



FACULTAD DE POSGRADOS

ESTUDIO COMPARATIVO DE MÉTODOS EXISTENTES PARA INTEGRAR LA
INFORMACIÓN ESTRUCTURADA Y NO ESTRUCTURADA DE UNA
INDUSTRIA ENFOCADO EN LA GENERACIÓN DE CONOCIMIENTO,
DESDE LA PERSPECTIVA DE UNA SOLUCIÓN INTEGRAL DE BIG DATA.

Trabajo de Titulación presentado en conformidad con los requisitos establecidos
para optar por el título de Magister en Gerencia de Sistemas y Tecnologías de la
Información.

Profesor Guía
Jaime Vinuesa

Autora
Sonia Carolina Guama Morales

Año
2015

DECLARACIÓN PROFESOR GUÍA

“Declaro haber dirigido este trabajo a través de reuniones periódicas con la estudiante, orientando sus conocimientos y competencias para un eficiente desarrollo del tema escogido y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación”

Ing. Jaime Augusto Vinueza Trujillo, MBA

CC: 1716028509

DECLARACIÓN DE AUTORÍA DEL MAESTRANTE

“Declaro que este trabajo es original, de mi autoría, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales que protegen los derechos de autor vigentes”

Ing. Sonia Carolina Guama Morales

CC: 1718304411

AGRADECIMIENTOS

Agradecimientos a Dios por el valor y la perseverancia para permitirme culminar esta meta académica

Agradecimientos al Ing. Jaime Vinuesa, por su amistad, apoyo y guía en el desarrollo del presente proyecto.

Agradecimientos a mi familia por su apoyo y preocupación para cumplir con este objetivo.

DEDICATORIA

Dedicado a Dios y a mi familia que son el pilar sobre el que construyo mis sueños.

RESUMEN

La generación de grandes cantidades de datos ha permitido que las industrias encuentren oportunidades de mejora a través del análisis e integración de datos estructurados y no estructurados elevando la fuente de conocimiento en el desarrollo de productos e innovación en las áreas críticas de negocio desde la perspectiva de Big Data.

Big Data es una solución empresarial con algunos componentes tecnológicos que permite generar valor para el negocio a partir de la toma de decisiones en base al análisis de los datos; las principales características de esta solución son el volumen, la variabilidad, la velocidad y la veracidad.

Existen varias técnicas y tecnologías que apoyan el trabajo de análisis e integración de diversas fuentes de datos para generar información que permita evidenciar tendencias y requerimientos puntuales de la comunidad para mejorar la toma de decisiones. Sectores como el gobierno y el turismo pueden aprovechar para mejorar su gestión en base al estudio y análisis de los datos.

El conocimiento genera valor y eleva las capacidades y habilidades para la toma de decisiones de una industria. La generación de conocimiento en base al análisis de datos requiere que la industria tenga claro el caso de negocio al que va a enfocar su estudio e investigación, al igual que contar con los datos necesarios para el análisis.

La implementación de Big Data en una industria requiere la consideración de ciertos factores que ayudaran al cumplimiento de los objetivos de este tipo de proyectos. Contar con un equipo de expertos, apoyo de la gerencia y de toda la organización, además de una mente abierta y una adecuada gestión de proyectos, es fundamental, para iniciar la implementación de Big Data.

Factores como la cantidad de datos, la heterogeneidad, la variabilidad de los datos y las necesidades del negocio ayudan a escoger la mejor técnica de análisis e integración para identificar las oportunidades y ventajas competitivas además de algunas ideas para innovación y mejora de procesos y productos.

PALABRAS CLAVE

TECNICAS DE ANÁLISIS DE DATOS/ INTEGRACIÓN DE LA INFORMACIÓN/
GENERACIÓN DE CONOCIMIENTO/ BIG DATA

ABSTRACT

Generation of large amounts of data has allowed industries to find improvement opportunities through analysis and integration of structured and unstructured data increasing knowledge for product development and innovation in the critical areas of business considering a Big Data perspective.

Big Data is a business solution with technological components that generate business value by making decisions based on data analysis; the main characteristics of this solution are volume, variability, speed and accuracy.

There are several techniques and technologies that support analysis and integration of various data sources to discover community trends and specific requirements to improve decision - making. Sectors such as Government and Tourism can improve their management based on these technics.

Knowledge creates value and raises the capabilities and skills for decision making in an Industry. The knowledge generation based on data analysis requires an industry clear business case to focus the research, as well as enough data to analyses.

A Big Data implementation requires specific factors to be considered in order to fulfil all Project goals. It is important to have an expert's team, management organization support as well as an open mind and a proper project management.

Factors such as the amount of data, heterogeneity, variability of data and business needs help choosing the best technical analysis and integration options to identify opportunities and competitive advantages as well as some ideas for innovation and improvement of processes and products.

ÍNDICE

PRÓLOGO	1
INTRODUCCIÓN	3
1. GENERALIDADES	5
1.1. Antecedentes.....	5
1.2. Marco Teórico	8
1.2.1. Definición del Problema	8
1.2.2. Justificación de Proyecto	10
1.2.3. Metodología de Investigación	11
1.3. Introducción a Big Data.....	12
1.3.1. Importancia de Big Data	18
1.3.2. Perspectiva de Big Data para la integración	21
1.3.3. Técnicas	25
1.3.4. Metodologías de gestión de proyectos de Big Data.....	33
2. Perspectiva del Negocio	39
2.3. Características de la organización	39
2.4. Aplicación de <i>Big Data</i> en la organización, perspectiva de negocio.....	43
2.5. Generación de Conocimiento desde la perspectiva de <i>Big Data</i>	50
2.6. Experiencias con Big Data	53
2.6.1. Gobierno.....	54
2.6.2. Turismo.....	59
2.7. Generalidades de las Tecnologías.....	65

3. Estudio Comparativo	75
3.1. Análisis de técnicas.....	75
3.2. Resumen y Análisis de resultados	92
CONCLUSIONES Y RECOMENDACIONES	98
GLOSARIO	103
REFERENCIAS	106
ANEXOS	113

ÍNDICE DE FIGURAS

<i>Figura 1.</i> Evolución tecnológica de Big Data	6
<i>Figura 2.</i> Crecimiento del mercado de Big Data en América Latina	9
<i>Figura 3.</i> Dimensiones de Big Data	15
<i>Figura 4.</i> Características de Big Data definidas por IBM, de acuerdo al volumen, velocidad y variedad – o simplemente las 3V.	17
<i>Figura 5.</i> Dificultades identificadas para la gestión de datos no estructurados.....	41
<i>Figura 6.</i> Etapas para la aplicación de <i>Big Data</i>	44
<i>Figura 7.</i> Resumen de las etapas para la aplicación de <i>Big Data</i>	49
<i>Figura 8.</i> Gestión del Conocimiento	51
<i>Figura 9a.</i> Razones de uso de Internet por área en el Ecuador	57
<i>Figura 9b.</i> Razones de uso de Internet por área en el Ecuador	57
<i>Figura 10.</i> Porcentaje de acceso a Internet en Ecuador	58
<i>Figura 11.</i> Personas que cuentan con celular y redes sociales en Ecuador ...	59
<i>Figura 12.</i> Cifras Turismo de Ecuador.....	62
<i>Figura 13.</i> Edad de turistas que visitan el Ecuador	63
<i>Figura 14.</i> Nivel de educación de turistas que visitan el Ecuador.	63
<i>Figura 15.</i> Triangulo de clasificación base de datos por propiedad.....	68
<i>Figura 16.</i> Taxonomía de las Bases de Datos NoSQL.....	69
<i>Figura 17.</i> Ranking bases de datos.....	70
<i>Figura 18.</i> El valor de Big Data para el negocio.	93
<i>Figura 19.</i> Estrategia para crear un plan exitoso para Big Data.....	95

ÍNDICE DE TABLAS

Tabla 1. Clasificación de definiciones de Big Data	13
Tabla 2. Datos generados por los medios sociales	21
Tabla 3. Comparación de ventajas y desventajas de un modelo SQL y NoSQL.	65
Tabla 4. 4 tipos de bases de datos <i>NoSQL</i>	66
Tabla 5. Descripción de las principales características de las bases de datos NoSQL	69
Tabla 6. Comparación de las soluciones para el procesamiento de datos desde la perspectiva de <i>Big Data</i>	73
Tabla 7. Niveles de madurez del proceso de análisis	77
Tabla 8. Técnicas de análisis estadístico	78
Tabla 9. Técnicas de análisis de visualización de datos	81
Tabla 10. Técnicas de análisis de automatización	82
Tabla 11. Técnicas de análisis semántico	82
Tabla 12. Técnicas de análisis predictivo	84
Tabla 13. Comparación de técnicas de análisis de datos aplicables en Big Data	85
Tabla 14. Valoración de características de las técnicas de Big Data	87
Tabla 15. Calificación de las técnicas de Big Data	88

PRÓLOGO

El presente estudio se basa en el análisis de las principales técnicas de análisis e integración de datos desde la perspectiva de Big Data para generar conocimiento dentro de una organización a través de los resultados generados.

Dentro de las organizaciones privadas y públicas de diferentes verticales existe una gran cantidad de información generada y recolectada a través de diferentes medios que puede ser de gran utilidad para el mejoramiento de oportunidades para el negocio, la sociedad, un país o el mundo entero, pero que lamentablemente no se han aprovechado por falta de conocimiento e información adecuada de las técnicas que se puede aplicar para sacar el mayor provecho del análisis de los datos considerando una adecuada estrategia de implementación.

La investigación del presente estudio se desarrolló en base a la exploración de diversas fuentes entre los cuales se destacan textos de autores especializados en Big Data y consultas en la web de sitios certificados como fuente de apoyo.

Durante la investigación se presentaron algunos inconvenientes para la recolección de la información dado que las fuentes certificadas son limitadas y la información está dispersa y existen diversas opiniones sobre un mismo tema lo cual dificultó la estructuración de este estudio

Esta investigación ha permitido desarrollar una guía para la introducción de Big Data en una organización, independientemente de su vertical, para generar conocimiento que les permita innovar, renovar o mejorar la visión de negocio que desean alcanzar. Un estudio comparativo de las principales técnicas de análisis e integración de los datos, consideraciones claves para la definición de una estrategia de implementación, generalidades de la metodología de implementación y arquitectura de Big Data.

Esta investigación está dirigida a empresas que deseen innovar y reemplazar una cultura basada en los datos estructurados y desee aprovechar la gran

cantidad de información disponible para optimizar las oportunidades de mercado basados en el análisis de los datos.

Está dirigido también a personas interesadas en ampliar su campo de conocimiento y que deseen fundamentar la implementación de Big Data dentro de sus ideas de negocio o de la organización en la que se desempeñen.

INTRODUCCIÓN

El conocimiento es experiencia, todo lo demás es información.

Albert Einstein.

El conocimiento es una de las nuevas fortalezas consideradas dentro de las organizaciones, sin embargo su generación es una debilidad que debe ser contemplada como un posible riesgo en el ámbito competitivo.

A nivel mundial el acceso a la tecnología ha permitido que se generen miles de millones de datos que pueden ser aprovechados en diferentes sectores como por ejemplo: el gobierno, la salud, manufactura e innovación en casi todos los ámbitos de estudio, entre otros; Estas grandes cantidades de datos se han ido introduciendo en diferentes modelos de análisis para generar valor y así lograr mayor aprovechamiento de las oportunidades y optimizar los procesos para llegar al cliente considerando sus verdaderas necesidades.

La Inteligencia de Negocios es una estrategia que se ha venido manejando en grandes sectores comerciales relacionados con la comercialización de productos, campañas de marketing, estudios de acceso a información en la web, banca, seguridad, acceso a tecnología, automotriz, etc. Pero, dado el volumen y variedad de datos generados y recolectados se ha venido complicando y reduciendo el campo de análisis de los mismos, por lo cual, se han visto en la necesidad de implementar técnicas orientadas al manejo de grandes cantidades de información que no se limiten ante la variabilidad de la fuente de datos y que procesen, analicen e integren la información a gran velocidad.

Big Data desde sus inicios se ha enfocado en el procesamiento de grandes cantidades de información (Volumen), provenientes de diversas fuentes (Variabilidad) a gran Velocidad, lo cual ha permitido ampliar el campo de análisis de los datos y generar resultados que han agregado valor al conocimiento adquirido de una organización para alcanzar los objetivos planteados con la implementación de Big Data.

El objetivo de este estudio se enfoca en la necesidad de identificar una técnica o método adecuado para integrar la información estructurada y no estructurada de una industria que permita construir una base de conocimientos ordenada y fácil de explotar, partiendo de la investigación de métodos existentes y aplicables al entorno nacional desde la perspectiva de Big Data.

El estudio está fundamentado en una exhaustiva investigación de las técnicas de análisis e integración de la información estructurada y no estructurada considerando la perspectiva de Big Data.

La propuesta desarrollada en base a la investigación pretende generar una base de conocimiento gerencial, identificando las principales características que debe tener el negocio para utilizar esta tecnología al igual que los lineamientos estratégicos que guiaran un proyecto de implementación de Big Data hacia un resultado exitoso.

