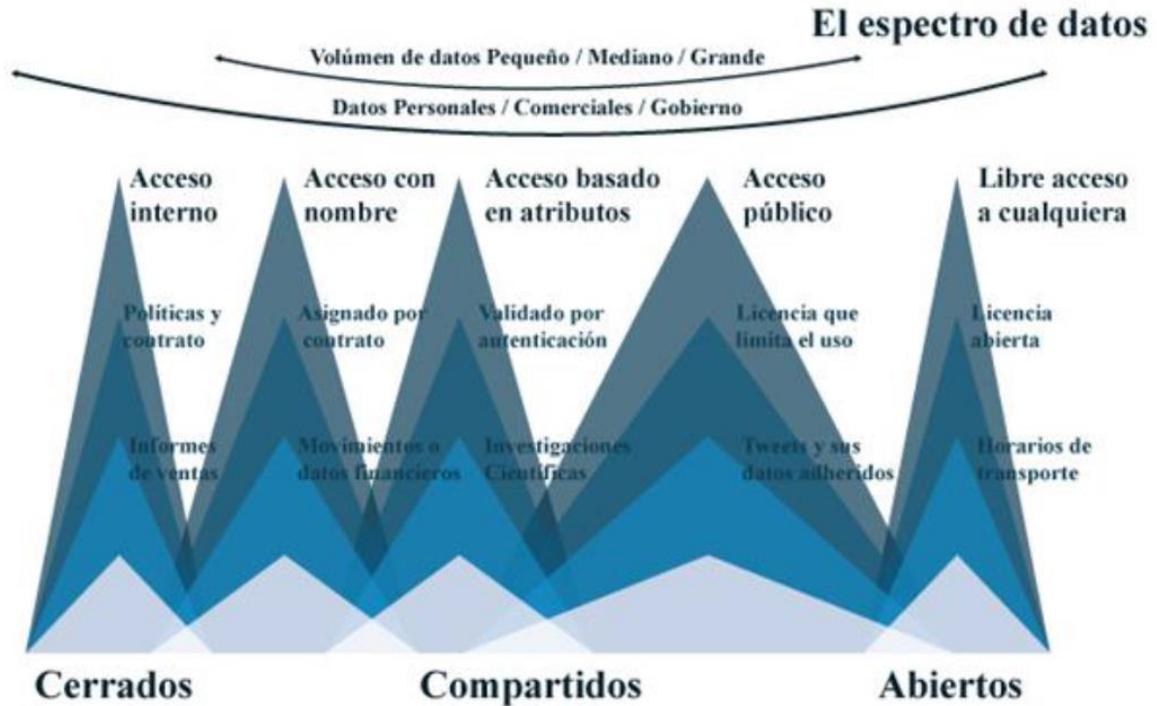
Conjuntamente se habla de la minería de datos; ésta se refiere a la localización de información útil en una extensa colección de datos (Mannila, 1996). Kamber (2005) lo define como “descubrir en conjuntos masivos de datos, patrones interesantes, significativos y comprensibles que se encontraban ocultos, para poder extraer conocimiento”.

## 2.2.1 Clasificación de los datos y datos abiertos

Los datos de entrada que alimentan el *big data* pueden ser datos cerrados, compartidos o abiertos, aunque existe otra clasificación aún más específica de los datos aportada por el Open Data Institute (ODI, s.f. y Broad, 2015), el cual clasifica los tipos de datos en cinco categorías (figura 2.2):

- Datos cerrados, de acceso interno: Son los datos más restrictivos, son privados o personales, es decir, solo tienen acceso a ellos ciertos miembros de una organización o determinadas personas especificadas bajo contrato o condiciones de uso.
- Datos de acceso autorizado: Son datos intermedios entre los cerrados y los compartidos, y se asignan exclusivamente bajo contrato.
- Acceso basado en grupos o comunidades: Estos datos no están disponibles para cualquier persona que lo desee, sino únicamente para grupos con características comunes y finalidad idéntica; se puede acceder a ellos a través de una autenticación.
- Acceso público, datos abiertos con condiciones: Datos accesibles para cualquier persona que lo desee, pero se limita el uso de los mismos.

- Datos abiertos: Puede acceder a ellos cualquier persona que lo desee, sin autorizaciones y para el fin que requiera.



Fuente: ODI (s.f.), Broad, (2015) y Cuantrix (s.f).

**Figura 2.2 Clasificación de tipo de datos**

En la guía de datos abiertos, Open Data Handbook (s.f.) se definen los datos abiertos como “aquellos que pueden utilizarse, reutilizarse y distribuirse libremente por cualquier persona, y que se encuentran sujetos al requerimiento de atribución y de compartirse de la misma manera en que aparecen”. Estos datos se definen por las siguientes características principales:

- Disponibilidad y acceso: La información debe estar disponible a un costo razonable y en una forma conveniente y modificable, de preferencia mediante descarga.
- Reutilización y redistribución: Los datos deben ser provistos bajo términos que permitan su reutilización, redistribución e integración con otros datos.
- Participación universal: Cualquier persona, sin discriminación alguna, debe poder utilizar, reutilizar y redistribuir la información.
- Interoperabilidad: Se refiere a la capacidad para trabajar en conjunto (interoperar); admite la integración de diferentes conjuntos de datos, es decir,

se pueden combinar distintas bases de datos, lo que permite construir sistemas más complejos y grandes.

## 2.3 Big data en el sector de transporte y logística

Las empresas de logística están entre las que más cantidad de datos generan y requieren al trabajar bajo la metodología *Just in Time*, por lo que obtener, manipular y estudiar estos datos permitirá optimizar el diseño de rutas, ahorros en gastos de operación, identificar tendencias y realizar predicciones más acertadas.

Existen diversos factores que se deben tomar en cuenta si se desea implementar un modelo de *big data* en cualquier sector.



Fuente: Elaboración propia con información de Vera (2015).

**Figura 2.3 Factores para la implementación de *big data***

Las fuentes de información en el sector logístico son diversas, y entre ellas se encuentran:

- Datos tradicionales de los sistemas de operación, como tiempos de tránsito, número de entregas, porcentajes de puntualidad, etc.
- Diagnóstico de vehículos, patrones de conducción y geolocalización. Por ejemplo, la instalación de GPS y sensores ayudará en la detección y prevención de averías; la reducción del consumo de combustible, y la geolocalización de vehículos.

- Datos de tráfico y clima, que al cruzarse con otros datos, ayudan en la toma de decisiones.



## 3 Modelos de referencia para el análisis de *big data*

---

El uso de *big data* en el sector logístico brinda numerosos beneficios, entre los que se encuentran:

- Mejora en los procesos: Se obtiene un mapa detallado de los procesos y permite un mayor control y anticipación de sucesos, así como la mejora de los activos al reducir inventarios, costos y tiempo. Además, contribuye en la mejora de la sostenibilidad del transporte al reducir combustible y emisiones de gases.
- Optimización del transporte: *El big data* con información proveniente de diversas fuentes como el tráfico, destinos, clima, hábitos de consumo, entre otros, permite generar información predictiva que ayuda en la optimización de las rutas (rutas inteligentes), una distribución más ágil y el seguimiento de los productos. De esta forma se reducen costos y tiempos.
- Gestión de información: El sector de transporte y logística genera gran cantidad de datos; sin embargo, el problema al que se enfrenta es la gran diversidad de los mismos, ya que cada entidad produce información de forma particular. La incorporación de un modelo de *big data* incluye un primer paso de obtención y homologación de los datos, lo cual permitirá la manipulación posterior, interpretación y generación de información para la toma de decisiones.
- Impulso de logística colaborativa: Facilita el uso de esta técnica al manipular mayor cantidad de datos en tiempo real, reduciendo traslados en vacío y aumentando la productividad.

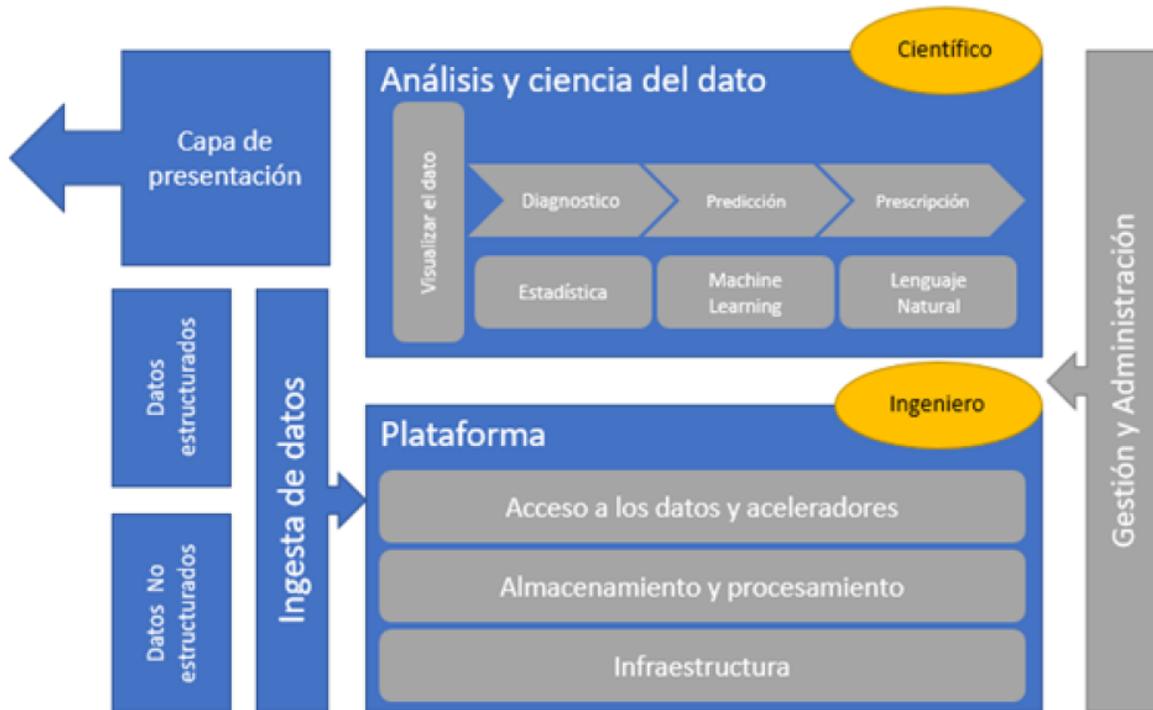
### 3.1 Análisis y ciencia del dato (*big data*)

Desde una perspectiva de análisis y ciencia del dato, se denomina *Big Data Analytics* o análisis de macrodatos al proceso de examinar grandes volúmenes de datos con la intención de descubrir patrones ocultos, correlaciones desconocidas y en general cualquier información que sea útil para llevar a cabo una mejor toma de decisiones, de acuerdo con SAS (2017).

Así mismo, esta tarea permite obtener valor a través de la reducción de costos; brindando a las organizaciones -a través de herramientas como Hadoop y el análisis basado en la nube-, la posibilidad de obtener ventajas del almacenamiento de datos

en volúmenes elevados, al permitirles identificar formas de hacer negocio de manera más eficiente.

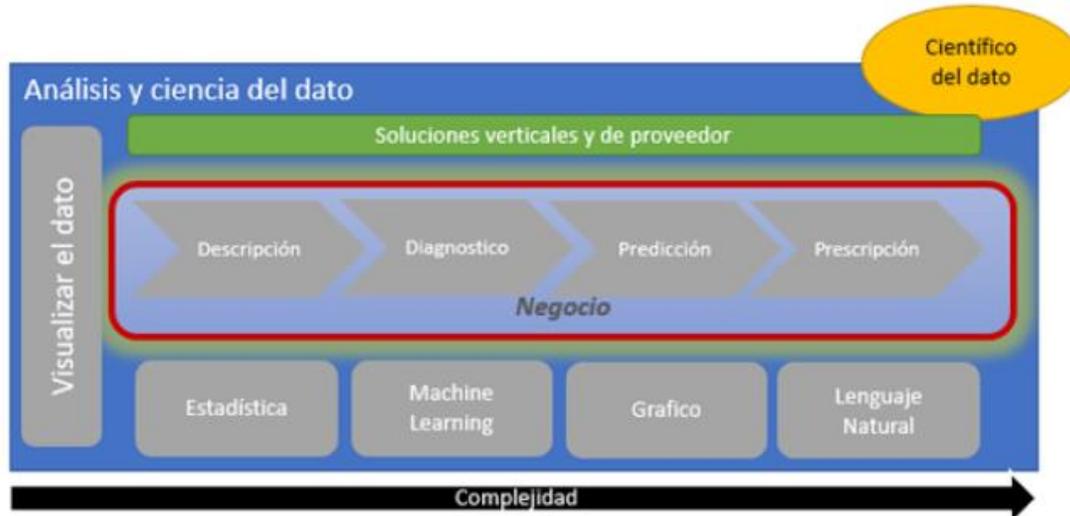
Además, el análisis de macrodatos contribuye a una más rápida y mejor forma de tomar decisiones, al disponer de la velocidad de dichas herramientas y la analítica en memoria. Adicional a la capacidad de analizar fuentes de datos diversas, las organizaciones pueden analizar la información inmediatamente y tomar decisiones basadas en lo que han aprendido. Posibilita la creación de nuevos productos y servicios, con la capacidad de medir las necesidades, gustos, y preferencias de los clientes. Y a través del análisis de grandes volúmenes de datos se hace cada vez más real la posibilidad de satisfacer los requerimientos de los clientes.