CONTENIDO

El documento se ha desarrollado en 3 capítulos que contienen información general sobre el análisis e integración de datos, la importancia de Big Data para la generación de conocimiento y un estudio comparativo de las técnicas de integración de la información desde la perspectiva de Big Data.

Capítulo I iniciativa para la investigación, justificación del estudio comparativo, introducción a Big Data, descripción de las técnicas de integración y metodología para la implementación de proyectos de Big Data

Capítulo II Perspectiva del negocio para la generación de conocimiento, estudio de casos para la aplicación de Big Data, generalidades sobre las principales herramientas para implementar Big Data.

Capítulo III Análisis comparativo de las técnicas de análisis e integración de los datos considerando las principales características, ventajas y desventajas para su aplicación dentro de una industria.

Al final se puede encontrar las conclusiones y recomendaciones del estudio, y demás Anexos importantes para la investigación

1. GENERALIDADES

1.1. Antecedentes

Los grandes cambios que han impactado sobre la forma de administrar un negocio, gestionar sus recursos y el crecimiento del mismo vienen acompañados de diferentes factores y exigencias que obligan a las empresas tomar decisiones y mejorar la estrategia con la que dirigen la misma, incluyendo actividades e información antes no valoradas por la empresa.

Durante años la información ha sido un factor de preocupación para las empresas por la inversión que deben realizar para almacenarla de forma segura, pero lejos de preocuparse por el valor que esta puede generar al negocio y la ganancia a nivel económico y estratégico que se puede obtener si se la analiza de forma adecuada.

Hace más de 20 años se viene manejando técnicas de Inteligencia de Negocios y Minería de Datos para aprovechar de forma creativa e inteligente los datos estructurados de una empresa con la finalidad de obtener información útil para la toma de decisiones sobre un ámbito específico y para establecer los lineamientos que lleven al éxito un negocio o lo mantengan bien posicionado dependiendo de su vertical; más el crecimiento de la información y su procedencia en muchos casos de fuentes no relacionales o estructuradas ha limitado su estudio y provocado cierta incertidumbre sobre lo que espera el consumidor o el usuario final sobre los productos y servicios ofertados en el mercado (Urquizu, 2009).

(O'Reilly Media, Inc., 2012), describe a Big Data como una de las tecnologías que ha evolucionado en el campo de los negocios y de TI, partiendo como una idea innovadora en el año 2012. En la actualidad se ha vuelto una de las soluciones con mayor trascendencia en el análisis de los datos estructurados y no estructurados como se muestra en la *Figura 1*:

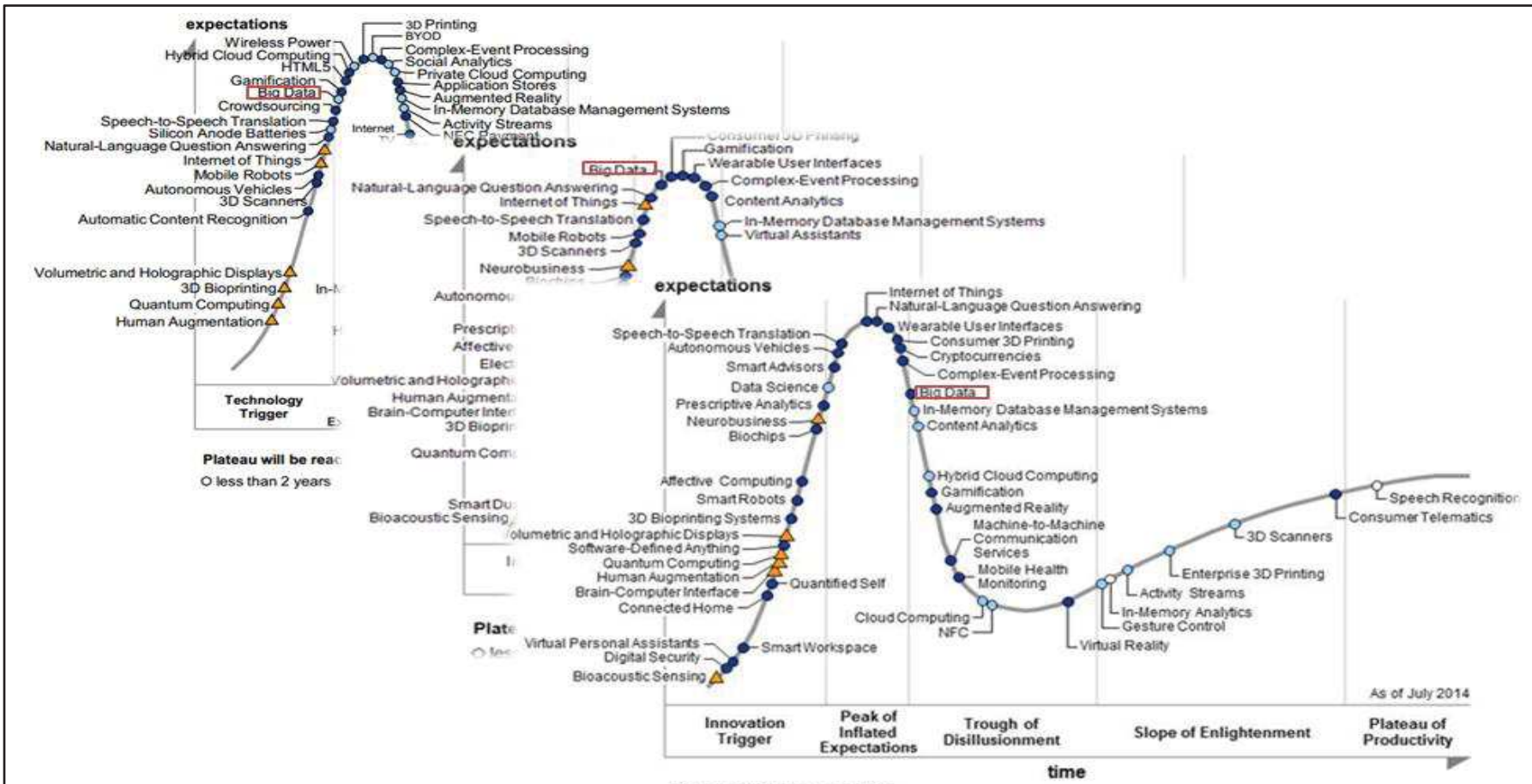


Figura 1. Evolución tecnológica de Big Data

Tomado de Gartner, Inc, "Hype Cycle for Emerging Technologies Identifies "Tipping Point" Technologies That Will Unlock Long-Awaited Technology Scenarios", 2012; Gartner, Inc, "Hype Cycle for Emerging Technologies Maps Out Evolving Relationship Between Humans and Machines", 2013; "Hype Cycle for Emerging Technologies Maps the Journey to Digital Business", 2014.

a. Representación histórica de la evolución de tecnologías, servicios y estrategias en el ciclo de Mercado y su aporte de valor al negocio y TI

El libre acceso a la tecnología en más de la mitad de la población mundial, la apertura de fronteras comerciales, el uso de Internet, las comunicaciones globales, el mercado electrónico, las redes sociales y factores ambientales, entre otros han provocado que el sector empresarial y los gobiernos se preocupen por gestionar los datos generados a través de las transacciones o publicaciones que se realizan por los medios mencionados, de manera que se analice los datos a un nivel más específico, predicciones de cambios en diferentes sectores, prevención en el campo de la salud, manejo estratégico de la oferta y demanda, etc.

Para alcanzar este objetivo se han desarrollado algunas herramientas y técnicas, además de una serie de términos que permiten al área de TI estudiar y aplicar en este nuevo campo, donde se pretende administrar grandes cantidades de datos, sin importar la estructura de los mismos para obtener información manejable a nivel gerencial y que sea una fuente de conocimiento para saber lo que espera la ciudadanía en el caso de los gobiernos, o el mercado en el caso de la industria.

Big Data es un término establecido hace aproximadamente 12 años y definido como Volumen, Velocidad y Variedad, que posteriormente se ha ido tornando en el manejo de grandes volúmenes de información a gran velocidad y de variadas fuentes de datos. Desde esta perspectiva las empresas han venido definiendo mejores prácticas para su implementación y métodos de manejo de la información a gran escala.

Al igual que el resto del mundo Ecuador cuenta con grandes sectores a nivel empresarial que estarían interesados en implementar técnicas relacionadas con *Big Data*, para mejorar las estrategias de negocio pero basados en una técnica o método de implementación fundamentado en un estudio objetivo de los métodos de aplicación existentes.

1.2. Marco Teórico

1.2.1. Definición del Problema

La innovación tecnológica es una de las principales causas de la generación de gran cantidad de datos que han dado lugar a la implementación de sistemas de información y tecnología que han facilitado la administración de los datos estructurados, al igual que se ha implementado herramientas de análisis de este tipo de datos para mejorar la toma de decisiones; sin embargo, la creación de las redes sociales y de otro tipo de fuentes de datos no estructurados han mostrado la necesidad de contar con herramientas y tecnologías que permitan integrar este tipo de datos con los datos relacionales manejados desde hace mucho tiempo atrás.

La necesidad de analizar los datos no estructurados se debe a la importancia que tiene para las organizaciones el relacionamiento con el cliente, su opinión y el poder que ha adquirido en el mercado actual, independientemente de la vertical o los intereses económicos que esta tenga. Orientar la inversión hacia este objetivo es el reto que deben asumir las organizaciones que desean generar valor al igual que empresas multinacionales como Amazon, Twitter y Facebook que han aplicado técnicas de análisis de datos y herramientas para hacer uso de los datos no estructurados para generar patrones de afinidad de acuerdo a la información proporcionada por los usuarios a través de sus portales.

Para la integración de los datos estructurados y no estructurados existen técnicas de *Big Data* que dependiendo de la organización pueden o no ser aplicadas, al igual que, software que requiere del trabajo de expertos para su administración y especialistas en datos para realizar análisis adecuados que generen resultados de utilidad para la toma de decisiones.

El reportaje realizado por la (Revista LÍDERES, 2013) transmite la opinión de los entrevistados quienes consideran que: “La implementación de proyectos de *Big Data* en el Ecuador no es una de las preocupaciones actuales para las organizaciones”, en especial para las instituciones públicas, a pesar de que ya

existen empresas que ofrecen este servicio, aunque no es una de las soluciones que más se ofertan.

Según el (*IDC Big Data Report, 2012*) países como Chile, Colombia y Perú hasta el año 2018 tendrán un sorprendente avance en el uso de esta tecnología y no es difícil imaginar el objetivo comercial detrás de esta innovación en sus empresas; ejemplo que puede ser atractivo para el Ecuador y aplicado a gran escala si se establece una estrategia de implementación y adaptación adecuada.



Una apropiada estrategia de implementación requiere de un plan de acción donde se establezcan objetivos claros sobre los resultados de los datos que se van a analizar, inversión, plazos, y un equipo de trabajo especializado entre los factores de mayor impacto en las metodologías para el correcto desarrollo de un proyecto.

Establecer la mejor estrategia para integrar la información es un gran reto que requiere de un alcance definido donde se identifiquen los resultados de la

integración de los datos y, el uso justificado de los mismos en los diferentes proyectos de emprendimiento, innovación y renovación que tenga planificado la empresa.

Este proyecto pretende establecer el marco teórico de las principales técnicas de integración de la información desde la perspectiva de Big Data, enfocados en la generación de conocimiento para el negocio.

1.2.2. Justificación de Proyecto

Según los estudios publicados por (Selvage, 2013) y (Buytendijk & Laney, 2014) las empresas no manejan un modelo de integración de la información y gestión del conocimiento; los datos se encuentran dispersos en documentos, bases de datos, redes sociales, etc., esto ha generado la existencia de islas de datos independientes que en algunas ocasiones limitan la visión de una empresa para la toma de decisiones y a esta complicación se suma la explosión del volumen de información para manejarla adecuadamente tanto en datos estructurados como no estructurados; realidad que actualmente también se encuentra reflejada en Ecuador.

Con la finalidad de presentar una visión uniforme y coherente de los datos, establecer mecanismos apropiados para el análisis de la información y el creciente volumen de data, entre otros, se propone la elaboración de un estudio para la identificación y análisis comparativo de una adecuada técnica o método que permita integrar la información estructurada y no estructurada para una industria, enfocada en la generación de conocimiento, partiendo de la investigación de métodos existentes y aplicables al entorno nacional desde la perspectiva de Big Data.

El desarrollo de este estudio comparativo es de gran relevancia para cualquier industria que desee generar valor al negocio a través de la adecuada gestión de los datos, integrándolos de forma que pueda analizar la información y beneficiar a la organización con estrategias para la toma de decisiones.

La integración de la gran cantidad de datos que se generan por diversos medios transaccionales y no transaccionales, se ha vuelto un gran reto para las empresas que desean enfocarse en brindar servicios y productos adecuados a las necesidades del consumidor y de los mercados en general; sin embargo, el no contar con una estrategia de integración ni técnicas o métodos adaptados a su realidad ha dificultado el análisis de la información. El desarrollo de este estudio comparativo pretende identificar una técnica o método en base a las mejores prácticas aplicadas por otros países y adaptarla a la realidad nacional.