Fuente: Elaboración propia con base en Accenture (2016).

**Figura 3.1 Diagrama de capas de la plataforma *big data***

Una de las capas internas mostradas en la figura 3.1 corresponde al análisis y la ciencia del dato que, de acuerdo con Accenture (2016), internamente se compone de diversos elementos según se indica en la figura 3.2.



Fuente: Elaboración propia con base en Accenture (2016).

**Figura 3.2 Análisis y ciencia del dato**

## 3.2 Tipos de análisis

La principal razón de ser de esta capa es lograr explotar los datos, trasladarlos a información útil y visualizarlos de forma que sea posible transformarlos en información. Esta tarea se efectúa a través de cuatro tipos de análisis (Oracle, 2013).

- Análisis descriptivo: responde a la pregunta ¿qué está ocurriendo? Y a través de métodos estadísticos es posible resumir los datos, haciendo posible entregables, por ejemplo, reportes.
- Análisis de diagnóstico: responde a preguntas del tipo ¿qué ocurrió? Este análisis permite estudiar los datos para validar o descartar hipótesis de negocio, lo que da como resultado, entregables que profundizan en los datos mediante análisis estadísticos y factoriales.
- Análisis predictivo: busca responde preguntas ¿qué es probable que ocurra? A través de la construcción de modelos estadísticos que utilizan los datos existentes para predecir los datos que no se tienen, permitiendo entregables del tipo modelado de datos y predicciones.
- Análisis prescriptivo: responder a la pregunta ¿qué debería hacer al respecto? Con este análisis, se busca construir un modelo prescriptivo que utilice no sólo los datos existentes, sino también los datos de acción y retroalimentación para guiar al tomador de decisiones a un resultado deseado.

Cada uno de estos análisis son relativos al negocio en cuestión y regularmente se realizan mediante soluciones verticales o de proveedor (*Herramientas BI*) que están apuntaladas por diversos métodos de análisis, por ejemplo: estadístico, Machine Learning (aprendizaje automático) y el lenguaje natural, cada uno de los cuales eleva el nivel de complejidad del análisis.

### 3.3 Métodos de análisis

Por cuanto a los métodos de análisis y ciencia del dato, de acuerdo con la figura 3.2, estos se clasifican en cuatro grandes grupos: estadísticos, método de aprendizaje automático, análisis de gráficos y procesamiento de lenguaje natural, mismos que se describen a continuación:

- **Método estadístico.** Busca el estudio, recopilación, organización, análisis, interpretación y presentación de datos con el fin de incrementar el valor de negocio. La extracción de informes (presentación de datos) es una parte fundamental del esfuerzo sostenido hacia la mejora de la inteligencia de negocio y la gestión del conocimiento. Con ello se busca, por un lado, resumir los datos de la población, describiendo lo que se observó en la muestra (media, desviación estándar, frecuencia, porcentaje) y por otro, utilizar patrones en los datos de la muestra para extraer inferencias sobre la población representada (estimación, correlación, regresión u otro).
- **Método de aprendizaje automático (*Machine Learning*).** Es el campo de estudio que da a las computadoras la capacidad de aprender sin ser programadas explícitamente para ello (Arthur Samuel, 1959). El aprendizaje de la máquina es un tipo de inteligencia artificial que se refiere a la construcción y el estudio de sistemas que pueden aprender de los datos y hacer predicciones de manera autónoma. Se basa en cuatro tipos de algoritmos diferentes: árboles de decisión, bosques al azar, regresión logística y redes neuronales.
- **Método de análisis de grafos.** Los grafos representan objetos de datos y sus conexiones; éstos pertenecen a la ciencia de la aplicación de algoritmos para responder preguntas sobre cómo se traducen los objetos en gráficos. Entender las relaciones permite poseer un enfoque más poderoso para el análisis de *big data*, a diferencia del tradicional que está orientado estáticamente. El análisis de grafos se puede hacer en la gama completa de datos y está demostrado que es particularmente eficaz en el aumento de las clases de *big data*, tales como: contenido web, registros de sistemas de gran alcance, comunicaciones usando mensajería instantánea, correos electrónicos, tuits, registros de llamadas, imágenes, fotos y videos.
- **Método de procesamiento de lenguaje natural.** Es un campo de la informática, en donde la inteligencia artificial y la lingüística se ocupan de las interacciones entre las computadoras y las lenguas humanas (naturales). Se refiere al uso y la capacidad de los sistemas para procesar las oraciones en

un lenguaje natural, en lugar de en un lenguaje informático artificial especializado. El objetivo es permitir que las máquinas entiendan el lenguaje humano y extraigan significado del texto. Se utiliza comúnmente en el aprendizaje automatizado y la minería de datos.



## 4 Estudio y resultados

---

### 4.1 Introducción

La fluidez en la cadena de suministro se refiere a un amplio rango de medidas de desempeño, que incluye: la confiabilidad en los tiempos de viaje, los costos, la cantidad de carga movida, el análisis de riesgo y la resiliencia del sistema de transporte. La medición del tiempo de viaje mediante el estudio de confiabilidad es una de las partes principales del análisis de fluidez. El estudio de confiabilidad ayuda a localizar y analizar los trayectos críticos del sistema con el propósito de desarrollar medidas que reflejen su desempeño real, analizando la variabilidad y estableciendo nuevos parámetros, con el objetivo de determinar velocidades de operación en el trayecto de análisis.

#### 4.1.1 Objetivo general

El objetivo general de este estudio de caso es conocer y desarrollar el análisis de confiabilidad de los tiempos de tránsito, en locaciones específicas del recorrido del autotransporte de carga, con la finalidad de analizar la variabilidad estacional en los tiempos de traslado en el recorrido de las principales rutas de carga del país. Para ello, se realizará un análisis que permita mejorar la comprensión y análisis del tiempo de traslado; se tomarán en cuenta datos respectivos a diferentes condiciones de operación del tramo en análisis por día, mes y periodo; se conectarán pares origen-destino en un corredor de transporte del país, y además, se utilizarán la información relativa a la velocidad, flujo vehicular y tiempo promedio por tramo.

#### 4.1.2 Alcance

En el presente estudio se pretende realizar el cálculo de las medidas de confiabilidad en sistemas de transporte de carga, tomando datos en cada uno de los puntos de interés a lo largo de todo el día. El análisis nos permite obtener las siguientes medidas:

**Percentiles de la distribución de probabilidad:** Se obtuvieron los percentiles del 95, 90, 75, 50, 25 y 10%. Las medidas representan un porcentaje del tiempo de viaje promedio.

**Coefficiente de variación:** Representa una medida de la dispersión de los tiempos de viaje.

**Índice de tiempo planificado (PTI, por sus siglas en inglés):** Es un índice para comparar la duración del viaje con el tiempo de viaje con flujo libre.

**Índice de holgura o amortiguación:** Se define como el tiempo extra expresado en minutos que necesita planearse con respecto al tiempo medio para asegurar llegar a tiempo el 95% de las ocasiones. La estadística está diseñada para tener en cuenta cualquier retraso inesperado.

**Parámetros de sesgo y amplitud:** El estadístico de sesgo permite cuantificar la magnitud y dirección del sesgo de los tiempos de viaje. El estadístico de amplitud nos permite saber la proporción del rango de valores que caen dentro del 80% alrededor de la media y la mediana del tiempo de viaje.

Para desarrollar el estudio de caso, los datos necesarios para el análisis de confiabilidad fueron: los tiempos de viaje entre un punto inicial y uno final, así como sus puntos intermedios. Para este ejercicio se establecieron como puntos de recolección aquellos utilizados por la Dirección General de Servicios Técnicos en su análisis de datos viales. De esta forma se utilizó la información relativa al tránsito diario promedio, así como la información del tiempo de traslado entre puntos, obteniendo mayores beneficios y facilidades para el análisis. Los puntos propuestos para la medición de 24 horas del 1 de enero al 31 de diciembre de 2016 se muestran en la tabla siguiente:

**Tabla 4.1 Puntos propuestos para la medición**

Nombre de puntos propuestos	Km	Latitud	Longitud
Caseta de cobro Tepetzotlán	43.01	19.715358	-99.207587
Ent. Jorobas	56.18	19.8229	-99.24619
Ent. Jorobas	56.18	19.823	-99.24625
Ent. Jorobas	56.18	19.82881	-99.257
T. Der. Tepeji del Río	69.36	19.87584	-99.3376
T. Der. Tepeji del Río	69.36	19.914617	-99.362453
T. Izq. Jilotepec	83.75	19.9758	-99.43607
T. Izq. Jilotepec	83.75	19.98567	-99.44364
X. C. Jilotepec – Maravillas	107	20.08084	-99.64616
X. C. Jilotepec – Maravillas	107	20.08266	-99.62512
T. Izq. Aculco	125	20.09747	-99.69294
T. Izq. Aculco	125	20.11698	-99.71785
Caseta de cobro Palmillas	145.8	20.296103	-99.928994

T. Izq. Toluca	148.5	20.29208	-99.92175
T. Izq. Toluca	148.5	20.302808	-99.93481
T. Izq. Toluca	148.5	20.30305	-99.935184
T. Der. Poniente San Juan del Río	164.1	20.37269	-99.99158
T. Izq. Acámbaro	170.45	20.44693	-100.07836
T. Der. Libramiento de Querétaro a San Luis Potosí	207	20.578879	-100.378216
T. Der. Libramiento de Querétaro a San Luis Potosí	207	20.584826	-100.343115

Fuente: Elaboración propia.

Para lo cual, se requiere de la siguiente información:

**Tabla 4.2 Información requerida para los puntos de medición**

Información	Descripción
<b>Km</b>	Kilómetro del punto generador antes referido.
<b>Latitud</b>	Coordenadas geográficas del lugar del aforo.
<b>Longitud</b>	Coordenadas geográficas del lugar del aforo.
<b>Velocidad</b>	Velocidad promedio, en los puntos de recolección de datos por hora.
<b>Flujo vehicular</b>	Tránsito promedio por hora.
<b>Longitud del tramo</b>	Longitud de punto a punto de recolección de datos.
<b>Hora</b>	Horario de recolección de datos (día:hora:min)
<b>Tiempo</b>	Tiempo promedio para trasladarse de un punto de recolección al siguiente, por hora del día.

Fuente: Elaboración propia.

## 4.2 Almacenamiento y procesamiento de los datos

Los datos utilizados en el presente estudio de caso se obtuvieron de cuatro empresas dedicadas a prestar servicios de transporte de carga por carretera. Los

datos provienen de dispositivos GPS de diversas compañías que prestan servicios de monitoreo de flotillas a dichas empresas, por lo que no se tiene un acceso directo a las bases de datos correspondientes.

Los datos recabados son de tipo GPS y la información se solicitó de tal manera que su contenido no vulnerará la seguridad de la empresa, por lo que solo se incluyeron los siguientes campos: un identificador del vehículo generador; la fecha y hora en que se generó la información, y por último la latitud y longitud de la posición.

A pesar de provenir de fuentes de datos estructurados, los datos se entregaron en diversos formatos, tanto de los tipos de archivos, como de su estructura y presentación. Por ejemplo, los archivos entregados van desde el tipo de copia de seguridad de base de datos SQL Server (extensión bak), libros de Excel (extensiones xlsx y xls) o archivos CSV (del inglés: *comma-separated values*). Cada compilado de información podía encontrarse en un solo archivo o dividido en varios y presentar distinto número de columnas de información, solo coincidiendo en los campos solicitados. Así, entre las columnas de información extra se encuentran datos de la velocidad del vehículo, dirección postal, día de la semana y odómetro, entre otros. También el formato de las columnas puede variar entre las empresas, a pesar de tratarse de la misma información, por ejemplo, la fecha puede encontrarse en formato de texto en un archivo y en otro encontrarse en un formato de fecha correcto. Dada la diversidad de formatos en que se presentan los datos se pueden considerar como semiestructurados.