El estudio de los métodos existentes para la aplicación de *Big Data* y la integración de los datos, la realidad de la industria y su nivel de madurez permitirá identificar una técnica o método considerando las mismas variables mencionadas pero adaptadas a la realidad de Ecuador; permitiendo integrar los datos enfocados en la generación de conocimiento, con fuentes veraces y casos exitosos demostrables a partir de la documentación oficial publicada.

1.2.3. Metodología de Investigación

El presente es un estudio exploratorio pero contiene algunos elementos descriptivos.

“Los estudios exploratorios se efectúan, normalmente, cuando el objetivo es examinar un tema o problema de investigación poco estudiado o que no ha sido abordado antes” (Metodología de la Investigación, 2014) como lo establece la definición citada las técnicas o métodos de integración de los datos para el adecuado análisis de la información y elevar la calidad en la toma de decisiones es un tema que no se encuentra completamente estudiado ni se han establecido procesos genéricos aplicables a cualquier realidad, pues, desde la perspectiva de Big Data depende mucho de factores que durante la exploración bibliográfica se han podido investigar y entre los cuales se puede citar: nivel de madurez de la empresa, la necesidad que se tenga de análisis de la información, recursos, automatización de procesos, acceso a la tecnología de última generación, etc., por lo cual el objetivo de este estudio está sujeto a la exploración de la información existente que, le permita al investigador obtener

las herramientas teóricas para identificar una técnica de integración de los datos, enfocada en la generación de conocimiento.

Los resultados de la exploración de la información permitirán identificar una técnica o método para la integración de la información en base a un estudio comparativo minucioso de las técnicas actualmente aplicadas por las principales industrias desde la perspectiva de Big Data, lo cual agregará un importante valor para definir cuál es mejor camino a seguir para integrar la información enfocados en la generación de conocimiento en base a una propuesta apoyada en esta investigación.

La investigación descriptiva busca especificar las propiedades importantes como “aspectos, dimensiones, o componentes del fenómeno o fenómenos que se pretende investigar” (Metodología de la Investigación, 1998); este tipo de estudios requieren un grado de conocimiento por parte del investigador sobre el tema que se está analizando, con la finalidad de poder describir el fenómeno en los términos esperados, el análisis de los conceptos y la información explorada permitirá describir de forma general la técnica o método para la integración de los datos independientemente de la fuente o del tipo de dato, enfocados en la generación de conocimiento y el fácil acceso a la información.

Los resultados de esta investigación permitirán identificar una técnica o método para la integración de la información estructurada y no estructurada enfocada en la generación de conocimiento, desde la perspectiva de una solución integral de Big Data, valiéndose de la exploración, estudio de la información existente y del criterio y análisis comparativos adquiridos de la investigación y del juicio de valor que se obtendrá a través de consultas y entrevistas a expertos en el análisis de datos.

1.3. Introducción a Big Data

Big Data

Big Data es una solución que ha trascendido del ámbito tecnológico para convertirse en una estrategia de negocio u oportunidad de mejora para la calidad de vida, el tratamiento de este tema aún en la actualidad es

controversial debido al desconocimiento del significado real de Big Data, o de una estrategia de implementación, o del planteamiento de objetivos para este tipo de proyectos, y otros factores que limitan su aplicación como son la planificación, los recursos económicos, el talento humano y lo más importante los datos.

De acuerdo a una encuesta realizada por IBM en el año 2012 (ver Tabla 1) para determinar la definición de Big Data en base a la opinión de un grupo selecto de encuestados se obtuvieron los siguientes resultados:

Tabla 1. Clasificación de definiciones de Big Data
Tomado de Schroeck, M., Shockley, R., Smart, Romero Morales, & Tufano, Analytics: el uso de *big data* en el mundo real, 2012, p. 3

18%	Un mayor ámbito de información
16%	Nuevos tipos de datos y análisis
15%	Información en tiempo real
13%	Entrada de datos procedentes de nuevas tecnologías
10%	Grandes volúmenes de datos
8%	La última palabra de moda
7%	Datos de Redes Sociales

Según el estudio publicado por el (Manyika, y otros, 2011), Big Data se refiere al conjunto de datos cuyo tamaño sobrepasa la capacidad de una base de datos común para almacenarlos, administrarlos y analizarlos y al igual que otras definiciones considera las siguientes dimensiones:

- **Volumen:** Esta es una de las características más atractivas de Big Data pues, como su nombre lo indica permite el análisis de masivas cantidades de datos que sobrepasan la capacidad de la infraestructura de TI tradicional para el análisis de datos.
- **Variedad:** Característica que permite el análisis de diferentes tipos de datos, entre los cuales se consideran datos estructurados, semiestructurados y no estructurados entre los que se encuentran bases

de datos relacionales, multimedia, documentos, y otros tipos de datos complejos.

- Velocidad: Capacidad para procesar, analizar e integrar grandes cantidades de datos en tiempo real. Según el informe publicado por (Lapkin, 2012) “la velocidad designa la rapidez con que se generan los datos y con la que deben procesarse para satisfacer la demanda”.

Existen otras dimensiones que se pueden considerar dentro de la definición de Big Data como son:

- Veracidad: Esta dimensión ya se considera como parte de la definición que algunos textos y expertos manejan abiertamente. Entre ellos se encuentra IBM explica que con la veracidad se asegura que los datos provengan de fuentes fiables; considerando a la incertidumbre parte fundamental para el análisis de datos de ciertos tipos de fuentes (Schroeck, Shockley, Smart, Romero Morales, & Tufano, 2012).
- Visión: La implementación de Big Data dentro de una industria debe tener un propósito que permita delimitar el campo de análisis y un plan donde se establezca los objetivos, estrategias, costos, equipo, plazos y resultados esperados (Schroeck, Shockley, Smart, Romero Morales, & Tufano, 2012).
- Validación: Dada la cantidad de datos analizados es necesario llevar el control sobre el cumplimiento del propósito del análisis de datos para evitar el tratamiento de datos fuera del alcance de un proyecto de Big Data o datos de fuentes inconsistentes o con altos niveles de incertidumbre (Schroeck, Shockley, Smart, Romero Morales, & Tufano, 2012).
- Verificación: Tanto para el análisis de los datos como para la generación de resultados adecuados es necesario asegurarse que los datos cumplan con las especificaciones o requerimientos del análisis (Schroeck, Shockley, Smart, Romero Morales, & Tufano, 2012).

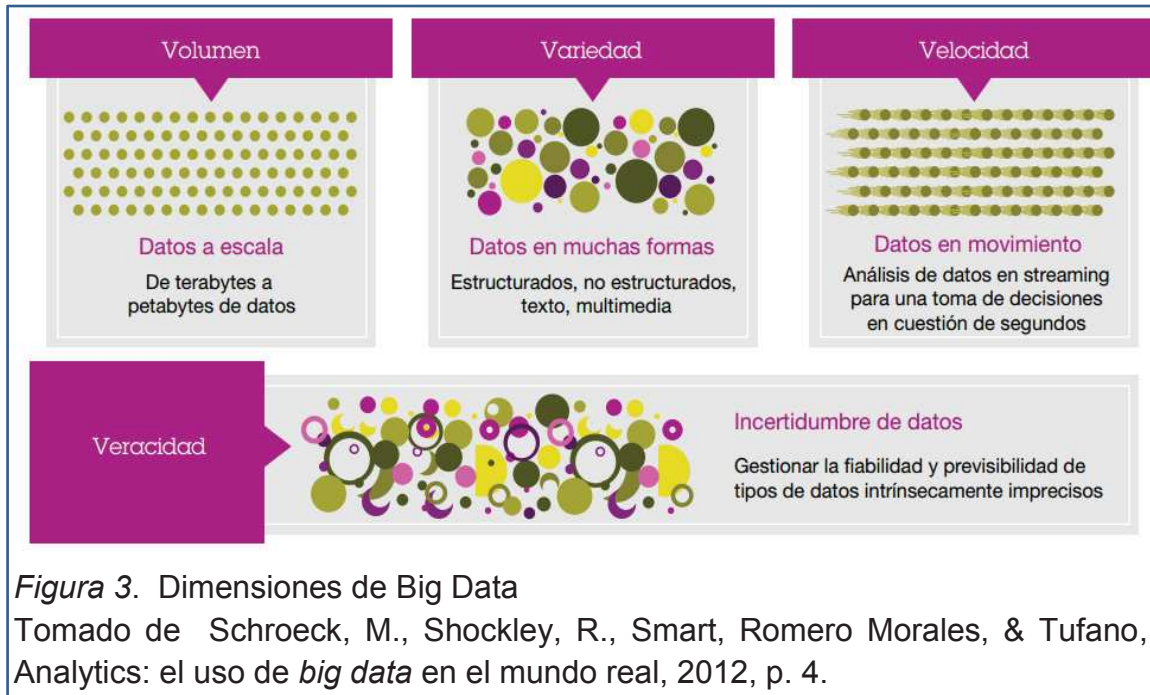


Figura 3. Dimensiones de Big Data

Tomado de Schroeck, M., Shockley, R., Smart, Romero Morales, & Tufano, Analytics: el uso de *big data* en el mundo real, 2012, p. 4.

Por tanto se puede decir que *Big Data* es una de las soluciones con mayores oportunidades de innovación en el ámbito en que se aplique considerando la velocidad con la que puede procesar, analizar e integrar gran cantidad de datos recolectados de fuentes variadas generando en la mayor parte de casos resultados en tiempo real (Schroeck, Shockley, Smart, Romero Morales, & Tufano, 2012).

Datos Estructurados

Los datos estructurados son datos relacionales, ordenados, consistentes y fáciles de manejar y ordenar en una tabla de base de datos (Loshin, 2013). La mayoría de organizaciones trabajan en la actualidad en base al análisis de datos generados a partir de sistemas transaccionales o capturados desde la web de acuerdo a ciertos niveles de ordenamiento y normalización.

Datos Semiestructurados

Los datos semiestructurados cuentan con una organización interna que facilita su administración. Como su nombre lo indica contiene características de los datos estructurados y no estructurados (Soumendra, Madhu, & Harsh, 2013), algunos ejemplos incluyen:

- “Archivos XML (*Extensible Markup Language*) y otros lenguajes de marcado,
- Correos electrónicos,
- Intercambio electrónico de datos por sus siglas en inglés EDI (*Electronic Data Interchange*), conjunto de estándares para el intercambios de información ente computadores.”

Datos No Estructurados

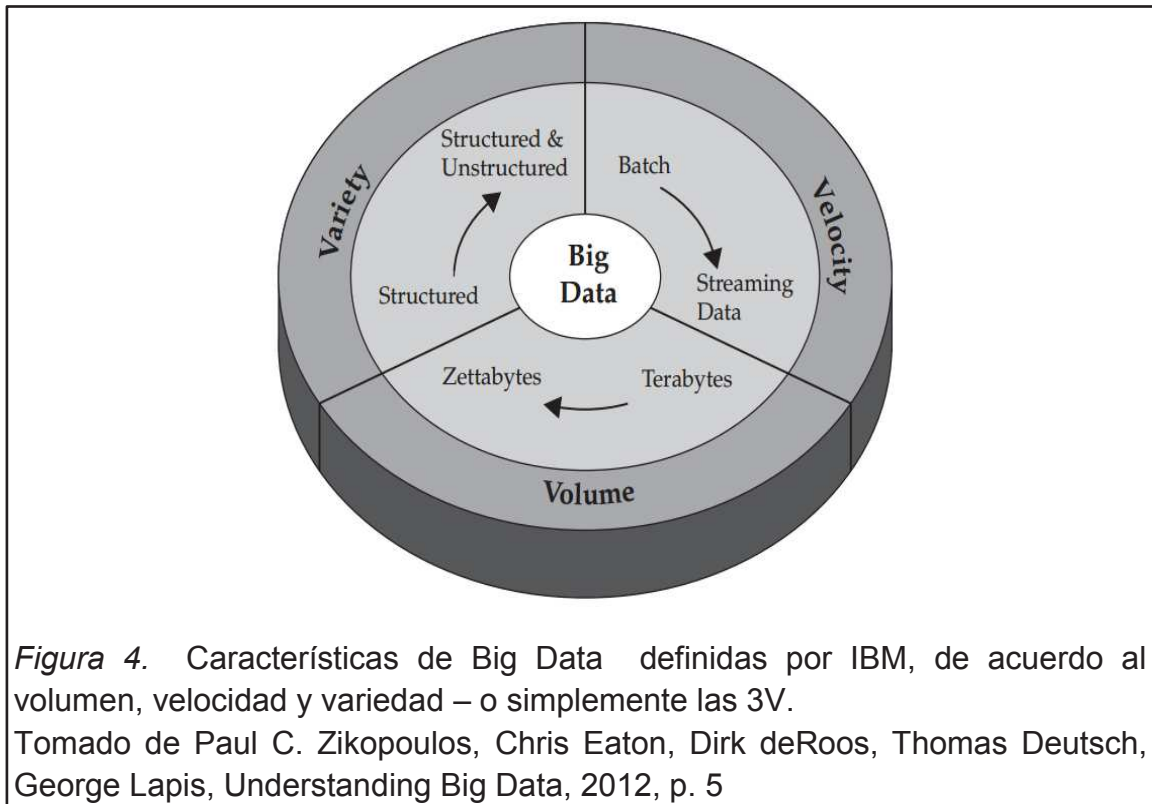
Los datos no estructurados son lo inverso a los datos estructurados. Son grandes cantidades de datos no relacionales, desordenados, cargados de texto y no son fáciles de representar en tablas tradicionales. Este tipo de datos ha sido recolectado por las organizaciones durante años y como menciona (Simon, 2013), los datos no estructurados comprenden más del 80% de todos los datos en las organizaciones y crecen 15 veces más rápido que los datos estructurados, pero, en su mayoría no están generando ningún tipo de valor al negocio.

Según (Williams, 2012) Las principales características que se puede encontrar en este tipo de datos son:

- Volumen y crecimiento superior al de los datos estructurados
- Orígenes de datos diversos
- Arquitecturas de almacenamiento que emplean herramientas de *Big Data*
- Terminología e Idioma diferentes para la invocación y generación de datos
- Complejidad para otorgar niveles de seguridad a los diferentes tipos de datos.

Respecto a los datos semiestructurados y no estructurados se debe señalar que son la base para el análisis empleando tecnología de Big Data y otro tipo de soluciones como bases de datos no relacionales e infraestructura de alta capacidad de almacenamiento.

El funcionamiento de Big Data está conformado por la estrecha relación entre cada una de las dimensiones consideradas en la *Figura 3*, y la variedad de las fuentes de los datos existentes, el tipo de análisis requerido, la velocidad para la generación de resultados y la cantidad de datos considerada para el caso de análisis.



Metadatos

Los metadatos en su definición más simple se refieren a datos sobre los datos, es decir, información estructurada acerca de los datos dentro de cualquier entorno (Krishnan, 2013).

Los metadatos son un término que se viene utilizando desde hace varios años atrás, como ejemplo de sus inicios se tiene a las bibliotecas que lo utilizan aún para la gestión de libros y otros textos, y han facilitado la gestión el control, entendimiento y preservación de los datos al pasar del tiempo.

Considerando el campo de aplicación de Big Data los metadatos constituyen una de las herramientas para la construcción de la información sobre los datos recopilados y su posterior procesamiento.

1.3.1. Importancia de Big Data

Big Data no es un nuevo concepto, pero es una de las soluciones que mayor expectativa ha generado respecto al análisis de grandes cantidades de datos en los últimos tiempos. Sectores como el gobierno, la salud y varias empresas dedicadas a la producción y venta de productos de algunos países de América y Europa se han preocupado por la investigación e indagación sobre este tema, algunos de ellos incluso han implementado tecnologías que permiten realizar este tipo de análisis, entre ellos el Gobierno de Estados Unidos de América, Organizaciones de Salud a nivel mundial y Amazon como principal ícono del comercio electrónico son algunos de los ejemplos claros del análisis e integración de datos estructurados y no estructurados (TechAmerica Foundation, 2012).

Para establecer de forma concreta la importancia de Big Data en la actualidad consideremos los siguientes principios:

- Big Data es una solución que además de permitir el análisis de los datos estructurados, soporta la integración de estos datos con datos semiestructurados y no estructurados de diferentes fuentes.
- Big Data es útil cuando se requiere analizar muestras de datos más grandes que las soportadas por las tecnologías de análisis actual.
- Big Data es una solución que permite mejorar la toma de decisiones debido a que aumenta el campo de análisis de los datos.

IBM al igual que otras marcas dedicadas a la investigación de Big Data aseguran que el 80% de los datos que una organización posee son no estructurados, por lo cual se puede entender que las decisiones basadas en el análisis de datos están en su mayoría solo considerando solo el 20% de los datos siendo estos estructurados (IBM, 2014), y en consideración a este análisis nace la importancia de la aplicación y estudio de *Big Data*.

La necesidad de incluir los datos no estructurados y semiestructurados en el análisis de datos realizado por las empresas ha llevado a dedicar tiempo y dinero para la investigación de esta tecnología como solución aplicable al

negocio y en base a la literatura estudiada se ha identificado que la importancia de Big Data no depende de la magnitud del proyecto, ni de la cantidad de datos si no del claro conocimiento del negocio y un enfoque adecuado sobre los resultados esperados para la toma de decisiones (EMC Corporation, 2011).

Las organizaciones deben contar con una estrategia para transmitir la importancia de Big Data en el negocio y demostrar de manera adecuada como se puede incrementar las oportunidades favorables para el mismo.

Big Data es un reto que deben enfrentar las organizaciones para participar de las oportunidades que ofrece el análisis de los datos. Su aplicación no debe ser vista como una obligación ni como un proyecto tecnológico; si bien su implementación requiere del uso de herramientas tecnológicas, a causa de la gran cantidad de datos que se debe extraer, almacenar y procesar, los objetivos del proyecto deben estar alineados a la estrategia de negocio y apalancados por la alta dirección.

La implementación de un proyecto de Big Data puede impactar de forma positiva o negativa en una organización, pues, si no está preparada para aceptar los cambios y renovar todas las líneas del negocio probablemente no consiga ningún beneficio de su aplicación. Una organización preparada y consciente de la importancia que tiene Big Data para el negocio enfocará los objetivos del análisis de datos, por ejemplo, en la innovación generando nuevos productos a través de la investigación y estudio de patrones obtenidos a partir del análisis de diferentes segmentos; otro enfoque puede ser el disminuir la probabilidad de fracaso de un nuevo proyecto, generar programas de inclusión económica y social en el caso de sectores gubernamentales, programas de salud para el control de epidemias y propagación de enfermedades, elevar niveles de seguridad y detección de fraude, entre otros.

Para el Ecuador la aplicación de este tipo de soluciones apoyaría a la gestión del sector privado para el desarrollo de proyectos y productos enfocados en las necesidades del mercado, y al sector público le permitiría innovar en la aplicación de sus políticas y el manejo de los recursos para extender su

cobertura a la mayor parte de individuos, al igual que conocer los segmentos minoritarios y establecer los patrones de comportamiento que estos tienen y así cubrir también sus necesidades.

Establecer segmentos de mercado, sociales y económicos, patrones de comportamiento, niveles de accesibilidad, probabilidad de éxito, predecir resultados o el impacto de una solución o producto en el mercado, reducción de costos y re-trabajo, trabajo colaborativo y elevar la curva de aprendizaje son varias de las razones por las que una organización puede decidirse a emprender un proyecto de Big Data, no obstante, cualquier emprendimiento de este tipo de soluciones requiere de investigación y el apoyo de toda la organización.

El conocimiento es otra de las razones por las que una organización puede implementar Big Data; para una empresa es fundamental conocer que es lo que quiere el cliente y ser el primero en proporcionárselo; para un Gobierno es el pilar de su gestión conocer las necesidades de los individuos con la finalidad de trabajar equitativamente y así establecer el bien común, el conocimiento genera valor al negocio, pudiendo convertirse en una ventaja competitiva y en una fuente de oportunidades.

Big Data aplicado de forma adecuada mejora la curva de aprendizaje de una organización e incrementa la base de conocimiento para la toma de decisiones. Big Data es un buen punto de partida para una organización que desea construir su base de conocimiento apoyada en el análisis de los datos.

Finalmente se debe reconocer que Big data a pesar de ser una solución presente en el mercado desde hace algunos años, todavía no ha logrado que se establezcan normas y estándares para su implementación, lo cual, ha dejado muchos vacíos respecto a su aplicación, gobernabilidad de los datos, privacidad de la información y niveles de participación de los individuos; por tanto es necesario que la organización dedique tiempo a la investigación de esta solución antes de iniciar un proyecto de Big Data.

1.3.2. Perspectiva de Big Data para la integración

El análisis de datos es una herramienta que se ha convertido en una estrategia para generar oportunidades e innovar en el entorno que se aplique, considerando sectores como la educación, la salud, el gobierno y diversas ramas de la industria (Manyika, y otros, 2011) donde se ha logrado mejoras en la calidad de servicios o generación de servicios acorde a las necesidades reales de la sociedad. A nivel mundial se generan millones de datos diarios a través de diferentes medios como se muestra en la Tabla 2:

Tabla 2. Datos generados por los medios sociales
Tomado de Ramírez Juan M., Camprubi Daniel, *Big Data: El Dorado del CRM*, (2014), p. 7.

Correo	Más de 144,8 millones al día
Facebook	Más de 648.000 trozos de contenido y de 34.000 <i>likes</i> en marcas cada minuto
Twitter	Más de 340 millones de tweess al día
Youtube	Más de 72 horas (259.200 segundos) de video cada minuto
Compras online	272.000 dólares cada día
Google	Más de 2 millones de consultas cada minuto
iTunes	Cerca de 47.000 descargas de aplicaciones por minuto
Tumblr	27.000 nuevos posts cada minuto
Instagram	3.600 fotos nuevas cada minuto
Flickr	3.125 nuevas fotos cada minuto
Foursquare	Más de 2.000 check-ins cada minuto
Páginas web	Se publican 571 cada minuto
WordPress	350 nuevas entradas cada minuto

La implementación de Big Data a diferencia del análisis de datos tradicional requiere de la aplicación de técnicas de análisis de datos que a través de una serie de algoritmos que permiten integrar los datos estructurados con datos semiestructurados y no estructurados ampliando el campo de opciones para optimizar la toma de decisiones.

En base a lo establecido por (Zikopoulos, Eaton, deRoos, Deutsch, & Lapis , 2012) y (Inc. TechRepublic, 2013) La perspectiva de Big Data para la integración de los datos considera algunos aspectos claves que ayudan a los expertos a definir el mejor momento para la aplicación de Big Data y se listan a continuación:

Análisis de datos

A diferencia del análisis de datos tradicional, Big Data como se establece en su definición permite el manejo de grandes cantidades de datos, considerando diversidad de fuentes basadas en modelos relacionales y no relacionales más conocidas como datos no estructurados.

La generación de datos no estructurados ha venido incrementándose de forma exponencial dentro de una organización en los últimos años, entre los datos no estructurados se consideran documentos, correos, documentos escaneados e imágenes, presentaciones, mensajes en las redes sociales, gráficos, audio, video, fotografías, imágenes y datos geoespaciales que integrados con la data estructurada expanden el campo de estudio para la toma de decisiones y multiplican las oportunidad de negocio.

Valoración de los datos

Es importante que las organizaciones tengan claro cuál es el valor que entregan los datos estructurados y no estructurados al negocio, pues, el almacenamiento de datos innecesarios puede convertirse en un gran problema al iniciar un proyecto de integración de datos e incluso volverse un dolor de cabeza para el negocio al generar grandes cantidades de ruido durante el análisis.

Gestionar de manera ordenada los datos disminuye la duplicidad de datos y mejora la visualización de los mismos, igualmente genera mayor y rentabilidad al negocio ya que controla la inversión en el almacenamiento de datos, pues, descarta datos que no generan valor al negocio.

El valor de los datos se determina alineando su contenido con los objetivos estratégicos del negocio, también se puede analizar su utilidad para la

generación de nuevos y mejores productos o el aporte que tienen en los procesos de negocio y en la creación de oportunidades.

Clasificación de los datos

Dada la cantidad de datos que se debe analizar es necesario clasificarlos de acuerdo al estudio que se va a realizar.

Una adecuada clasificación de los datos permite reconocer los tipos de análisis que se puede realizar, la técnica que se puede aplicar para el análisis, el tamaño de los datos, la frecuencia de generación, los tipos y orígenes de datos y lo más importante, identificar las áreas que consumen los datos.

La clasificación de los datos eleva la calidad de los resultados obtenidos en el análisis y mejora la organización de los mismos.

Aplicación de técnicas

Desde la perspectiva de Big Data existen varias técnicas que se pueden aplicar dependiendo del tipo de negocio, del tipo y formato de los datos, de la necesidad de integración de los datos (objetivos claros de la necesidad) y de los requerimientos de procesamiento en tiempo real o casi real.

A través de la aplicación de estas técnicas las organizaciones pueden establecer patrones que les permitan orientar sus objetivos al desarrollo de productos y soluciones con altos niveles de rentabilidad y poca probabilidad de un bajo impacto en el mercado, igualmente se puede proyectar y predecir el alcance que tendrá un nuevo emprendimiento e identificar nuevas oportunidad de negocio y de innovación.

Depuración de los datos

Para asegurar resultados de calidad es necesario analizar datos de calidad. Los procesos de depuración aseguran a las organizaciones la calidad de los datos pues disminuyen el ruido en los mismos y reducen los costos de almacenamiento.

La depuración de los datos también ayuda a eliminar errores producidos por la manipulación y utilización diaria de los datos, al igual que evita la distorsión de los resultados para una mejor toma de decisiones.

Desarrollo de nuevos productos

Como resultado de una adecuada integración de datos las organizaciones pueden emprender proyectos de innovación para el desarrollo de nuevos productos, o mejorar la calidad de los productos existentes en el mercado.

Sectores como el gobierno y la salud pueden utilizar *Big Data* para enfocar su gestión en mejorar la calidad de vida de sus países y generar beneficios a largo plazo para diferentes segmentos e identificar necesidades claras para el direccionamiento de recursos.