Para el proceso de ingesta de los datos, se presenta una gran oportunidad para la elaboración de una herramienta que automatice este proceso. Las copias de seguridad de bases de datos son las más sencillas de ingresar, dado que existe una cantidad de herramientas y documentos donde se explica el proceso de migración de datos entre distintos sistemas de gestión de bases de datos.

En cuanto al proceso de ingesta de los datos provenientes de libros de Excel y archivos CSV, se desarrolló una macro que se puede adaptar a cada caso particular, lo que permite de esta manera concentrar, limpiar y dar el formato correcto a los datos de las distintas empresas. Para el almacenamiento de los datos se empleó el sistema de gestión de bases de datos PostgreSQL y su módulo PostGIS para el soporte de objetos geográficos. Así, se almacenó un total de 84,086,714 puntos provenientes de 797 dispositivos GPS diferentes.

A objeto de preparar la información para el análisis de confiabilidad de los tiempos de viaje, se usaron como puntos de referencia las coordenadas de las estaciones de aforo definidas en el Libro de Datos Viales de la DGST.

El proceso para la obtención de los nuevos datos para el análisis de confiabilidad es el siguiente:

1. Se toma una carretera definida de acuerdo con la información publicada por la DGST.

2. Se toman todos los puntos GPS que pasan por la carretera.
3. De dichos puntos se determinan los diferentes dispositivos GPS a los que pertenecen.
4. También se toman los distintos puntos de aforo que pertenecen a dicha carretera. Estas estaciones dividen cada tramo carretero en segmentos.
5. Se toma uno a uno los dispositivos, se determina el primer punto GPS correspondiente y se sigue la ruta transitada por el vehículo en la carretera seleccionada.
6. Conforme se avanza en la ruta, se va verificando el cruce con alguna estación de aforo, y si es el caso, se toman los dos puntos más cercanos que se encuentren en lados opuestos respecto a la estación de aforo.
7. Con dichos puntos GPS, se calcula el tiempo estimado en que el vehículo debió cruzar por dicha estación de aforo (se calculan distancias, tiempos y velocidades entre las tres posiciones involucradas).
8. Se repiten los pasos del 1 al 7 por cada dispositivo diferente que circuló por la carretera seleccionada.
9. Se repiten los pasos del 1 al 8 por cada carretera definida por la DGST y de interés para el estudio.

### 4.3 Análisis y ciencia del dato

Los datos de las casetas utilizadas se muestran en la siguiente tabla:

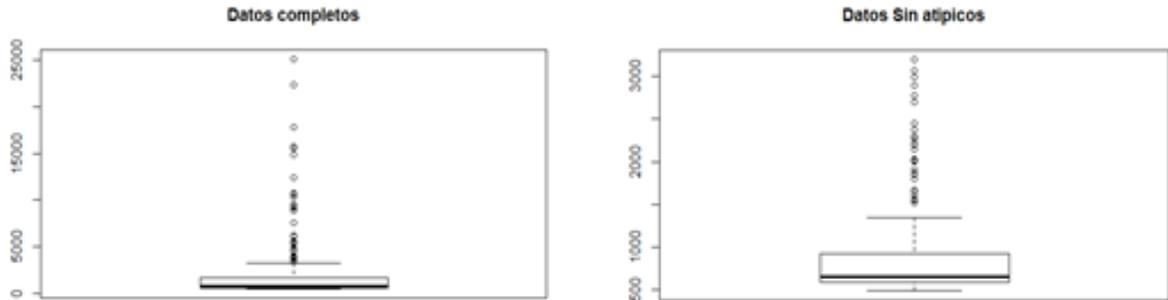
**Tabla 4.3 Datos de casetas analizadas**

	Nombre	Km	Latitud	Longitud	Id datos	Vel. diseño
<b>Punto inicial</b>	Tepotzotlán	43.01	-99.207587	19.715358	2053	110
<b>Punto final</b>	Ent. Jorobas	56.18	-99.24619	19.8229	2054	110

Fuente: Elaboración propia.

Se analizaron en total 221 datos correspondientes a viajes que comenzaron en la caseta de Tepotzotlán en el km 43.01 de la autopista México – Querétaro, identificada con el número 2053, y como punto final de análisis en la caseta Jorobas, en el km 56.18, identificada con el número 2054. El análisis de datos se realizó con el software de análisis estadístico R y se validaron en Minitab-16.

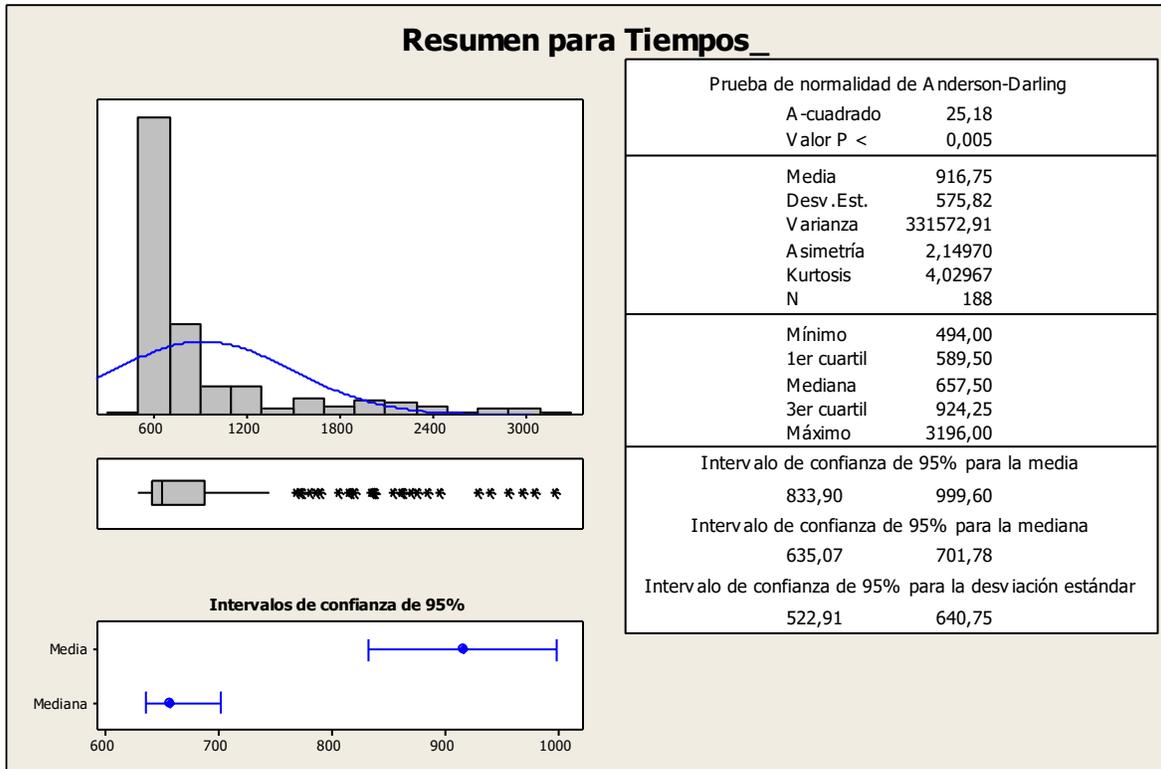
En lo sucesivo se muestran los resultados obtenidos por Minitab-16 y, en los casos que así se requiera, se presentan gráficas elaboradas a través de R. De un total de 221 datos, se elaboraron los gráficos de caja para los datos completos, así como para los datos después de la eliminación de datos atípicos, los cuales después de la eliminación fueron un total de 188 datos (ver grafica 4.1).



Fuente: Elaboración propia

**Gráfica 4.1 Diagrama de caja de los datos obtenidos antes y después de la eliminación de datos atípicos.**

Un análisis general de los datos arrojó el siguiente resumen estadístico básico que se presenta en la gráfica 4.2.



Fuente: Elaboración propia.

### Gráfica 4.2 Resumen estadístico básico.

El primer paso consistió en adaptar el conjunto de datos a una distribución de probabilidad, sin embargo, no se encontró ningún ajuste adecuado de todas las distribuciones utilizadas dentro de Minitab-16. Todos los estadísticos P (nivel de significancia) son menores que un nivel de significancia alfa = 0.05. La información detallada de los ajustes se presenta en la siguiente tabla.

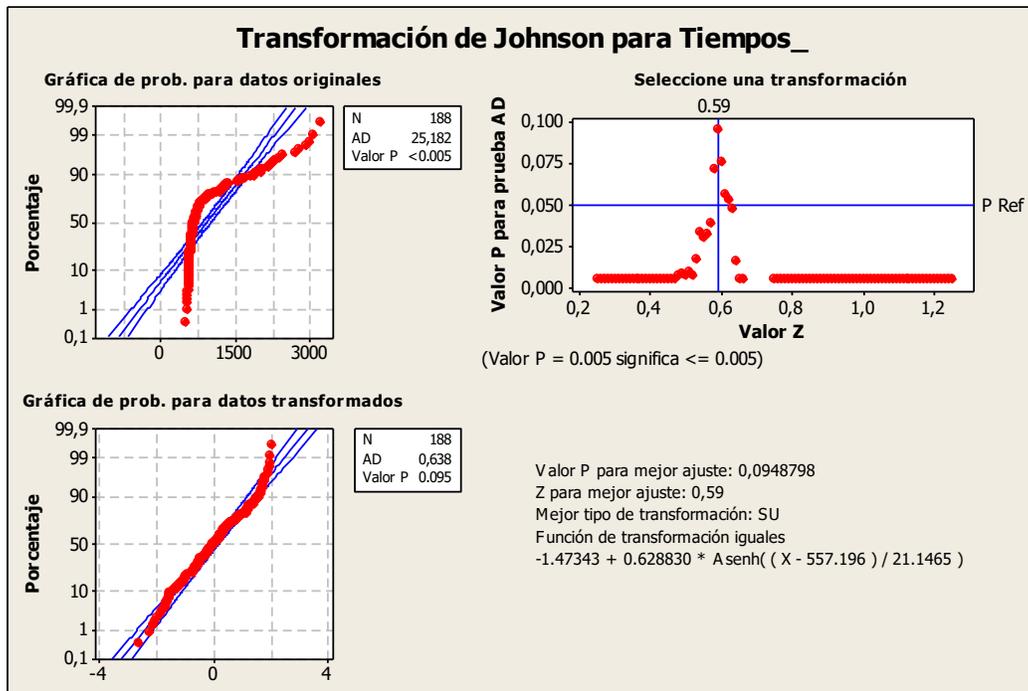
**Tabla 4.4 Información de los ajustes de los datos**

Distribución	AD	P	LRT P
<b>Normal</b>	25.183	<0.005	

<b>Transformación Box-Cox</b>	5.080	<0.005	
<b>Log-normal</b>	16.050	<0.005	
<b>Log-normal de tres parámetros</b>	4.099	*	0,000
<b>Exponencial</b>	32,498	<0,003	
<b>Exponencial de dos parámetros</b>	12.335	<0,010	0,000
<b>Weibull</b>	20.137	<0,010	
<b>Weibull de tres parámetros</b>	8.014	<0.005	0,000
<b>Valor extremo más pequeño</b>	29.406	<0,010	
<b>Valor extremo más grande</b>	19.259	<0,010	
<b>Gamma</b>	19.265	<0,005	
<b>Gamma de tres parámetros</b>	9.832	*	0,000
<b>Logística</b>	20.614	<0,005	
<b>Log logística</b>	13.138	<0,005	
<b>Log logística de tres parámetros</b>	3.382	*	0,000

Fuente: Elaboración propia.

Debido a que las transformaciones de variables sencillas no fueron satisfactorias, se procedió con la transformación de Johnson, cuya función de transformación se muestra en la gráfica siguiente.