Los procesos de una organización y la plataforma tecnológica se consideran instrumentos de apoyo fundamentales para la integración de datos desde la perspectiva de Big Data y se los describe a continuación:

Procesos

Una adecuada gestión de procesos dentro de una organización ayudará a identificar los mejores escenarios o casos para el análisis de datos y generar nuevas oportunidades para el negocio e incrementar el valor que los datos entregan al mismo.

Big Data eleva las ventajas competitivas y contribuye a la objetividad de la toma de decisiones dentro de los procesos de la organización permite la innovación en la construcción y reingeniería de procesos enfocados en cubrir necesidades de un segmento específico o de las masas.

Plataforma

Uno de los principales retos que tienen las industrias para aplicar Big Data es la implementación de una plataforma que permita analizar los datos de forma rápida sin las limitaciones que generalmente se suelen presentar respecto a la capacidad de almacenamiento y velocidad de procesamiento, igualmente es necesario considerar la optimización de tiempos para administrar la plataforma teniendo en cuenta que los requerimientos para su configuración, gestión y

mantenimiento deben ser mínimos, considerando tiempos de respuesta casi imperceptibles para el negocio.

Dada la cantidad de datos que se debe almacenar y procesar es necesario contar con una plataforma de alto rendimiento y escalable, preparada para soportar el crecimiento y con la capacidad de reaccionar ante los cambios sin disminuir el rendimiento ni la calidad de sus servicios.

Muchas organizaciones están optando por soluciones cloud que desvinculan de forma parcial o total a la misma de la adquisición, administración y mantenimiento de la plataforma, sin embargo, es necesario que cada modelo de negocio conozca los objetivos con los que se realiza el análisis de los datos para decidir la mejor opción para la organización, pues, para este tipo de decisiones se debe tomar en cuenta factores fundamentales como el costo, la seguridad y privacidad de los datos.

1.3.3. Técnicas

Existen varias técnicas de análisis de la información desde la perspectiva de Big Data que permiten integrar los diferentes tipos de datos en base a los criterios establecidos en el estudio que se desee realizar y de la cantidad y calidad de los datos con los que cuente la organización.

Estas técnicas están fundamentadas en modelos estadísticos, algoritmos de computación, matemáticos y económicos, pueden ser empleadas dependiendo del volumen y la variedad de datos o del valor que se desee generar a partir de los mismos.

Considerando el dinamismo de la generación de los datos una organización puede aplicar una técnica de integración específica o establecer un conjunto de técnicas dentro de un mismo proyecto dependiendo del tipo de escenario.

De acuerdo a lo establecido por (Manyika, y otros, 2011) los principales modelos de análisis de datos aplicables en Big Data son:

Estadísticos

Las técnicas y métodos estadísticos generalmente se relacionan con algoritmos que prueban de una hipótesis considerando una población de datos referencial. La aplicación de este tipo de técnicas obliga a la organización a realizar el análisis con datos veraces ya que, al considerarse solo una muestra de datos podría afectarse gravemente el resultado generado.

Visualización de Datos

La visualización de datos permite mostrar los resultados de forma adecuada para que la organización pueda hacer uso adecuado de ellos. Existen herramientas como Pentaho, Tableau y QlickView que permiten estructurar la forma en que se va a mostrar el resultado del análisis de datos, sin embargo, es necesario que el especialista que trabaja utilizando esta técnica tenga amplio conocimiento de las necesidades del negocio al que el análisis de Big Data está apalancando.

Automatización

La automatización ha revolucionado los mecanismos utilizados para el análisis de datos, permitiendo alcanzar niveles, donde, los datos para el análisis provienen de las acciones humanas y de la interacción con las maquinas, sensores, nanotecnología entre otras.

Semántica

La semántica se basa en tres pasos para el análisis de datos: aproximación al posible significado de un objeto de datos, combinación de los resultados y organización de los datos en clases, considerando en el paso inicial la identificación del objeto de datos, un valor para el dato y un descriptor para este valor.

Este tipo de técnicas se utiliza generalmente para el análisis de datos semiestructurados y no estructurados.

Análisis Predictivo

El análisis predictivo como su nombre lo indica permite analizar un conjunto de datos estructurados o no estructurados para predecir lo que puede pasar en

uno o varios escenarios, principalmente en el área de marketing donde generalmente se utiliza este tipo de técnicas para establecer recomendaciones, perfiles y productos.

Descripción de Técnicas aplicables a Big Data

Dentro de los modelos de análisis se encuentran algunas técnicas que pueden ser aplicadas en diferentes verticales de una organización e incluso algunas de ellas son utilizadas para el análisis de pequeños conjuntos de datos en proyectos que no corresponden a Big Data; Según el informe publicado por (Manyika, y otros, 2011), las principales técnicas de análisis que se pueden aplicar en Big Data son las siguientes:

Minería de Datos.- Estas técnicas se basan en la aplicación de métodos estadísticos que combinados con modelos de aprendizaje automático permiten identificar patrones en grandes conjuntos de datos. “Estas técnicas incluyen el aprendizaje de Reglas de Asociación, Análisis de Clúster, Clasificación y Regresión” (Manyika, y otros, 2011) citado por (usun, 2011). Un ejemplo de su aplicación es la identificación de características de los mejores empleados en el caso de las áreas de Talento Humano, o, el análisis de las compras realizadas por un cliente en el supermercado para establecer un patrón de comportamiento (Tecnologías de la Información, 2009), ver ANEXO 1.

Clasificación.- Son un grupo de técnicas que mediante el análisis de datos previamente categorizados permiten determinar conjuntos de datos similares. Esta técnica puede ser utilizada para analizar el comportamiento de un grupo de clientes segmentados. A estas técnicas también se les llama de aprendizaje supervisado. Un ejemplo de aplicación se visualiza en las decisiones que toma un consumidor para comprar un producto u otro (Manyika, y otros, 2011).

Reglas de asociación.- Son un conjunto de técnicas que permiten descubrir casos comunes dentro de un determinado conjunto de datos. Estas técnicas son el resultado de la aplicación de varios algoritmos que generan y ponen a prueba todas las combinaciones posibles de las variables (Manyika, y otros, 2011).

Un ejemplo de aplicación es el análisis de las compras en el supermercado, donde se analizan los productos que son comprados conjuntamente y se pueden dar reglas como la siguiente, “si un cliente compra cebollas y vegetales es muy probable que también compre carne” (Agrawal, Imielinsk, & Swami, 1993).

Análisis de Cluster.- Es un método estadístico que permite segmentar a un conjunto de objetos en grupos homogéneos, generando conjuntos más pequeños que contengan objetos comunes entre sí. Un ejemplo de Análisis de Clúster es segmentar a los clientes de una empresa de acuerdo a sus características comunes, para ofrecerles productos personalizados dependiendo del segmento al que pertenezca el cliente (Manyika, y otros, 2011).

Regresión.- Son técnicas estadísticas utilizadas para evaluar el comportamiento del valor de una variable dependiente cuando se modifica una o más de las variables independientes incluidas en el estudio. Se utiliza para Minería de Datos. Un ejemplo de su aplicación es la predicción del volumen de ventas en base ciertas variables económicas y a los cambios en el mercado (Manyika, y otros, 2011) citado por (usun, 2011).

Redes Neuronales.- “Son modelos computacionales inspirados en la estructura y el funcionamiento de las redes neuronales biológicas - (Células y sus conexiones al cerebro) - para encontrar patrones dentro de los datos. Las redes neuronales son adecuadas para la búsqueda de patrones no lineales y su optimización” (Manyika, y otros, 2011) citado por (usun, 2011). Entre las aplicaciones prácticas de esta técnica está la identificación de los clientes de alto valor que pretenden cambiar de proveedor.

Aprendizaje Colaborativo/Simultaneo(Ensemble).- El análisis de los datos se realiza aplicando varias veces un algoritmo base que permita obtener un resultado más preciso, ya que, se cuenta con la posibilidad de analizar todos los resultados generados y seleccionar el que más se aproxime a la realidad del negocio. Esta técnica utiliza múltiples modelos predictivos para obtener

mejores predicciones de rendimiento que al utilizar los modelos predictivos individualmente. Este es un tipo de aprendizaje supervisado (Manyika, y otros, 2011).

A/B Testing.- Esta técnica compara un grupo de muestra con uno o varios grupos de prueba para determinar los cambios o procesos que se deben aplicar a la variable para lograr una mejora.

Big Data permite procesar y analizar un gran número de pruebas que garantizan que el tamaño de los grupos sea el correcto e identificar estadísticamente las diferencias existentes entre la muestra y los grupos de prueba. Un ejemplo de aplicación es determinar que imágenes, fuentes, colores o textos mejorarán el nivel de aceptación de una página Web para convertirla en un sitio de comercio electrónico (Manyika, y otros, 2011).

Análisis de Redes.- “Son técnicas utilizadas para caracterizar las relaciones entre los nodos de un gráfico o una red” (Manyika, y otros, 2011). Al analizar las conexiones de un individuo de una comunidad en las redes sociales se puede analizar como viaja la información y como está genera influencia en el resto de individuos. Una aplicación práctica de esta técnica se encuentra en la gestión de los cuellos de botella para los flujos de información de una organización.

Estadística.- “Es la parte de la Matemática que se encarga de la recolección, organización e interpretación de datos. Incluye el diseño de encuestas y experimentos; las técnicas estadísticas se utilizan a menudo para analizar las relaciones existentes entre las variables y reducir la probabilidad de errores de tipo falso negativo y falso positivo” (Manyika, y otros, 2011).

Reconocimiento de patrones.- “Es un conjunto de técnicas de aprendizaje automático, que asignan algún tipo de valor de salida, para los valores de entrada considerados en el set de datos de acuerdo a un algoritmo específico”. Dentro de este grupo se encuentra a la técnica de clasificación (Manyika, y otros, 2011) citado por (usun, 2011).

Análisis espacial.- Son un conjunto de técnicas principalmente estadísticas que permiten realizar el “análisis de propiedades topológicas geométricas o geográficas codificadas dentro de un conjunto de datos”. Normalmente estos datos son capturados gracias a los sistemas de información geográfica (GIS), por ejemplo coordenadas de longitud y latitud. Incorporando datos espaciales en regresiones espaciales se puede averiguar la correlación que existe entre los clientes que requieren adquirir un producto y su localización (Manyika, y otros, 2011).

Aprendizaje supervisado.- Las técnicas de aprendizaje automático permiten inferir una relación o función a partir de un conjunto de datos de entrenamiento. (Manyika, y otros, 2011).

Aprendizaje no supervisado.- Son un conjunto de técnicas que trabajan principalmente con datos no etiquetados y facilitan la identificación de su estructura oculta. (Manyika, y otros, 2011) citado por (usun, 2011).

Análisis de Series de Tiempo.- Utilizadas para el análisis de secuencias de datos basado en la aplicación de modelos estadísticos y procesamiento de señales, que permiten visualizar características importantes de la información a través de los valores sucesivos en el tiempo que se generan. Ejemplos de análisis de series de tiempo son “el número de pacientes diagnosticados con una enfermedad determinada todos los días” (Manyika, y otros, 2011) citado por (usun, 2011).

Mapas de Calor.- Muestran los valores de un conjunto de datos en un mapa geográfico, pero adicionalmente muestra la intensidad relativa respecto a un parámetro específico (Merino, 2014).

Visualización.- Técnicas utilizadas para el modelamiento de “imágenes, diagramas y animaciones para comunicar, entender y mejorar los resultados de los análisis de grandes datos” (Manyika, y otros, 2011) citado por (usun, 2011)

Aprendizaje automático.- También conocida como Inteligencia Artificial, esta técnica se especializa en el estudio y desarrollo de comportamientos en

computadores basados en datos empíricos, que se logran a través del desarrollo de algoritmos. Su aplicación para la toma de decisiones basado en los datos es parte importante de la evolución de las máquinas y su capacidad de autoaprendizaje que le permite identificar hasta los patrones más complejos. El procesamiento del lenguaje natural es un ejemplo de este tipo de técnicas (Manyika, y otros, 2011).

Procesamiento de lenguaje natural.- Estas técnicas se basan principalmente en la Inteligencia Artificial y en la lingüística, utilizan algoritmos computacionales para facilitar el análisis del lenguaje humano. El aprendizaje automático ejemplifica a estas técnicas. (Manyika, y otros, 2011).

Análisis de Textos.- Es una técnica de análisis, organización y recuperación de la información que tiene la finalidad de extraer conocimiento de grandes volúmenes de datos no estructurados.

Utiliza métodos de bases de datos, inteligencia artificial, aprendizaje automático, procesamiento de datos, estadística y gestión de datos para construir una estructura comprensible en base a los datos parametrizados en el análisis (Loshin, 2013).

Análisis de Sentimientos.- Consiste en “la aplicación del procesamiento de lenguaje natural, y otras técnicas analíticas que permiten identificar y extraer información subjetiva de diversas fuentes. Uno de los aspectos claves de este análisis es la identificación de los sentimientos generados hacia un producto u objeto pudiendo ser positivo, negativo o neutro”. Cuando las empresas realizan un análisis de sentimiento en las redes sociales para medir el nivel de aceptación o no aceptación de la marca frente a sus posibles consumidores o clientes utilizan técnicas de análisis de sentimientos (Manyika, y otros, 2011).