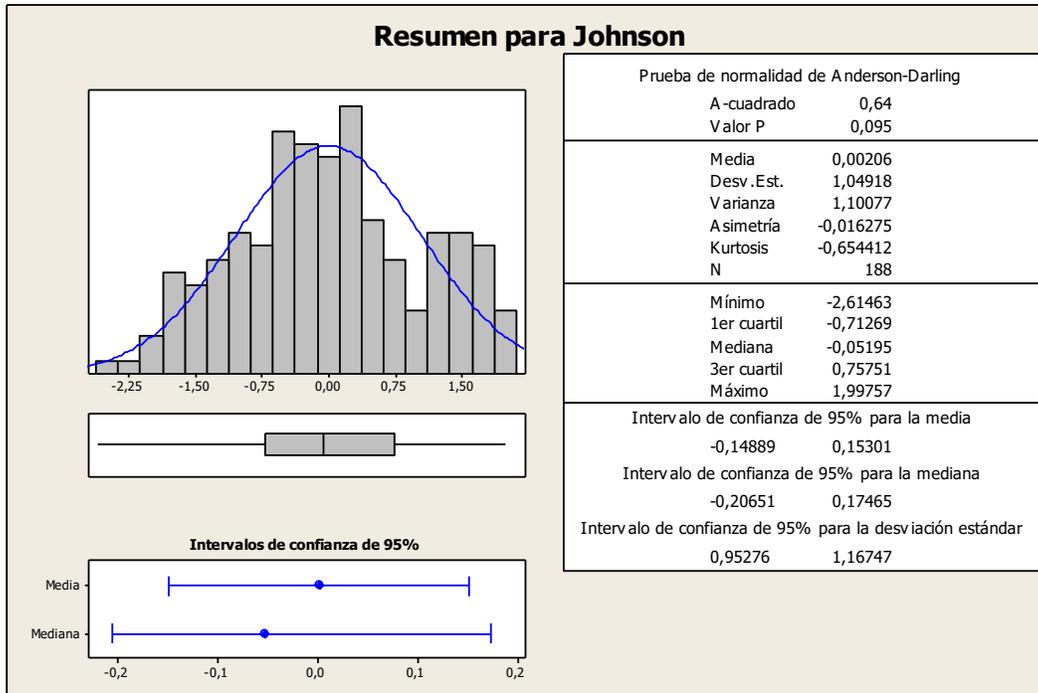


Fuente: Elaboración propia.

### Gráfica 4.3 Transformación de Johnson.

Se observa que el estadístico de prueba P presenta un valor de 0.095, el cual es mayor que el 0.05 requerido para aceptar la hipótesis, la cual no rechaza la distribución normal como función de distribución de los datos de viajes.

Las características del conjunto de datos después de la transformación de Johnson son los siguientes (ver grafica 4.4):



Fuente: Elaboración propia.

**Gráfica 4.4 Resumen estadístico básico después de la transformación de Johnson**

Al convertir los datos transformados para la realización de los cálculos, se despejó la variable de la función de transformación arrojada por Minitab-16. A continuación, se presenta la función original de transformación obtenida mediante Minitab-16 seguida de la función resultante del despeje de los valores x originales, la cual se utilizó para obtener los valores de los percentiles de los datos originales.

$$f(x) = f. \text{ gamma} + f. \text{ eta} * \text{Asinh} \left( \frac{x - f. \text{ epsilon}}{f. \text{ lambda}} \right)$$

$$x = (f. \text{ gamma}) \left( \text{Sinh} \left( \frac{(x + (-1)) * f. \text{ gamma}}{f. \text{ eta}} \right) \right) + f. \text{ epsilon}$$

El resumen estadístico básico para los datos de tiempos realizando la transformación inversa de Johnson, se presenta a continuación.

**Tabla 4.5 Resumen estadístico de los datos**

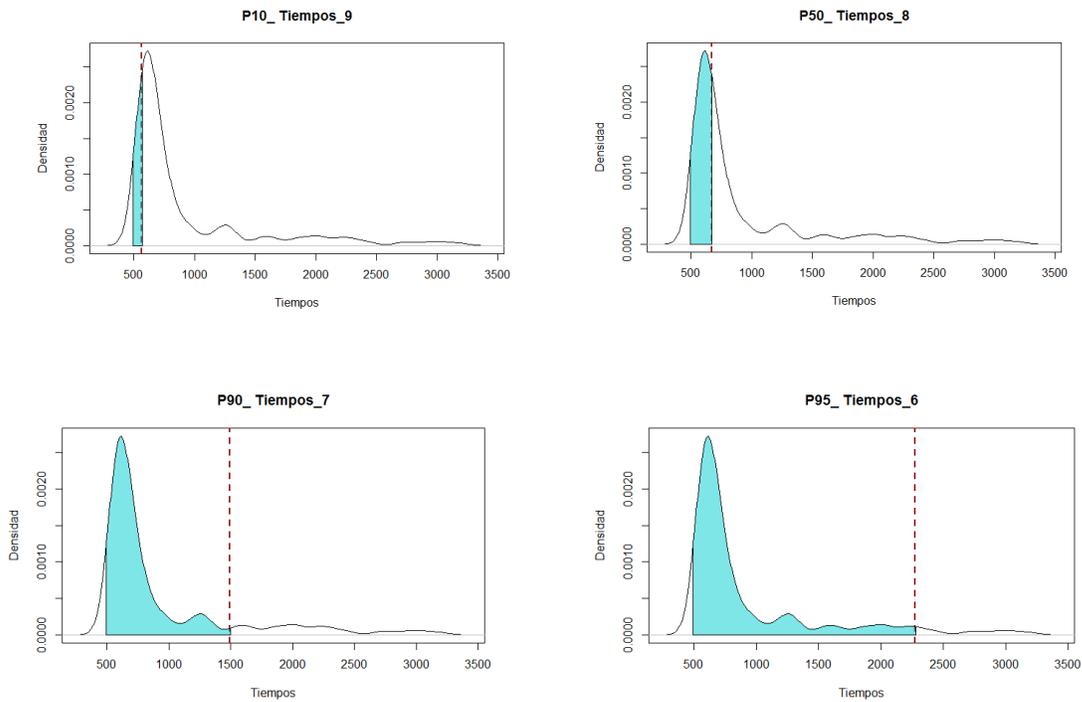
Medida	Min	1er Qu.	Mediana	Media	3er Qu.	Max	Ds	Var
--------	-----	---------	---------	-------	---------	-----	----	-----

<b>Valor</b>	494	589.92	657.47	666.65	897.25	3196	1141.01	1190.96
--------------	-----	--------	--------	--------	--------	------	---------	---------

Fuente: Elaboración propia.

### 4.3.1 Resumen estadístico de las medidas de confiabilidad

A continuación, se presenta una representación gráfica de los percentiles involucrados en el cálculo de la probabilidad del 95%, 90%, 50% y 10%.



Fuente: Elaboración propia.

**Gráfica 4.5 Percentiles de probabilidad**

**Tabla 4.6 Percentiles de probabilidad**

Percentil	P10	P50	P90	P95	P>95
<b>Valor</b>	561.63	666.65	1494.32	2275.45	2750.22

### 4.3.1.1 Coeficiente de variación

El tiempo de viaje debe estimarse como una variable, aleatoria denominada coeficiente de variación, la cual se describe como la relación entre la desviación estándar y la media de los datos de tiempos de viaje (Cedillo -Campos *et al.* 2017).

Medida	Media	Ds
Valor	666.65	1141.01

$$\text{Coeficiente de variación} = \left( \frac{\text{Desviación estándar}}{\text{Media}} \right) * 100 \quad (1)$$

$$\text{Coeficiente de variación} = \left( \frac{1,141.01}{666.65} \right) * 100 = 171.15\%$$

Ésta es una medida de la variabilidad dentro del conjunto de datos. Un mayor valor del coeficiente, significa que el conjunto de datos presenta mayor heterogeneidad, es decir que existe una variedad de datos con características diferentes con respecto a la media. Para el estadístico antes calculado, se tiene que existe una gran variabilidad en los valores de la variable aleatoria que representa la muestra de los tiempos de viaje.

### 4.3.1.2 índice de amortiguación (*Buffer Index*)

De acuerdo con información de la Administración Federal de Carreteras (*Federal Highway Administration, FHWA*) de Estados Unidos, el índice de amortiguación se expresa como un porcentaje y su valor aumenta a medida que empeora la confiabilidad.

$$\text{Índice de amortiguación} = \left( \frac{\text{Percentil del 95\%} - \text{Media}}{\text{Media}} \right) \quad (2)$$

$$\text{Índice de amortiguación} = \left( \frac{2,275 - 666.65}{666.65} \right) * 100 = 241.3246\%$$

La cifra anterior significa que, para un tiempo promedio de 666.6531 segundos u 11.11 minutos (666.6531 x 2.4132 = 1608.76 o 26.8127 minutos), el viajero debe presupuestar otros 2.41% (1608.76 segundos) para asegurar la llegada puntual la mayor cantidad de ocasiones. En este ejemplo, los 26.81 minutos se denominan:

tiempo *buffer* o tiempo de amortiguación, el tiempo necesario para asegurar llegar a tiempo el 95% de las ocasiones.

Tiempo planificado = P95%

Tiempo planificado = 2,275.451

Percentil	P95
Valor	2,275.451

#### 4.3.1.3 Índice de tiempo planificado (*Planning Time Index*)

La ruta en análisis es la siguiente:

- Caseta de Tepetzotlán km 43 (2053) – Caseta Jorobas km 56 (2054)

Se analizaron en total 188 datos correspondientes a viajes que comienzan en la caseta de Tepetzotlán en el km 43.01 de la autopista México-Querétaro, identificada con el número 2053, y con punto final de análisis en la caseta Jorobas perteneciente a la autopista antes mencionada en el km 56.18, identificada con el número 2054.

Con base en las distancias calculadas mediante el paquete “geosphere” cargado en R, se obtiene una distancia de caseta a caseta de 12.57365 km (como referencia se menciona que un tiempo estimado de llegada por Google Maps es de 13 minutos). El tiempo de flujo libre se calcula utilizando la velocidad de diseño de la carretera.

Para el caso de este estudio, la velocidad de diseño del segmento de la carretera Federal 57 en este segmento es de 110 km/h y se supone una velocidad de operación máxima de las unidades en análisis de 90 km/h, razón por la cual, se utilizará tal velocidad para el cálculo del Índice de tiempo de viaje. Suponiendo un flujo libre donde el usuario no se encuentra con intervenciones en su viaje, se obtiene el siguiente tiempo:

$$\text{Tiempo de viaje con flujo libre} = \frac{12.57 \text{ km}}{90 \text{ km/h}} = 502.946 \text{ seg} = 8.38 \text{ min}$$

El índice de tiempo planificado compara el peor de los casos con el tiempo de viaje con flujo libre, de acuerdo con la FHWA:

$$\text{Índice de tiempo planificado} = \frac{\text{Percentil del 95\%}}{\text{Tiempo de viaje con flujo libre}} \quad (3)$$

$$\text{Índice de tiempo planificado} = \frac{2\,275.45}{502.94} = 4.52$$

El índice de tiempo planificado, es un factor aplicado al tiempo de viaje con flujo libre, se distingue del índice de amortiguación (*Buffer Index*) debido a que incorpora los retrasos no habituales. El índice anterior indica que un viajero tendrá que planear un viaje total de  $4.524244 * 502.946 = 2275.451$  segundos (37.92418 min), con respecto al tiempo de viaje en flujo libre para asegurar llegar a tiempo.

### 4.3.2 Estadísticos de amplitud y sesgo

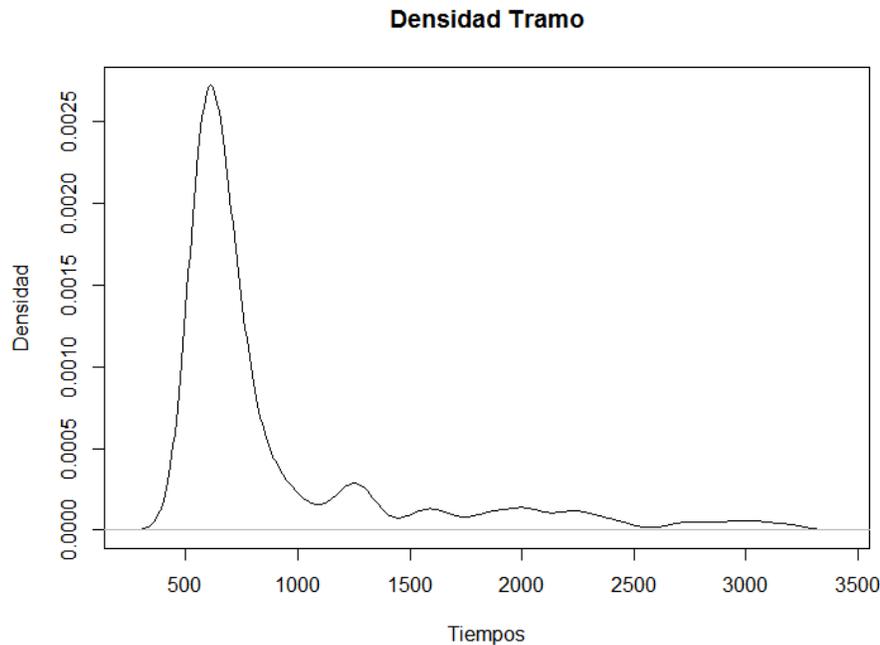
Las suposiciones de normalidad en los datos es un práctica común; sin embargo, a pesar de ser la estrategia de confiabilidad más utilizada, ha recibido varias críticas sobre la suposiciones de una distribución simétrica. La evidencia empírica muestra que las distribuciones simétricas son más frecuentes en las condiciones de flujo libre, de igual forma, la desviación estándar del conjunto de datos es severamente afectada por los valores atípicos. De acuerdo con Van Lint y Van Zuylen (2014), se propone considerar el sesgo y la amplitud de la distribución del conjunto de datos para calcular la confiabilidad. A continuación, se presentan ambos índices.

#### 4.3.2.1 Estadísticos de sesgo ( $\lambda$ skew)

El estadístico de sesgo, es una medida de la asimetría en la distribución del conjunto de datos, indicando de esta forma el nivel y dirección del sesgo de la distribución. En el caso de valores altos del estadístico, éstos indican que existe una mayor posibilidad de encontrar valores mayores a la mediana de la distribución, además de revelar que la forma de la distribución está fuertemente sesgada hacia la izquierda, lo cual debe interpretarse como poca confiabilidad en el tiempo de viaje, al indicar que al menos el 10% de los tiempos será significativamente mayor que la mediana. Valores superiores a 1 significan que los viajeros considerados como desafortunados por tener tiempos de viaje altos perderán más tiempo que los viajeros con tiempos de viaje cercanos a la mediana. Valores altos también indican que una alta proporción de la confiabilidad se debe a eventos no recurrentes, tales como incidentes, zonas de trabajo, clima o acontecimientos extraordinarios. Por el contrario, valores muy pequeños indican que la distribución está fuertemente sesgada hacia la derecha. Si el estadístico es igual a 1 la distribución es simétrica y, por lo tanto, se debe recurrir al estadístico de amplitud para generar conclusiones respecto a su confiabilidad.

Es importante hacer una aclaración en este punto. Este estadístico representa la asimetría de la concentración de los datos con respecto a la mediana de la distribución. Es por ello que se representa como la razón de la mitad superior de los datos entre la mitad inferior de los mismos. Dependiendo de su valor representará la dirección de la acumulación de los datos en la parte derecha o izquierda de la distribución. No está relacionado con el estadístico de sesgo comúnmente descrito

en estadística, el cual puede tomar valores positivos o negativos, dependiendo de la dirección del sesgo.



Fuente: Elaboración propia.

**Gráfica 4.6 Densidad de probabilidad de los datos**

**Tabla 4.7 Percentiles de los datos.**

Percentil	P10	P50	P90	P95	P>95
<b>Valor</b>	561.6308	666.6531	1494.325	2275.451	2750.222

$$\lambda_{skew} = \frac{P90 - P50}{P50 - P10} \quad (4)$$

$$\lambda_{skew} = \frac{1494.32 - 666.65}{666.65 - 561.63} = \frac{827.67}{105.02} = 7.80$$

El estadístico de sesgo mayor que 1 indica que la distribución está sesgada hacia la izquierda (esto se puede observar en la gráfica 4.6, la cual nos muestra gráfica de densidad del conjunto de datos). De acuerdo con Van Lint y Van Zuylen (2014), valores altos en el estadístico de sesgo deben interpretarse como poco confiables; esto implica que al menos en el 10% de los casos los tiempos de viaje son

significativamente más largos que la media. Estos valores son los presentes en la cola superior de la distribución, la cual podemos observar en la gráfica de densidad 4.6.

#### **4.3.2.2 Estadístico de amplitud ( $\lambda$ var)**

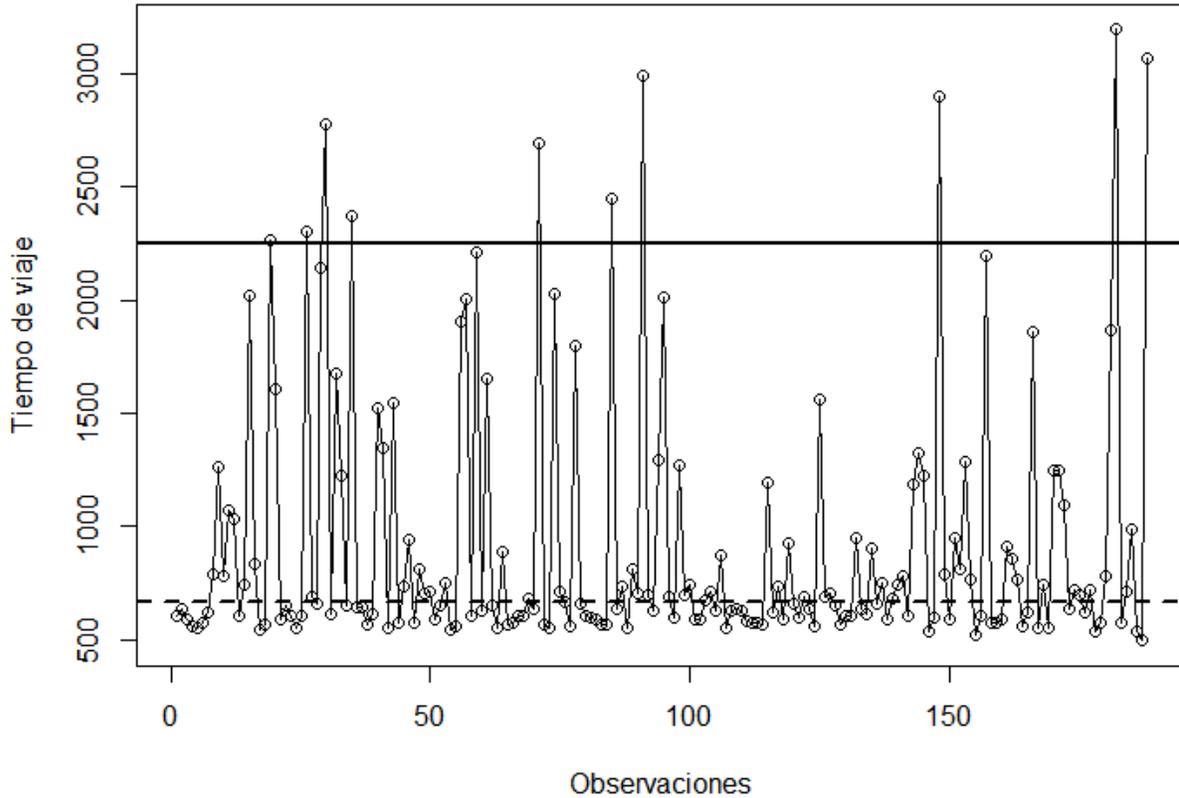
Esta medida representa la relación donde el 80% de las observaciones caen dentro del rango de la mediana y el tiempo medio de viaje. Valores altos de esta medida indican que la amplitud de la distribución del tiempo de viaje es relativamente amplia con respecto a su mediana. El índice se calculó de la siguiente forma:

$$\lambda \text{ var} = \frac{P90 - P10}{P50} \quad (5)$$

$$\lambda \text{ var} = \frac{1\,494.35 - 561.63}{666.65} = \frac{932.69}{666.65} = 1.39$$

Valores altos de este índice como el obtenido implican que la amplitud de la distribución del tiempo de viaje es relativamente extensa con respecto a la mediana y, por lo tanto, la confiabilidad en el tiempo de viaje puede clasificarse como baja.

A continuación, puede observarse una gráfica de variabilidad con respecto a la media (línea punteada) y la probabilidad del 95% del conjunto de datos.



Fuente: Elaboración propia.

#### Gráfica 4.7 Variabilidad con respecto a la media.

De acuerdo con la gráfica 4.7, se observa alta variabilidad en la muestra. Se recomienda un análisis exhaustivo del comportamiento de este segmento, tomando como punto de partida los casos que involucran tiempos por encima del percentil del 95% con el propósito de visualizar y determinar las causas de la variabilidad en los tiempos de viaje.

### 4.3.3 Análisis de confiabilidad

La información recolectada permite el análisis y el cálculo de los diferentes índices de confiabilidad. Al segmentar la información por hora del día, día de la semana, semana del mes y mes del año, los conjuntos de datos resultantes presentan muestras menores que los 30 datos. Lo anterior es la razón por la cual se dividió únicamente el conjunto de información en segmentos de operación a.m. (00:12:59 horas) y p.m. (13:23:59 horas).

A continuación, se presenta una tabla resumen con todos los estadísticos obtenidos en el análisis de confiabilidad del tramo México Querétaro para los segmentos de operación a.m. y p.m.

**Tabla 4.8 Resumen estadístico de las medidas de confiabilidad (p.m.)**

X	A2053_2054	A2054_2058	A2058_2060	A2060_2063	A2063_2065
<b>N</b>	188.0	142.0	381.0	642.0	1018.0
<b>Mín.</b>	494.0	379.0	466.1	735.0	251.0
<b>P25</b>	589.9	484.9	636.03	993.0	321.0
<b>Mediana</b>	657.4	505.0	670.8	1053.0	330.0
<b>Media</b>	666.6	506.2	672.0	1159.3	340.3
<b>P75</b>	897.2	539.1	703.8	1148.7	344.0
<b>Máx.</b>	3196.0	765.0	804.1	2378.0	503.0
<b>Ds</b>	1141.1	561.5	720.8	303.2	37.5
<b>Var.</b>	1190.9	566.1	726.2	91979.6	1409.7
<b>P10</b>	561.6	466.8	592.4	930.1	313.0
<b>P50</b>	666.6	506.2	672.02	1053.0	330.0
<b>P90</b>	1494.3	592.3	735.5	1641.0	392.3
<b>P95</b>	2275.4	643.5	758.7	1865.05	423.1
<b>Peores2</b>	2750.2	697.2	807.2	2109.06	457.3
<b>CV</b>	1.71	1.11	1.07	0.26	0.11
<b>BI</b>	2.41	0.27	0.13	0.61	0.24
<b>BT</b>	1608.8	137.2	86.7	705.6	82.8
<b>MI</b>	4.1	1.3	1.2	1.8	1.3
<b>PT</b>	2275.4	643.5	758.7	1865.05	423.15
<b>Distancia</b>	12.5	11.2	15.12	23.05	7.28
<b>TVFL</b>	502.9	449.04	604.9	921.9	291.2
<b>PTI</b>	4.5	1.43	1.25	2.02	1.45
<b>IF</b>	1.3	1.13	1.11	1.26	1.17
<b>Sesgo</b>	7.8	2.19	0.80	4.78	3.66
<b>Amplitud</b>	1.4	0.25	0.21	0.68	0.24
<b>Origen</b>	Tepetzotlán	Ent. Jorobas	T. Der. Tepeji del Río	T. Izq. Jilotepec	X. C. Jilotepec - Maravillas
<b>Destino</b>	Ent. Jorobas	T. Der. Tepeji del Río	T. Izq. Jilotepec	X. C. Jilotepec - Maravillas	T. Izq. Aculco

Fuente: Elaboración propia.