Fusión e integración de los datos.- Permiten la integración y el análisis de datos de fuentes variadas con el objetivo de identificar comportamientos de manera eficiente y que sean potencialmente más efectivos que al realizarlos en base a una sola fuente.

Una atractiva campaña de marketing puede ofrecer la integración de los datos de las redes sociales con los reportes de ventas de una organización para determinar el efecto de la misma sobre los clientes y el mercado en general (Manyika, y otros, 2011).

Procesamiento de Señales.- “Son conjunto de técnicas de ingeniería eléctrica y matemática aplicada, originalmente desarrollado para analizar las señales continuas y discretas, es decir las representaciones de las magnitudes físicas analógicas como señales de radio, sonidos e imágenes. Esta técnica se la puede utilizar para separar una señal del ruido” (Manyika, y otros, 2011) citado por (usun, 2011).

Optimización.- Son técnicas numéricas que en base a la aplicación de dimensiones objetivas permiten mejorar el rendimiento en la estructuración y rediseño de procesos y sistemas complejos. Ejemplos de aplicación son el mejoramiento de los procesos a nivel operativo y la distribución de una fábrica. Entre estas técnicas se encuentran los algoritmos genéticos (Manyika, y otros, 2011) citado por (usun, 2011).

Simulación.- “Modelado de sistemas complejos que se utilizan para la prevención, predicción y planificación de escenarios futuros. El método de Monte Carlo, por ejemplo, es una secuencia de algoritmos que se basa en un muestreo repetido al azar, es decir, miles de simulaciones, cada una basada en supuestos diferentes. Una aplicación es la evaluación de la probabilidad de alcanzar las metas financieras dadas las incertidumbres sobre el éxito de varias iniciativas” (Manyika, y otros, 2011) citado por (usun, 2011).

Modelos Predictivos.- Son un conjunto de técnicas que permiten realizar predicciones de escenarios en base a la aplicación de modelos matemáticos para disminuir la probabilidad de obtener un resultado negativo. Por ejemplo en la relación con los clientes el uso de modelos predictivos ayuda a determinar la rotación de los clientes o la probabilidad de que un cliente adquiera un producto con otro proveedor. La regresión es un ejemplo de este tipo de técnicas (Manyika, y otros, 2011).

Crowdsourcing.- Se puede definir como Colaboración Abierta Distribuida, es una técnica que generalmente se aplica en redes sociales o Internet y funciona de forma colaborativa entre comunidades semejantes o que guarden intereses similares. (Manyika, y otros, 2011). Esta técnica es un tipo de colaboración masiva que utiliza Internet 2.0 (Chui, Mille, & Roberts, 2009).

Algoritmos genéticos.- Esta técnica se inspira en la evolución natural, es decir, la supervivencia de lo más apto en el medio ambiente. “En esta técnica las posibles soluciones se codifican y se pueden combinar y mutar como “cromosomas”. Estas posibles soluciones individuales se seleccionan para la supervivencia en un modelo que determina la idoneidad o el rendimiento de cada individuo dentro de la población”. Generalmente las soluciones para los problemas no lineales se obtienen en base a esta clase de algoritmos (Manyika, y otros, 2011) citado por (usun, 2011). Algunos ejemplos de aplicación son en la gestión de tareas a través de una adecuada planificación o en las inversiones optimizando la cartera de clientes.

1.3.4. Metodologías de gestión de proyectos de Big Data

La gestión de proyectos de Big Data es fundamental para el éxito del proyecto ya que la se considera la integración de grandes infraestructuras, herramientas de software, procesos, proveedores y clientes que requieren de un estudio personalizado respecto a la planificación, equipo de trabajo, y gestión en cada una de las etapas del proyecto.

Según (Crawford, 2013) varios informes han reportado que entre el 65 y 100% de proyectos de Big Data han fallado debido a que no se han planificado de forma adecuada, los objetivos no son claros, incumplimiento en los plazos o se han quedado sin presupuesto, riesgos que están directamente relacionados con la experiencia de la organización en proyectos de este tipo y las tecnologías de la información.

La metodología para gestionar un proyecto de *Big Data* a diferencia de otro tipo de proyectos, debe basarse en un enfoque ágil, que, permita generar resultados a corto plazo, considerando que pueden o no ser de alto impacto,

pues, se debe reconocer que este tipo de proyectos puede desarrollarse en varias etapas, siendo generalmente la primera etapa la de la investigación.

Respecto al desconocimiento es necesario aclarar que a diferencia de otro tipo de proyectos existen factores que impactan directamente en el desarrollo del proyecto y que los especialistas deben considerar durante todo el proyecto. La inclusión del personal adecuado es uno de los factores más críticos que tiene el proyecto, determinar el tamaño del equipo y su conformación es vital para la adecuada gestión y seguimiento del mismo.

Establecer si el negocio está preparado para la implementación de *Big Data* y que datos se van a analizar, con que datos cuenta el negocio – análisis de valor para el negocio y Depuración de Datos - y que datos necesita recopilar; reconocer si existe la gobernabilidad de los datos en la organización; determinar cómo se va a manejar la documentación del proyecto y el relacionamiento y comunicación con el *Sponsor* y los *Stakeholders*; conocer que requerimientos de infraestructura son necesarios para implementar *Big Data*, son algunos factores técnicos y estratégicos que bridan un panorama más claro respecto a la gestión de este tipo de proyectos.

El impacto que este tipo de proyecto genera al interior y hacia afuera de la organización es un factor a considerar dentro de la gestión de proyectos para una adecuada planificación.

La investigación es fundamental, pues, no existe un manual o guías para la implementación de este tipo de proyectos, el basarse en un caso de éxito de otra organización puede no ser una muy buena opción, ya que, el modelo de negocio y las necesidades no siempre son las mismas en las organizaciones. Para emprender un proyecto de *Big Data* es necesario que quienes trabajen en el de forma directa e indirecta tengan claro conocimiento del negocio, sean expertos en el análisis de datos e integración y transmitan de forma clara los resultados generados plasmándolos sobre los objetivos empresariales.

Los *Stakeholders* del proyecto deben estar conscientes de los cambios que podría enfrentar la organización luego de la implementación de *Big Data*,

conocer respecto a la priorización de criterios y resultados dependiendo del valor que estos pueden generar al negocio, además de mantener una mentalidad abierta para el crecimiento que puede generarse respecto a la línea de aprendizaje, competitividad, oportunidades, y gestión de los datos.

Dada la magnitud del proyecto y sus implicaciones es necesario contar con el apoyo externo de especialistas por lo cual, la organización debe considerar la posibilidad de contar con proveedores y conocer un adecuado método de selección. Las especificaciones básicas para la selección de proveedores contemplan una amplia experiencia y equipo adecuado, propuestas que incluyan soluciones de *cloud* e integración con la infraestructura física tradicional, orientadas a vender la solución al cliente no basadas en un modelo estrictamente técnico y que incluyan la participación de todos los niveles del negocio.

Por otro lado es importante considerar que la organización debe tener conocimiento sobre el análisis predictivo y la gestión de datos desde la perspectiva de *Big Data*. Para esto es necesario contemplar programas de capacitación para el personal, conocimiento sobre tecnologías, estrategia y generación de valor.

Finalmente antes de seleccionar una metodología para la gestión de proyectos de *Big Data* es necesario el establecimiento de acuerdos de nivel de servicio con los grupos participantes en el proyecto para dimensionar los requerimientos y estimar los plazos para cumplir el alcance del mismo.

Metodologías

Existen varias metodologías aplicables para la gestión de proyectos en general; para el caso de este estudio se va a revisar la perspectiva para la aplicación de PMBOK y Scrum.

PMBOK “es una guía de fundamentos para la dirección de proyectos, enfocada en la colaboración y gestión adecuada de cada una de las fases establecidas en proyecto” (Project Management Institute, Inc. , 2008). Sin embargo, no es

una de las metodologías que se puede aplicar a un proyecto de *Big Data* debido a las implicaciones que requiere este tipo de proyectos.

Los proyectos de *Big Data* se fundamentan en el análisis de grandes volúmenes de datos, cada día se generan millones de ellos, por lo cual, el entorno de análisis es cambiante y esto exige flexibilidad ante los cambios en las fases del proyecto y en los plazos para la entrega de resultados.

Las limitaciones son mínimas respecto a la gestión del conocimiento y la colaboración entre los grupos participantes en el proyecto, las barreras encontradas pueden ser analizadas y derribadas si así el negocio lo requiere, soportan escalabilidad desordenada sin que esto signifique que la gobernabilidad no es necesaria y las líneas de tiempo son definitivamente cortas. La comunicación es un factor clave para mantener claros los objetivos del proyecto; de esta manera el control excesivo, plazos y recursos fijos pueden convertirse en un riesgo que el proyecto debe enfrentar restando importancia a su implementación.

El ciclo de vida de un proyecto de *Big Data* es más corto que el de un proyecto en cualquier otra área y se puede retroalimentar de la experiencia adquirida durante la investigación, es iterativo y exploratorio, aparentemente muestra poco control, poca comunicación y planificación ante lo cual la flexibilidad permite soportar la modificación de sus objetivos o el redimensionamiento de la hipótesis con un esfuerzo mínimo en la gestión de cambios del mismo.

La comunicación debe ser colaborativa e interactiva con los *Stakeholder* y la organización requiere un tratamiento especial de modo que se reconozca al proyecto como una solución estratégica para el negocio.

La seguridad es un factor clave que se debe considerar y dedicar esfuerzos en su gestión. Se requiere el establecimiento de tareas para delimitar el acceso y la visibilidad, enmascaramiento de los datos, encriptación además de pruebas que aseguren que los sistemas y la información no se encuentran expuestos al uso inadecuado.

Los tiempos de implementación deben ser flexibles pero no ilimitados, durante la planificación se debe considerar tiempos adecuados para pruebas, análisis, adquisición, implementación y estabilización de la infraestructura y las herramientas.

La organización debe dedicar recursos para la contratación de personal experto y con cualidades que agreguen valor al proyecto y ayuden a generar nuevos expertos.

Para la gestión y mitigación de riesgos la experiencia y conocimiento de *Big Data* es fundamental, pues, ayudará a establecer riesgos específicos sobre el proyecto evitando esfuerzos innecesarios sobre aspectos que no son propios del mismo.

Los proyectos de *Big Data* pueden iniciar con un alcance limitado pero pueden crecer de acuerdo a los niveles de investigación. La gestión del proyecto debe ser fácil y no desgastar recursos de forma indiscriminada al igual que considerar que el establecimiento de fases no es un requerimiento específico para el desarrollo del proyecto.

Scrum es una de las metodologías ágiles más importantes y más utilizadas en la actualidad que se basa en el trabajo iterativo y colaborativo para el desarrollo de aplicaciones y la gestión de proyectos (Crawford, 2013).

Los proyectos de *Big Data* pueden gestionarse de mejor manera con esta metodología ya que admiten el trabajo iterativo, es flexible ante el manejo de los cambios, gestión del factor humano, colaboración con el cliente y el desarrollo iterativo para la generación de resultados en proyectos con requerimientos cambiantes y plazos cortos.

Scrum permite gestionar el factor humano dividiéndolo en equipos para distribuir la carga de trabajo de una mejor manera, en base a la investigación realizada se considera que de acuerdo a las necesidades de los proyectos de *Big Data* podría establecerse los siguientes equipos:

- Equipo de Desarrollo

- Equipo de Pruebas de unidad e integración
- Equipo de control de Calidad y pruebas de funcionamiento
- Equipo de Infraestructura (Interna y Externa)
- Equipo de gestión de aplicaciones
- Equipo de redes, monitorización y virtualización
- Equipo de Soporte

Además que permite gestionar el conocimiento entre los equipos y la organización para apalancar el proyecto y fortalecer sus bases.

La coordinación de actividades muestra una adecuada gestión sobre las reuniones diarias y los resultados reportados por cada miembro del equipo, distribución adecuada del equipo para cada tarea y trabajo colaborativo, auditorías diarias del avance del proyecto para facilitar el cierre de tareas y lo más importante no se preocupa por temas irrelevantes o de mínima importancia.

Considerando las características previamente mencionadas respecto a los requerimientos para la gestión de proyectos de *Big Data* se puede concluir que *Scrum* es una metodología aplicable por una organización que desea incurrir en la implementación de este tipo de tecnología.

2. Perspectiva del Negocio

La aplicación de técnicas de análisis e integración desde la perspectiva de *Big Data* dependerá de la visión que tenga organización para determinar la necesidad de implementar este tipo de soluciones.

Visualizar con claridad el propósito del uso de *Big Data* en una organización es fundamental; establecer si se considera como la evolución del análisis de datos tradicional al análisis de los datos no estructurados o como una estrategia de cambio para obtener el valor oculto en los conjuntos de datos que aún no se han analizado, evidencia que, se ha constituido una idea base para desarrollar un proyecto de *Big Data*.

Si el propósito de la implementación no es claro y la organización está consciente de que no cuenta con una cultura fundamentada en la toma de decisiones en base al análisis de datos o con las habilidades, el liderazgo adecuado y la infraestructura adecuada se puede decir que la organización no está lista para *Big Data*.

Contar con una idea base para el inicio de un proyecto de *Big Data* es crítico para establecer los objetivos y el alcance de modo que se asegure la culminación del mismo determinando de forma adecuada los recursos necesarios para su ejecución y se reduzca el riesgo de fracaso del mismo.

2.3. Características de la organización

Las organizaciones que desean emprender en la aplicación de *Big data* deben considerar ciertos criterios que les ayudaran a determinar si están preparadas para este tipo de emprendimiento y si el negocio realmente requiere de este tipo de soluciones.

Big Data no es para todos, podría sonar como una propaganda que desestima el uso de esta solución, pero es una realidad que debe ser analizada antes de iniciar con proyectos de este tipo; entre los factores a analizar se encuentran

ciertas características que deben tener las organizaciones que desean aplicar *Big Data*.

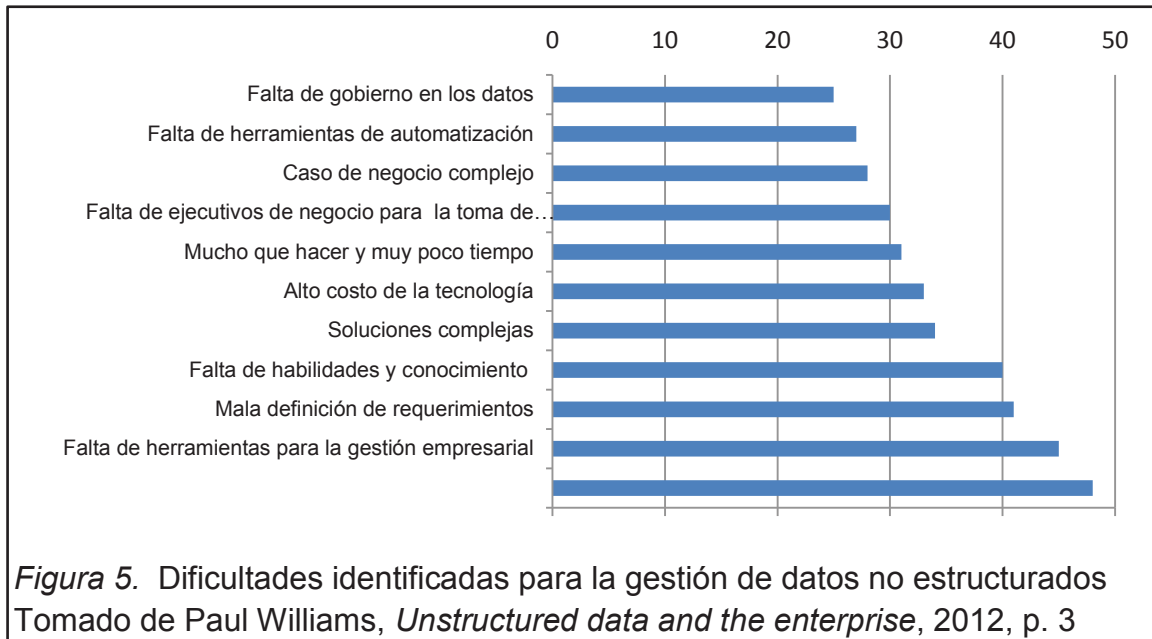
En base al estudio realizado sobre los casos de aplicación de *Big Data* se ha identificado las siguientes características que de forma general debe cumplir una organización para establecer si Big Data generará oportunidades al negocio o se convertirá en un dolor de cabeza.

Para emprender con una iniciativa de *Big Data* es necesario el apoyo incondicional de la organización, ya que si la decisión viene desde los niveles gerenciales es importante que el resto del personal esté interesado en el proyecto y vean las mismas oportunidades para el negocio. Contar con el compromiso de toda la organización es clave para soportar el desarrollo de un proyecto de *Big Data* pues en el caso de que la propuesta surja desde TI y no se logre el interés necesario de la Alta Gerencia será inútil intentar avanzar con el proyecto.

La aplicación de *Big Data* requiere de tiempo ya que, no es un proyecto que generará un producto específico y durante su ejecución puede presentarse un cambio parcial o total de los objetivos y el alcance del mismo con implicaciones directas sobre los plazos y recursos estimados inicialmente. Pueden presentarse situaciones en las que el proyecto de *Big Data* se desarrolle en fases donde la primera solo establezca la investigación de la solución para lo cual la organización debe ser paciente para lograr resultados satisfactorios.

Los emprendimientos en Big Data deben considerar la existencia de procesos como una de las claves para el éxito, pues, ayudan a identificar las áreas en las que podría mejorar la organización, mejora e innovación de productos, falencias y puntos de optimización de los procesos, áreas de impacto y oportunidades para el negocio. Contar con una adecuada estructura por procesos acelerará la identificación del propósito y enfoque del negocio para la aplicación de *Big Data*.

La Figura 5 muestra las dificultades durante el análisis de datos no estructurados



Los datos son la base para determinar si realmente una organización requiere de la aplicación de *Big Data* o simplemente innova las soluciones de BI o del incremento en las características de la infraestructura. La organización debe analizar si además de contar con grandes cantidades de datos, las soluciones implantadas ya no soportan los requerimientos de análisis del negocio y existen datos que no están generando valor, sin embargo, implican inversión para su almacenamiento.

Los recursos y su administración son importantes para la gestión de un proyecto de *Big Data*, sin embargo, respecto al equipo de trabajo la organización debe contemplar la búsqueda de expertos en el análisis de datos, manejo de herramientas de análisis y procesamiento de datos al igual que personal con conocimiento y experiencia en la comprensión de los resultados para exponerlos a la alta gerencia. Identificar el personal adecuado contribuirá de manera estratégica con el desarrollo del proyecto pues ayudará a determinar los conjuntos de datos que se debe analizar dependiendo del caso de estudio establecido; la comprensión de los resultados y su alineación con las estrategias del negocio es clave para establecer las oportunidades de negocio encontradas.

La organización debe ser consciente que contar con un equipo de expertos en *Big Data* implicará inicialmente una fuerte inversión, la misma que, debe ser

retribuida con la generación de nuevos expertos a partir de conocimiento compartido de los expertos hacia el resto del equipo que trabaje en el proyecto.

El conocimiento es uno de los productos que indudablemente obtendrá una organización al finalizar un proyecto de *Big Data*; concientizar a la organización del valor que tiene la generación y adquisición de conocimiento es un trabajo que se debe realizar e implica tiempo y recursos que se deben invertir para fomentarlo en todos los niveles organizaciones. Por lo cual es importante que esta cultura ya exista en las organizaciones que van aplicar *Big Data*, pues, al hacerlo durante la planificación y ejecución del proyecto enfocará el esfuerzo en lograr este objetivo y descuidará el verdadero propósito del proyecto.

Dentro de las fortalezas relacionadas con el conocimiento la organización debe contar con experiencia previa en el manejo de herramientas de Inteligencia de Negocios y una cultura orientada a la toma de decisiones basada en el análisis de datos.

Determinar si la infraestructura con la que cuenta la organización es adecuada para almacenar y procesar grandes cantidades de datos incluyendo datos no estructurados es vital para establecer si la organización debe o no incluir recursos para la adquisición de una nueva infraestructura o, mejorar las características de la que posee. Analizar soluciones de *Cloud* para el almacenamiento puede abaratar costos y permitir el desarrollo de propuestas más atractivas para la alta gerencia; ante este último punto se debe aclarar la importancia de los acuerdos de nivel de servicio y las políticas de seguridad y privacidad de los datos.

Finalmente la organización debe tener presente que, si al terminar el proyecto no se obtienen los resultados esperados, no ha fracasado en la aplicación de *Big Data*, pero si ha fortalecido el conocimiento respecto a esta tecnología en el modelo de negocio sobre el que se haya trabajado. Al igual que cuenta con la oportunidad de continuar con la investigación para lograr el análisis e integración de los datos desde la perspectiva de *Big Data*.

2.4. Aplicación de *Big Data* en la organización, perspectiva de negocio

Big Data es una de las soluciones de negocio más innovadoras y atractivas en la actualidad; según información de algunas firmas como *Gartner* y *Forrester* la implementación de *Big Data* ha crecido en un alto porcentaje e incluso empresas como Yahoo, Google y Facebook basan el análisis de datos en estas soluciones (Hopkins, 2013).

La generación de grandes cantidades de datos en especial no estructurados, las limitaciones de la infraestructura de procesamiento y el apareamiento de soluciones de *Cloud* han enfocado la mirada de las organizaciones en la implementación de *Big Data* para ampliar el campo de análisis de los datos y generar valor a partir de la integración de los datos no estructurados.

Para iniciar la implementación de un proyecto de *Big Data* a diferencia de otro tipo de proyectos de negocio y tecnología la organización debe considerar el caso de estudio, los datos, los recursos y la iniciativa del negocio que son requerimientos clave para el análisis de datos desde la perspectiva de *Big Data*.

La clave del éxito de un proyecto de *Big Data* no puede describirse en un listado que cualquier organización puede seguir y tener resultados satisfactorios, pero si se puede establecer pautas para que una organización visualice si es correcta la decisión de emprender o no en un proyecto de *Big Data*.

Al principio una organización debe establecer un propósito de implementación al que se le debe sumar el enfoque primario de negocio o tecnológico con el que se desarrollará el proyecto.

El enfoque de negocio puede establecer la iniciativa en los procesos organizaciones pensando en cambios de estructura, determinar los interesados del proyecto y las áreas de impacto directo e indirecto del proyecto; o, enfocarse en el capital humano, identificando si existe al interior de la organización o es necesario contratar nuevo personal.

El enfoque tecnológico se basa estrictamente en la infraestructura para soportar el análisis de grandes cantidades de datos, considerando características del almacenamiento y requerimientos de red y equipos necesarios.

La organización debe delinear una estrategia para la aplicación de *Big Data* considerando de forma general 3 etapas a las deberá enfrentarse durante el desarrollo del proyecto.

Las etapas identificadas son la planificación, la ejecución y el seguimiento (Post-implementación) que se apoyaran en una estrategia de comunicación para mantener informada a la organización de lo que sucede en el proyecto, las actividades desarrolladas, los avances logrados y los resultados o productos generados. Ver *Figura 6*.



Figura 6. Etapas para la aplicación de *Big Data*

Comunicación

La comunicación es un pilar muy importante para los proyectos de *Big Data* no hacerlo conllevará a la pérdida de tiempo y recursos e inclusive al fracaso del mismo.

Durante la planificación es esencial comunicar la visión que tiene el negocio respecto a la aplicación de *Big Data*, los objetivos estratégicos y su alineación con las metas del negocio; el caso de aplicación debe ser establecido por toda la organización de modo que se identifique a los responsables, interesados y afectados del proyecto. Mantener una comunicación adecuada en esta etapa

ayuda a definir de mejor manera el tiempo y los recursos necesarios para su ejecución.

Comunicar los hallazgos encontrados durante la fase de ejecución retroalimenta el conocimiento de la organización sobre *Big Data* y su aplicación efectiva en el caso de estudio. La mejora continua es otro beneficio que proporciona la comunicación en esta etapa ya que, en caso de encontrar ineficiencias en la planificación es posible optimizar y renovar ideas de acuerdo a la experiencia adquirida por la organización y el trabajo colaborativo de todos sus integrantes.

Al finalizar el proyecto se debe comunicar a toda la organización de los éxitos logrados y/o los objetivos incumplidos (en los casos que se presenten situaciones de incumplimiento de objetivos por hipótesis mal elaboradas o limitaciones de recursos y tiempo), con la finalidad de alentar a la organización a dar continuidad a la implementación de otros proyectos de *Big Data* y fomentar en la organización una cultura orientada a la colaboración y la toma de decisiones basadas en el análisis de datos.

Etapa 1: Planificación.

En esta etapa se establece los objetivos del proyecto, el alcance y la inversión en tiempo y recursos que realizará la organización para lograr un proyecto de calidad.

Según (Desouza, 2014) para la elaboración de una adecuada planificación la organización debe considerar dos factores clave:

Definir un caso de aplicación

Identificar un caso de aplicación para el análisis de datos desde la perspectiva de *Big Data* es la base para establecer el alcance del proyecto, los riesgos a los que se puede enfrentar en determinadas situaciones, los desafíos que debe enfrentar la organización y el nivel de complejidad del caso seleccionado y de los conjuntos de datos a analizar.

Contar con un caso de aplicación ayuda a los expertos a evidenciar datos innecesarios para el análisis al igual que fuentes que se deben adquirir; además, el trabajo con expertos en análisis de datos y *Big Data* contribuye enormemente a la investigación que se desarrolla en busca de resultados satisfactorios.

Conocer los datos que van a incluirse en el análisis determina los niveles de seguridad y privacidad que la organización debe establecer para su adecuada manipulación y uso.

Big Data como estrategia de negocio

Si la organización ha decidido emprender en un proyecto de *Big Data* es necesario que la iniciativa cuente con el auspicio de la Alta Gerencia o de un patrocinador con la autoridad suficiente para la toma de decisiones en momentos de alto riesgo y criticidad del proyecto. Este requerimiento no debe ser opcional, pues, a diferencia de otro tipo de proyectos la falta de interés recaerá en el fracaso del proyecto o el abandono del mismo.