**Tabla 4.9 Resumen estadístico de las medidas de confiabilidad (p.m.)  
(continuación)**

X	A2065_2069	A2069_2072	A2069_2068	A2072_2073	A2073_2074
<b>N</b>	959.0	718.0	808.0	1170.0	989.0
<b>Mín.</b>	907.0	373.0	31.0	447.0	1137.0
<b>P25</b>	1225.7	495.7	80.5	531.2	1423.0
<b>Mediana</b>	1376.0	539.5	119.3	547.0	1468.0

<b>Media</b>	1369.8	538.4	179.9	551.6	1529.5
<b>P75</b>	1492.2	593.6	187.4	569.0	1576.0
<b>Máx.</b>	1897.0	840.0	525.0	677.0	2208.0
<b>Ds</b>	169.4	624.0	114.4	35.6	178.2
<b>Var.</b>	28721.3	622.7	13095.6	1273.4	31757.9
<b>P10</b>	1141.2	465.3	56.3	515.0	1386.0
<b>P50</b>	1376.0	538.4	119.3	547.0	1468.0
<b>P90</b>	1589.6	657.2	313.5	599.0	1811.2
<b>P95</b>	1646.5	703.0	456.6	619.0	1927.2
<b>Peores2</b>	1735.5	771.0	489.5	637.72	2047.7
<b>CV</b>	0.12	1.16	0.64	0.06	0.12
<b>BI</b>	0.20	0.31	1.54	0.12	0.26
<b>BT</b>	276.7	164.6	276.7	67.36	397.65
<b>MI</b>	1.27	1.43	2.72	1.16	1.34
<b>PT</b>	1646.5	703.0	456.6	619.0	1927.2
<b>Distancia</b>	32.19	10.71	0.88	12.23	31.55
<b>TVFL</b>	1287.5	428.2	35.1	489.2	1262.1
<b>PTI</b>	1.28	1.64	13.0	1.27	1.53
<b>IF</b>	1.06	1.26	5.12	1.13	1.21
<b>Sesgo</b>	0.91	1.62	3.08	1.63	4.19
<b>Amplitud</b>	0.33	0.36	2.16	0.15	0.29
<b>Origen</b>	T. Izq. Aculco	T. Izq. Toluca	T. Izq. Toluca	SJR	T. Izq. Acámbaro
<b>Destino</b>	T. Izq. Toluca	T der. P. SJR	Palmillas	T. Izq. Acámbaro	Libramiento Qro-SLP

Fuente: Elaboración propia.

La utilidad de los índices de fluidez y el índice de tiempo planificado yace en la comparación entre el desempeño real del subsegmento y el tiempo de viaje con flujo libre (Eisele *et al.*, 2011). El índice de fluidez para el segmento completo es de 1.58. Con base en lo anterior, se requieren  $1.58 \times 6272.30 = 9894.75$  segundos (164.91 minutos o 2.74 horas). El tiempo de recorrido de un viaje medio es de 7514.09 segundos (125.23 minutos o 2.08 horas), y existe una diferencia de 39.67 minutos entre el tiempo promedio y el calculado con los índices de confiabilidad. Al subdividir el tramo, se observa que los siguientes presentan los mayores índices de fluidez e índice de tiempo planificado:

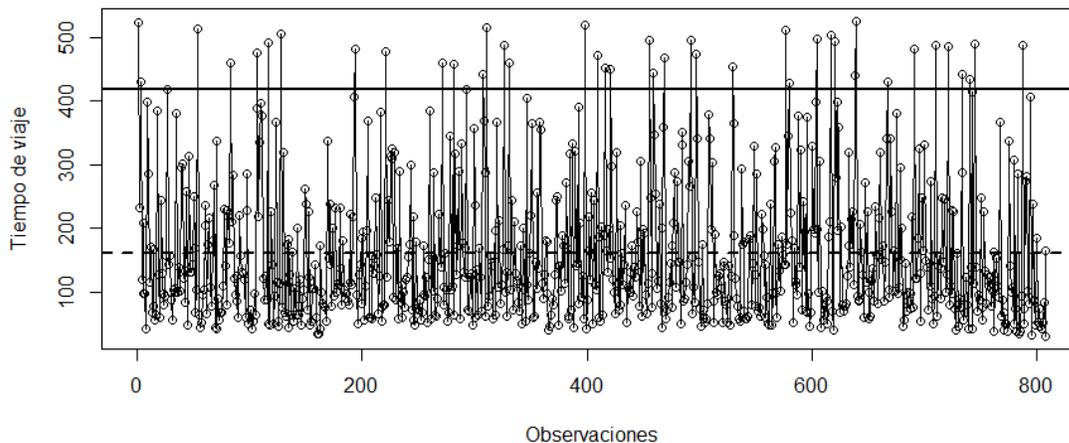
**Tabla 4.10 Índices más altos del índice de fluidez**

Id	Tramo inicial	Tramo final	FI	PTI
<b>A2069_2068</b>	T. Izq. Toluca	Caseta de cobro Palmillas	5.12	13.00
<b>A2053_2054</b>	Tepetzotlán	Ent. Jorobas	1.33	4.52
<b>A2060_2063</b>	T. Izq. Jilotepec	X. C. Jilotepec-Maravillas	1.26	2.02

Fuente: Elaboración propia.

Con base en lo anterior es correcto inferir que los sub-segmentos que presentan mayor variabilidad de acuerdo con el índice de fluidez son los mostrados en la tabla anterior (ver tabla 4.10). En comparación con los demás subsegmentos, éstos muestran un rendimiento menor, de acuerdo con la información recolectada, razón por la cual se procede a una descripción que, aunque limitada por los datos, muestra una gran utilidad del cálculo de los índices de confiabilidad relacionada con el estudio de fluidez.

### 4.3.3.1 Índice de fluidez: A2069\_2068

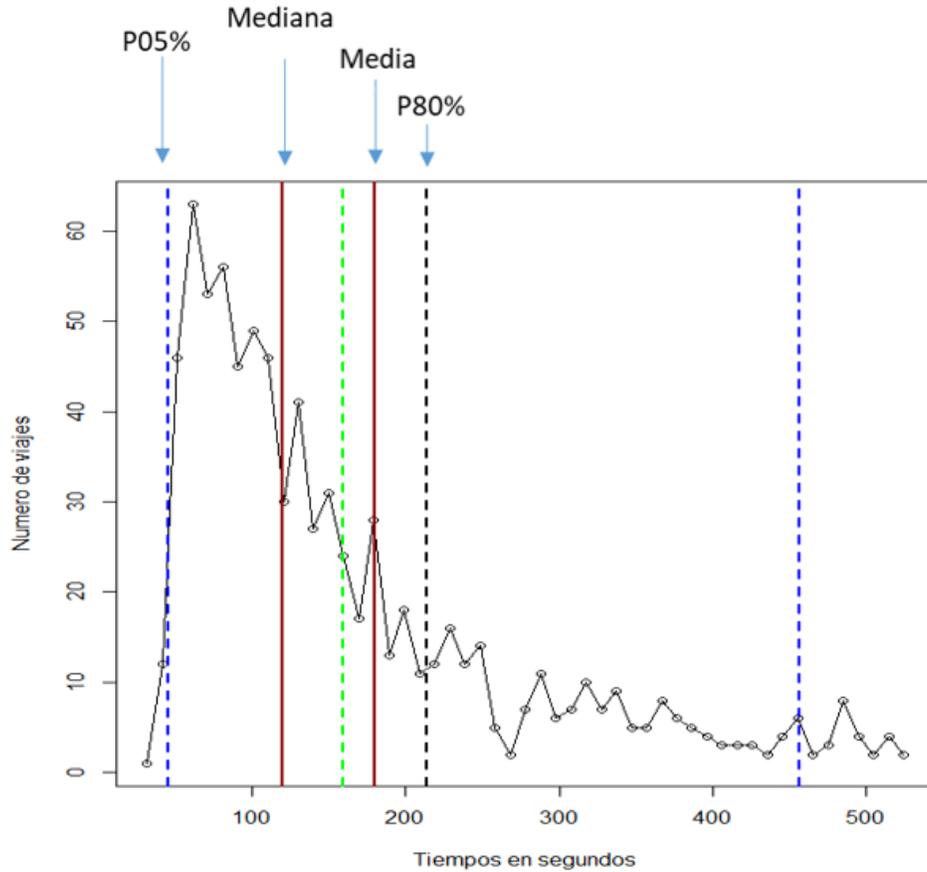


Fuente: Elaboración propia.

#### Gráfica 4.8 Gráfica de variabilidad del segmento

En la siguiente gráfica, se puede observar la variabilidad en los tiempos de viaje, donde muchos de los viajes superan incluso el P95%. Además, se presenta el valor del índice de fluidez de 5.12, un valor extremadamente alto, considerando que el índice de fluidez compara el tiempo promedio de viaje con respecto al tiempo de viaje con flujo libre.

Lo anterior indica que el tiempo promedio de viaje puede llegar a ser 5.12 veces más grande que el tiempo promedio. Además, presenta un PTI de 13.00 lo cual significa que para los viajes que requieren ser entregados de forma urgente se requiere de  $13 \times 35.11 = 456.43$  segundos (7.60 minutos). Al ser comparado con el tiempo promedio 179.97 (2.99 minutos), ofrece la posibilidad de ilustrar el desempeño del segmento del corredor el cual realmente necesita de más 130% del tiempo reportado por el promedio del tiempo.



Fuente: Elaboración propia.

#### Gráfica 4.9 Distribución de los tiempos de viaje

Como se puede observar, aproximadamente el 80% (646) viajes son menores que 213 segundos (3.55 minutos). De acuerdo con este porcentaje, uno de cada cinco viajes se encuentra en un rango de valores mayores que aumentan tanto la desviación estándar como los percentiles. La media (179.9) se encuentra considerablemente alejada de la mediana (119.9), lo que sugiere un estudio a fondo de los datos típicos por encima de éstos para identificar la causa de los tiempos de viaje más largos de lo habitual, por ser los que “empujan” los valores de la media hacia límites más altos.

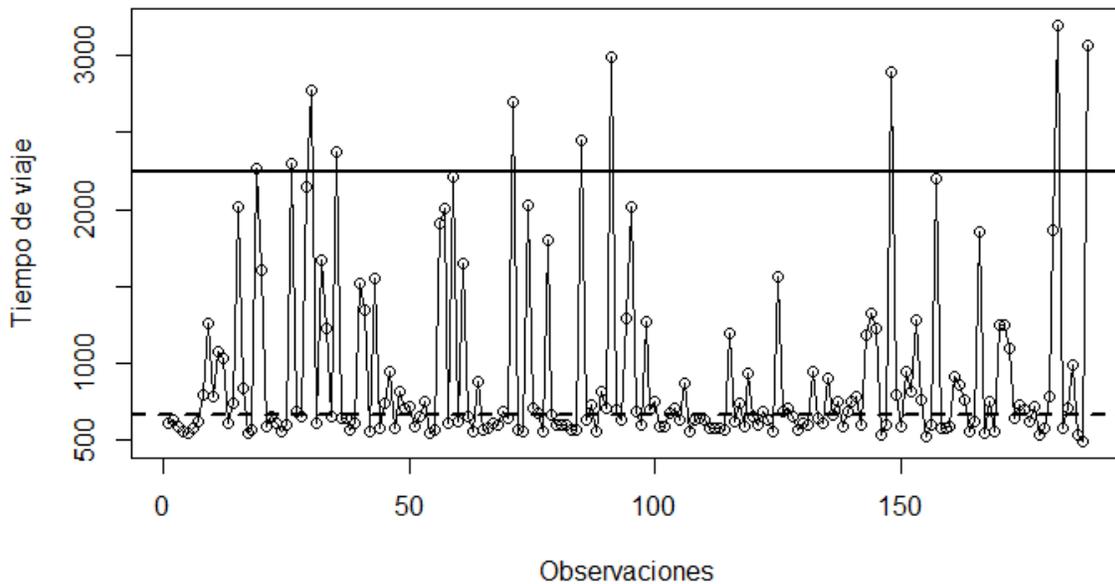
Al realizar un análisis de los tiempos que sobrepasan el P80%, se revelaron los siguientes patrones de conducta:

Para los viajes más allá del P95%, se observan colas en la caseta 2068 (caseta de cobro Palmillas), aunque no se cuenta con suficientes datos para proporcionar una hipótesis sobre los horarios más concurridos; los tiempos de viaje más prolongados entre las 18 y las 23 horas.

En el caso de los viajes que se encuentran entre el P80% y el P95%, se observan 98 ocasiones de un total de 167 oportunidades que coinciden en que el periodo de tiempo con mayores tiempos de viaje corresponde al horario de las 17 a las 23 horas, aunque no se detectaron patrones concurrentes en las congestiones, por lo cual se requiere de un estudio más a fondo, con información más específica sobre el viaje.

#### 4.3.3.2 Índice de fluidez: A2053\_2054

En este segmento se observa una gran variabilidad, con al menos 10 viajes que superan el P95% (2272.4 segundos u 11.11 minutos). El tramo en cuestión tiene un índice de fluidez de 1.33, lo que significa que un viaje que tomará 502.95 segundos (8.38 minutos) en condiciones de tiempo de viaje con flujo libre, tardará hasta  $1.33 \times 502.94 = 666.65$  segundos u 11.11 minutos.

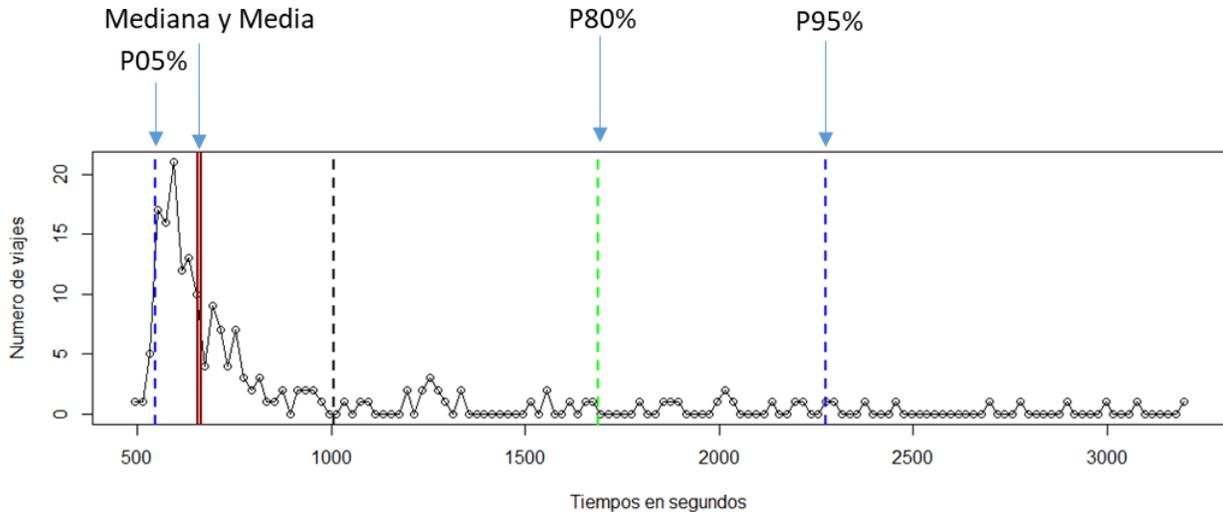


Fuente: Elaboración propia.

#### Gráfica 4.10 Gráfica de variabilidad del segmento

En la gráfica 4.11 se observa que aproximadamente el 80% (150) viajes son menores que 1006 segundos (16.76 minutos). Con base en este porcentaje, uno de cada cinco viajes se encuentra en un rango de valores mayores, que aumentan tanto la desviación estándar como los percentiles.

La media (666.653118) no se encuentra alejada de la mediana (657.4706), aunque 38 viajes se encuentran fuera del rango de P80%. Esta cantidad equivale a casi un quinto de la muestra, por lo que se sugiere un estudio a fondo de los datos típicos por encima de éstos, para identificar la causa de los tiempos de viaje más largos de lo habitual.



Fuente: Elaboración propia.

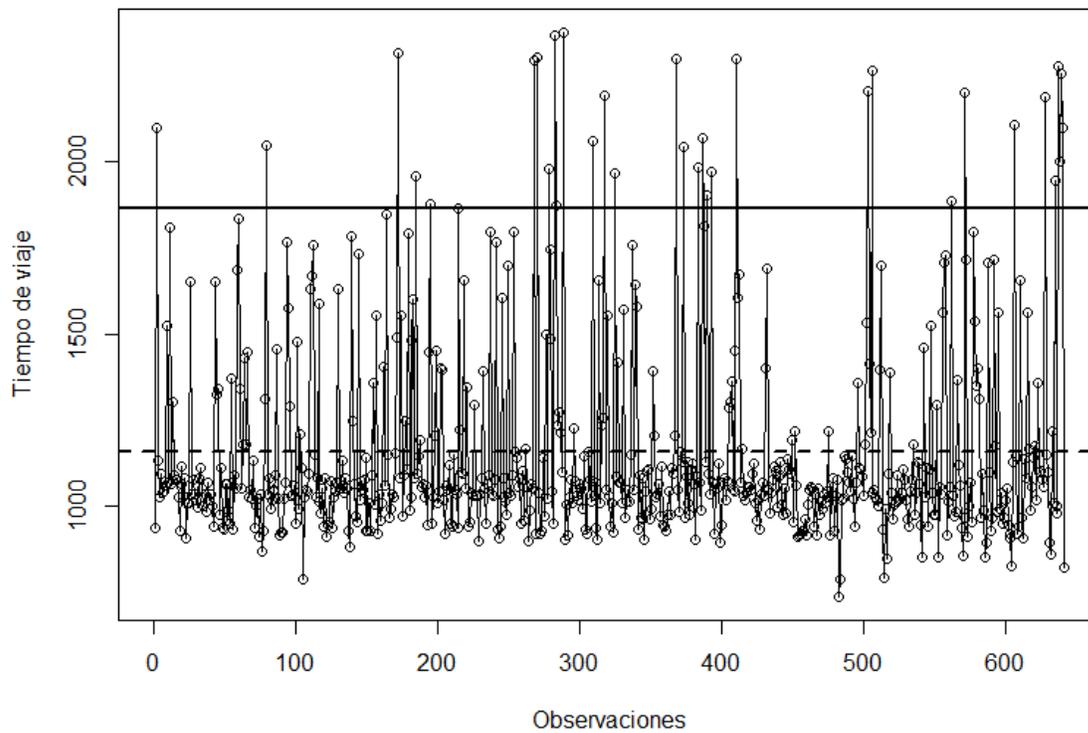
#### Gráfica 4.11 Distribución de los tiempos de viaje

Al realizar un análisis de los tiempos que sobrepasan el P80%, se revelaron los siguientes patrones de conducta:

De aquellos viajes que superan el P80% (1006.8), 28 están agrupados en un horario de 17-22 horas. Los viajes muestran congestiones similares a la mostradas en las imágenes en el paso a la segunda caseta con número 2054 (Ent. Jorobas). Los viajes superiores al P95% presentan congestionamientos importantes en la caseta 2054 (Ent. Jorobas), aunque no es posible definir las causas con la información disponible. Sin embargo, se recomienda un estudio de valores atípicos con una mayor cantidad de información.

#### 4.3.3.3. Índice de fluidez: A2060\_2063

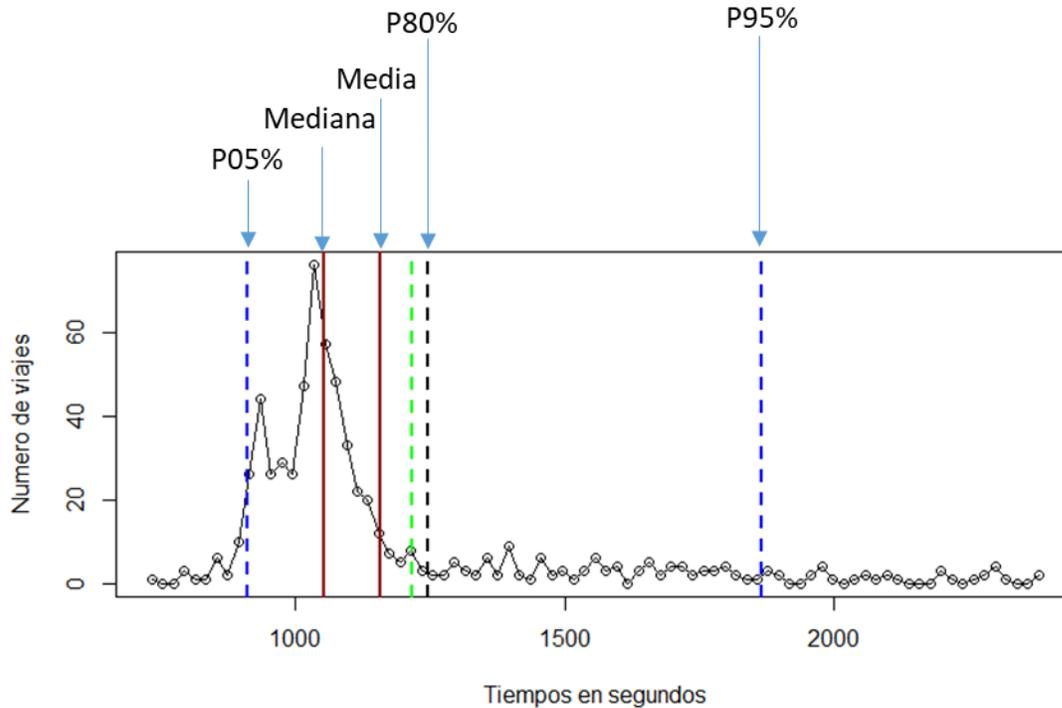
De los 642 datos analizados, en la gráfica 4.12 puede observarse que el segmento presenta alta variabilidad, con al menos 30 viajes que superan el P95% (1865.05 segundos o 31.08 minutos). El tramo en cuestión tiene un índice de fluidez de 1.257, lo que significa que un viaje que tomará 921 segundos (15.35 minutos) en condiciones de tiempo de viaje con flujo libre tardará hasta  $(1.257 \cdot 921.90 = 1158.82)$  1158.82 segundos o 19.31 minutos.



Fuente: Elaboración propia.

### Gráfica 4.12 Gráfica de variabilidad del segmento

Del total de viajes que superan tanto el P80% como el P95%, 68 de los 129 viajes se encuentran en el rango de las 18 a 22 horas. Como puede observarse, de acuerdo con la gráfica 4.13, se muestra que, aproximadamente el 80% (513) de los viajes son menores al 1247.4 segundos (20.7 minutos). De acuerdo con este porcentaje, uno de cada cinco viajes se encuentra en un rango de valores mayores que aumentan tanto la desviación estándar como los percentiles. La media (1159.3) no se encuentra considerablemente alejada de la mediana (1053); sin embargo, 129 viajes se encuentran fuera del rango de P80%. Esta cantidad equivale a casi un quinto de la muestra, por lo que se sugiere un estudio a fondo de los datos típicos, para identificar la causa de los tiempos de viaje más largos de lo habitual, ya que éstos son los que desvían los valores de la media hacia límites más altos.



Fuente: Elaboración propia.

**Gráfica 4.13 Distribución de los tiempos de viaje**

Al realizar un análisis de los tiempos que sobrepasan el P80%, no se revelaron patrones de conducta de las unidades en cuestión, por lo que se requiere de una cantidad mayor de información para evaluar el desempeño en este segmento.

### 4.3 Conclusiones del caso práctico

El estudio de confiabilidad en sistemas de transporte es de suma importancia para medir su desempeño y generar estrategias tanto de mitigación de eventos disruptivos como de mejora de los puntos débiles de la red. El caso práctico se basó en un análisis de las variaciones en el flujo del segmento de un corredor de transporte importante para México. El análisis de tiempos de viajes de unidades de transporte de carga se llevó a cabo mediante un estudio de confiabilidad. Se propuso un enfoque de desarrollo de medidas de fluidez en sistemas de transporte desde un punto de vista de análisis del desempeño.

Los índices de confiabilidad describen el tramo en análisis utilizando información correspondiente al segmento a.m. (antes meridiano) y p.m. (pasado meridiano). Estos periodos del día presentan condiciones de operación similares, aunque distintas, lo que pudo comprobarse al comparar tanto su índice de fluidez como el tiempo de recorrido total por segmento del día (ver tabla 4.11). Para el análisis que

cubre el segmento p.m., el tramo de la caseta de Tepotzotlán km 43 (2053) a la caseta Jorobas km 56 (2054) es uno de los tres principales generadores de variabilidad en la ruta, siendo el segundo en cuanto a la magnitud del IF (ver tabla 4.11).

Al analizar el segmento a.m. no se encontró dentro de los índices de fluidez más alto. En el segmento a.m. el segmento Tepotzotlán-Jorobas es sustituido por el tramo T. Izq. Toluca (2068) T. Der. Poniente San Juan del Río (2072). En la tabla los valores más altos para el índice de fluidez aparecen en color gris, con lo que se demuestra que los índices de confiabilidad y fluidez están en función del horario y las condiciones de operación del segmento en análisis. De la misma forma, existen índices que disminuyen en función del horario de análisis, y se experimenta una variabilidad menor en relación con el segmento del día en análisis.

**Tabla 4.11 Comparación de valores del índice de fluidez**

		2053	2054	2058	2060
		-	-	-	-
		2054	2058	2060	2063
	<b>Media</b>	666.65	506.24	672.02	1159.39
<b>p.m.</b>	<b>IF</b>	1.33	1.13	1.11	1.26
<b>a.m.</b>	<b>IF</b>	1.25	1.11	1.13	1.34
	<b>Media</b>	632.49	501.93	688.29	1239.83

		2063	2065	2068	2069	2072	2073
		-	-	-	-	-	-
		2065	2069	2072	2068	2073	2074
	<b>Media</b>	340.35	1369.81	538.48	179.97	551.64	1529.55
<b>p.m.</b>	<b>IF</b>	1.17	1.06	1.26	5.12	1.13	1.21
<b>a.m.</b>	<b>IF</b>	1.28	1.03	1.67	4.58	1.12	1.14
	<b>Media</b>	374.47	1333.65	718.01	161.14	550.74	1446.45

Fuente: Elaboración propia.

Se pueden obtener dos importantes conclusiones del análisis: La primera apunta a que, los tiempos promedio no representan adecuadamente los tiempos de viaje; se observan importantes diferencias entre el tiempo promedio de 125.23 minutos y el índice de tiempo planificado, el cual reporta 188.64 minutos, necesarios para asegurar la llegada el 95% de las ocasiones. Entre ambas estimaciones existe una diferencia de 63.41 minutos. Esta diferencia en tiempo es de gran importancia para la planeación de sistemas JIT y JIS, así como para la planeación de las operaciones de producción. Estas calificaciones de desempeño ayudarán a la toma de decisiones sobre el tiempo necesario para el traslado de carga dependiendo de la segmentación utilizada para el tramo. De esta forma, se pueden evitar penalizaciones por retrasos inesperados.

La segunda conclusión indica la necesidad de un modelo conceptual para la recolección, preparación, análisis y reporte de extensas bases de datos en el marco de la utilización de macrodatos (*big data*). Los datos aquí utilizados fueron tratados para su posterior análisis y el reporte de conclusiones utilizando técnicas de minería de datos y *big data*. Conjuntos de información tan grandes como los aquí presentados marcan un precedente para la utilización de modelos de *big data* y su creciente importancia.



## 5. Conclusiones

---

El presente informe de investigación es el resultado de una búsqueda bibliográfica exhaustiva de propuestas metodológicas que permitieron proponer un modelo de gestión de estrategias de *big data*, que a su vez, permita la ejecución de proyectos de macrodatos en el IMT. En este sentido, el objetivo se centró en la construcción de un modelo conceptual generalizable apropiado para el desarrollo de investigaciones futuras en materia de *big data* en el sector transporte.

El modelo aquí conceptualizado se basa en cuatro etapas: el primer paso consiste en la captura del dato. Esta etapa describe el proceso de la consecución del dato y su preparación previa. La segunda etapa comprende el procesamiento de los datos, es decir, según se reciban los datos, se prosigue a darles un tratamiento que permita la homogenización de la información. La etapa tres, denominada ciencia del dato, consiste en realizar el análisis de los datos y crear conclusiones útiles para los tomadores de decisión. Finalmente, la cuarta etapa corresponde a la presentación del dato, lo cual consiste en construir una plataforma de uso amigable para que el usuario final pueda comprender de manera natural los análisis que ahí se le presenten.

Si bien el modelo es perfectible y deberá pulirse con el desarrollo de más casos de investigación como el aquí presentado en la sección cuatro, el modelo conceptual propuesto permite desarrollar proyectos de investigación en materia de *big data*, de los cuales se pueden generar experiencias que evidencien las necesidades y problemáticas vigentes en el contexto nacional e internacional. Es de interés para el grupo de trabajo que las investigaciones en esta materia sigan avanzando en el IMT, para que dichas experiencias se compartan con la finalidad de construir un modelo cada vez más robusto que permita trabajar con múltiples fuentes de datos de forma segura, tanto para el generador de la información como para el responsable de su tratamiento.

Finalmente, se presenta una línea de trabajo futuro con la intención de mejorar el método a través de la inclusión de un mayor número de variables, que permita brindar mayor certeza a los administradores al momento de tomar decisiones. De acuerdo con Mc Donnell (2016), en materia logística los siete principales proyectos de *big data* son los siguientes:

1. Análisis de volumen de paquetes: Identificar patrones que ayuden a predecir el volumen de paquetes, permite mejorar la asignación de recursos y presupuestos.
2. Datos sobre paquetes en entornos controlados: Permite gestionar el transporte de productos que requieren un entorno controlado y encontrar rutas alternativas.

3. Elección de rutas y generación de rutas personalizadas: Con el objetivo de elegir rutas más económicas, viables y fiables, así como, con la obtención de datos provenientes de dispositivos móviles de los usuarios, se puede ofrecer un servicio de transporte más personalizado y retroalimentar el mismo flujo de información.
4. Volúmenes de datos no estructurados: El hecho de que los datos provengan de diversas fuentes influye en el volumen y la variedad de los datos, lo que dificulta su manipulación, por lo que homologar y gestionar estos datos supone un área de gran interés en materia de *big data*.
5. Certeza en tiempos de traslado y entrega: La manipulación de diferentes tipos de variables, incluso en tiempo real, permite generar información más precisa acerca del traslado y entrega de productos.
6. Analítica web: Las redes sociales, páginas web y aplicaciones móviles generan gran cantidad de información útil para la gestión de logística.
7. Servicio al cliente: A través de datos en tiempo real como la geolocalización, se puede conocer el estado exacto de los productos ofreciendo a los receptores certeza del transporte.





## Bibliografía

---

Bernal, C., y Landaverde, J. (1995). *Proyecto de modernización de las pyme. Gestión tecnológica*. Santa Fe de Bogotá: Servicio Nacional de Aprendizaje.

Bogers, E., Van Lint , H., y Van Zuylen, H. (2008). "Reliability of Travel Time: Effective Measures from a Behavioral Point of View." *Transportation Research Record: Journal of the Transportation Research Board*, pp. 27-34.

Casassus, J. (2000). *Problemas de la gestión educativa en América Latina*. Santiago de Chile: UNESCO.

Castañón, L. D. C. A., y Ricoy, M. T. P (2011). *Una valoración de la gestión del conocimiento, para el desarrollo organizacional en empresas locales manufactureras del estado de Guanajuato, México. XVI Congreso Internacional de Contaduría, Administración e Informática*. Obtenido el 4 de mayo de 2017, de: <http://132.248.164.227/congreso/docs/xvi/docs/15T.pdf>

Castro, J. A. A. (2003). "Tecnologías y sistemas de información en la gestión de conocimiento en las organizaciones." *Revista Venezolana de Gerencia*, vol. 8, núm. 24.

Cedillo Campos, M. G., Lizarraga-Lizarraga, G., y Martner-Peyrelongue, C. D. (2017). "MiF3 method: Modeling intermodal fluidity freight flows." *Research in Transportation Economics*, pp. 1-10.

Diego, D. G. (2016). "¿Por qué aplicar el *big data* en el sector de la logística?" IMF Business School. Recuperado el 19 de agosto de 2017, de: <http://www.imf-formacion.com/blog/logistica/cadena-de-suministro/big-data-sector-logistico/>.

Dogson, M., Gann, D., y Salter, A. (2008). *The management of technological innovation*. Oxford, Nueva York.

Eisele, W. L., Tardif, L.-P., Villa, J., Schrank, D., y Lomax, T. (2011). "Evaluating Global Freight Corridor Performance for Canada." *Journal of transportation*, Institute of Transportation Engineers, vol. 1, núm. 1, pp. 39-58.

Foucault, M. (2013). *El poder, una bestia magnífica. Sobre el poder, la prisión y la vida*. Siglo Veintiuno Editores. México, Distrito Federal.

Macintosh, A. (1997). *Position Paper on Knowledge Management*. Adinburgh: Artificial Intelligence Applications Institute.

Mc Donell, F. (2016). "Big data y logística en siete proyectos tangibles." *Revista de logística mantenimiento y almacenaje*. Recuperado el 20 de agosto de 2017, de: <http://www.manutencionyalmacenaje.com/es/notices/2016/11/big-data-y-logistica-en-siete-proyectos-tangibles-39492.php#.WZuLYz4jHIU>.

Pérez, D., y Dressler, M. (2007). *Tecnologías de la información para la gestión del conocimiento*. Intangible Capital, enero-marzo de 2007, vol. 3, núm. 15, pp. 31-59.

Picciano, A. (2012). "The evolution of the Big Data and learning analytics in the American Higher School." *Journal of Asynchronous Learning Networks*, vol. 16, núm. 9.

Prieto, I. M., y Revilla, E. (2004). "La naturaleza dual de la gestión del conocimiento: *The dual nature of managing knowledge*." Academia. Revista Latinoamericana de Administración, núm. 32.

Team, C. P. (2002). "Capability Maturity Model® Integration (CMMI), Version 1.1-- Continuous Representation."

Solleiro, J. L. (1988). *La gestión y la administración de tecnología*. México: Cuaderno del Instituto de Investigaciones Jurídicas.

Van Lint, J., y van Zuylen, H. (2005). "Monitoring and Predicting Freeway Travel Time Reliability: Using Width and Skew of Day-to-Day Travel Time Distribution." *Transportation Research Record: Journal of the Transportation Research Board*, pp. 54-62.

Vera, S. F. L. (2015). "Tips para implementar el Big Data en logística." *Revista Énfasis Logística*. Recuperado el 18 de agosto de 2017, de: <http://www.logisticamx.enfasis.com/articulos/72655-tips-implementar-el-big-data-logistica>.

Zorrilla, H. (1 de enero de 1997). Recuperado el 20 de abril de 2015, de "La gerencia del conocimiento y la gestión tecnológica" en: [http://bv.ujcm.edu.pe/links/cur\\_comercial/PensaGestEstrategica-13.pdf](http://bv.ujcm.edu.pe/links/cur_comercial/PensaGestEstrategica-13.pdf) .

Oracle (2015), obtenido el 21 de agosto de 2017, de: [http://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529\\_esa.pdf](http://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529_esa.pdf).

SAS (2017). "Big Data analytics, what it is and why it matters", obtenido el 22 de agosto de 2017, de: [https://www.sas.com/en\\_us/insights/analytics/big-data-analytics.html](https://www.sas.com/en_us/insights/analytics/big-data-analytics.html).

Accenture (2016). "Big Data & Analytics Overview", obtenido el 22 de Agosto de 2017, de: <https://kxdocuments.accenture.com/?k=data%20analytics&page=1&0248=Yes>.

Oracle (2013). "Big Data Analytics Advanced Analytics in Oracle Database", obtenido el 22 de Agosto de 2017, de: <https://kxdocuments.accenture.com/contribution/dfb59467-037c-45c3-82d2-a6f5186b7eea?referrer=https://kxdocuments.accenture.com>.

Tableau (2017). "Tableau for the Enterprise: An Overview for IT", obtenido el 22 de agosto de 2017, de: [https://www.tableau.com/sites/default/files/media/whitepapertableauforenterprise\\_0.pdf](https://www.tableau.com/sites/default/files/media/whitepapertableauforenterprise_0.pdf).

Tableau (2017). "Tableau Server 10.0 High Availability: Delivering mission-critical analytics at scale", obtenido el 22 de Agosto de 2017, de: [https://www.tableau.com/sites/default/files/whitepapers/high\\_availability\\_10.0\\_0.pdf](https://www.tableau.com/sites/default/files/whitepapers/high_availability_10.0_0.pdf)

Tableau (2017). "Escalabilidad de Tableau Server Descripción introductoria: cómo escalar Tableau Server en toda su empresa", obtenido el 22 de agosto de 2017 de: [https://www.tableau.com/sites/default/files/whitepapers/scalability\\_overview\\_es-es.pdf](https://www.tableau.com/sites/default/files/whitepapers/scalability_overview_es-es.pdf)

Tableau (2017). "Recomendaciones y requisitos de hardware mínimos para Tableau Server", obtenido el 22 de agosto de 2017, de: [http://onlinehelp.tableau.com/current/server/es-es/server\\_hardware\\_min.htm](http://onlinehelp.tableau.com/current/server/es-es/server_hardware_min.htm)









Km 12+000 Carretera Estatal 431 "El Colorado-Galindo"  
Parque Tecnológico San Fandila  
Mpio. Pedro Escobedo, Querétaro, México  
CP 76703  
Tel +52 (442) 216 9777 ext. 2610  
Fax +52 (442) 216 9671

[publicaciones@imt.mx](mailto:publicaciones@imt.mx)

<http://www.imt.mx/>

Esta publicación fue desarrollada en el marco de un sistema de gestión de calidad  
certificada bajo la norma ISO 9001:2015