La alineación de los objetivos del proyecto con los objetivos estratégicos de la organización es fundamental para demostrar la importancia de su aplicación; un claro enfoque de aplicación para el mejoramiento o innovación de los procesos, estructura organizacional, infraestructura y productos es un excelente punto de partida.

Los Indicadores clave de rendimiento (*KPI* por sus siglas en inglés), permitirán establecer patrones de medición para evidenciar los avances del proyecto contribuyendo a la adecuada gestión de los recursos.

Como parte de la estrategia de negocio se debe contemplar la elaboración de un Plan de mitigación de riesgos dado el nivel de complejidad del proyecto; la organización debe considerar que la salida de un miembro del equipo puede desequilibrar la investigación o incluso detener el avance del proyecto. También es importante tener en cuenta las acciones en caso de disminución de recursos o cambios en los objetivos establecidos en la etapa de planificación.

Etapa 2: Ejecución

Como parte de la ejecución del proyecto de *Big Data* se debe contemplar algunos factores de los que dependerá el éxito o fracaso del proyecto.

Entre los factores a considerar están los siguientes:

Equipo de Trabajo

La organización debe considerar la integración al equipo del proyecto de expertos en datos; este personal se dedicará exclusivamente al análisis de los datos y de los resultados generados para presentarlos a la alta gerencia. No se recomienda trabajar con personal sin experiencia, pues, de esto dependerá la calidad del análisis y de la investigación.

Trabajo colaborativo

Un proyecto de *Big Data* afectará varias áreas independientemente del objetivo general establecido en la planificación. El trabajo colaborativo es importante para evitar los obstáculos y minimizar los riesgos para la organización a causa de la resistencia al cambio del personal.

Por otro lado con la finalidad de obtener todo el valor que *Big Data* puede entregar al negocio, TI debe convertirse en un proveedor de servicios, donde se incluya el uso de las herramientas de BI y la aplicación de modelos y metodologías para que los dueños de los procesos experimenten con ellos y extraigan el valor de los datos.

Gobierno de los datos

Se deben establecer políticas para el manejo de los datos, manipulación y usabilidad. Las políticas deben contemplar niveles de accesibilidad y establecer los lineamientos para considerar un conjunto de datos dentro del análisis.

Los niveles de seguridad y privacidad deben ser parte del contenido de las políticas para evitar la exposición, pérdida o corrupción de datos a causa de ataques cibernéticos, manejo irresponsable y daños en la infraestructura.

Infraestructura

El análisis de la infraestructura debe partir de la revisión de las características de la arquitectura con la que cuenta la organización frente a los requerimientos del análisis de datos que se va a realizar.

Es necesario determinar las características de la infraestructura requerida en base al caso de aplicación establecido. Las propuestas de infraestructura pueden incluir opciones como la adquisición de una infraestructura privada, la combinación de infraestructura de la organización con integración a una plataforma *cloud* para el procesamiento y almacenamiento o definitivamente una infraestructura completamente en *cloud* dependiendo de la capacidad económica y las necesidades del negocio y los niveles de seguridad que se desee proporcional a los datos.

Etapa 3: Post – implementación

Al finalizar la implementación de un proyecto de *Big Data* inicia una etapa de evaluación y seguimiento para evidenciar los aciertos y fracasos en la toma de decisiones al igual que nuevas oportunidad para la aplicación de *Big Data*.

El impacto del proyecto sobre los procesos de la organización, y el cumplimiento de los KPIs es otro factor de análisis en esta etapa, pues, determinaran el nivel de afectación que este tuvo sobre el negocio.

Identificación de nuevos proyectos de *Big Data*

Los resultados generados al finalizar el proyecto permitirán a la Organización identificar nuevas oportunidades de aplicación para *Big Data*; no obstante, es recomendable considerar un espacio de tiempo entre un proyecto y otro, con la finalidad que el equipo asimile el conocimiento y habilidades adquiridas.

Conocimiento generado

El conocimiento es un producto intangible que obtiene la organización desde el inicio del proyecto; la investigación y el estudio de los resultados del análisis de datos elevan el valor que el conocimiento puede aportar al negocio, contar con personal experto y capacitado eleva las oportunidades para emprender proyectos exitosos.

Cada etapa del proyecto debe contar con la documentación respectiva, este activo de información será de utilidad para próximos emprendimientos como insumo para la toma de decisiones y el desarrollo del proyecto.

En la *Figura 7* se muestra un resumen de las consideraciones y resultados de cada una de las etapas mencionadas para lograr los objetivos de un proyecto de *Big Data*.

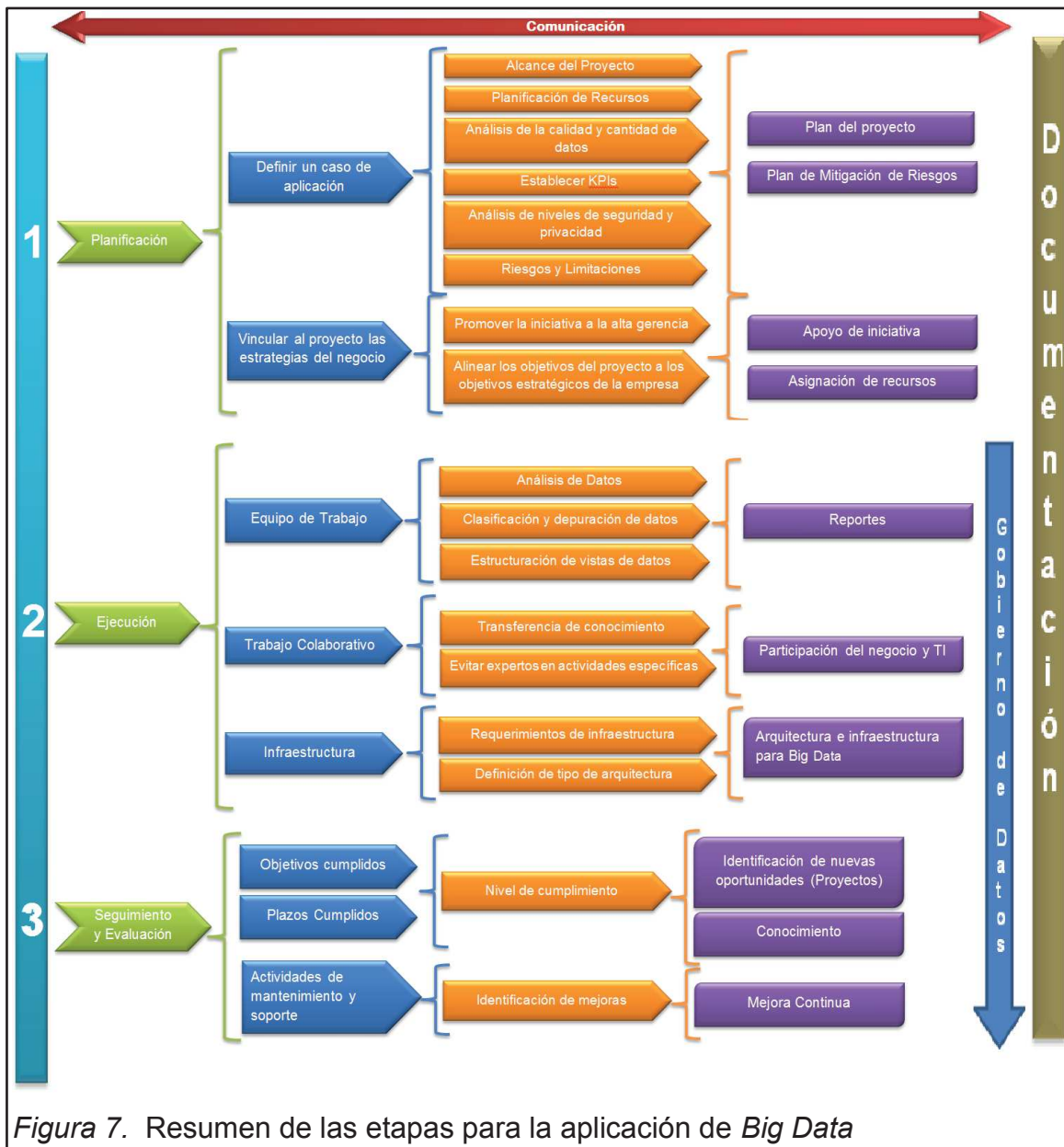


Figura 7. Resumen de las etapas para la aplicación de *Big Data*

2.5. Generación de Conocimiento desde la perspectiva de *Big Data*

Uno de los activos de información más importantes es el conocimiento, es fundamental para investigar nuevas áreas, desarrollo de proyectos estratégicos e innovación; en la actualidad el conocimiento es un factor determinante para la toma de decisiones y es el que mayor valor genera a la organización si se gestiona de manera adecuada (Rajpathak & Narsingpurkar, 2013).

Es necesario entender la criticidad en la gestión del conocimiento, en especial cuando la organización basa su toma de decisiones en el análisis de datos y más si ha iniciado un emprendimiento relacionado con Big Data.

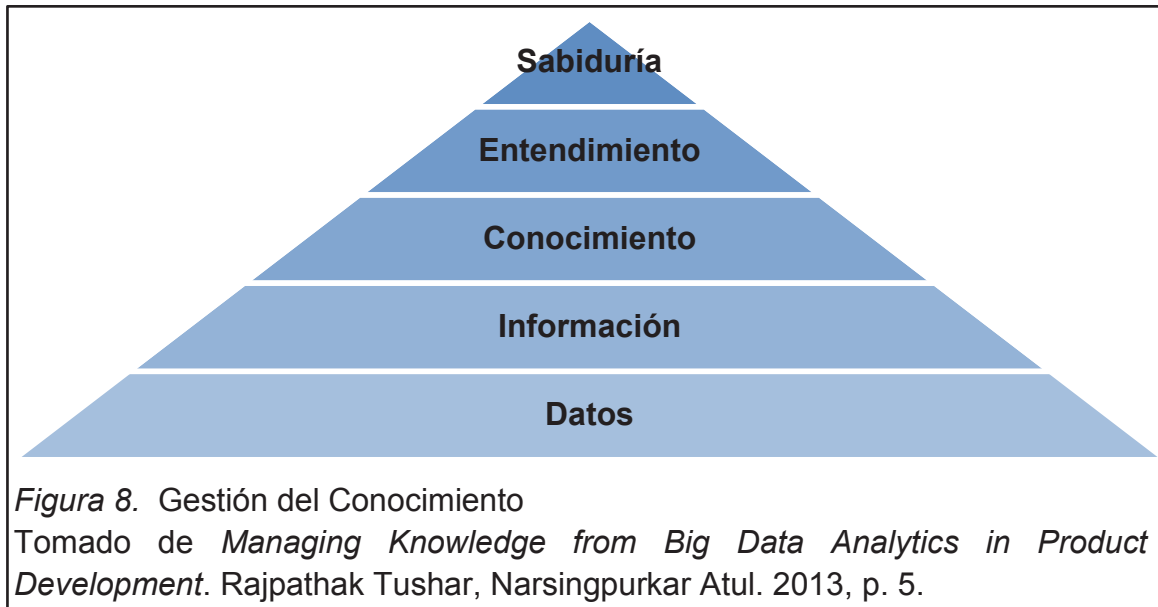
La aplicación de nuevas tecnologías en una organización al igual que cualquier proceso de innovación implica los siguientes beneficios intangibles para el negocio:

- Crecimiento de la base de conocimientos
- Adquisición y mejora de habilidades en el recurso humano
- Documentación útil para el estudio de oportunidades de mejora

Big Data incrementa el valor del conocimiento que adquiere una organización durante la aplicación de esta tecnología, y la innova proporcionando información en base al exhaustivo análisis de datos, incluyendo datos no estructurados; se considera innovador el proceso de generación de conocimiento debido a la cantidad de datos procesados y a la participación del negocio para determinar las estrategias de integración y el modelamiento de patrones para la toma de decisiones en tiempo real o casi real.

A diferencia de otro tipo de proyectos la generación de conocimiento nunca se detiene, al igual que el campo de análisis se expande cada día; los criterios para la evaluación y entendimiento del conocimiento adquirido con miras a la generación de sabiduría deben ser precisos y alinearse a las estrategias del negocio; una organización no identificará los beneficios de la adquisición de conocimientos que no aportan valor al negocio.

Como se muestra en la *Figura 8* la generación de conocimiento a partir de los datos debe pasar por diferentes fases para convertirse en sabiduría es decir generar expertos en el área de estudio.



El nivel de entendimiento y experiencia del equipo dedicado al análisis e integración de los datos es fundamental para la generación de información que al estudiarla con criterios adecuados permitirá reconocer las relaciones existentes con el objetivo de negocio.

A partir del conocimiento generado se debe trabajar en la identificación de patrones que sean de utilidad para el establecimiento de nuevas oportunidades para el negocio y el descubrimiento de principios y fortalecen la estructura, procesos y habilidades de la organización.

Una organización interesada en la generación de conocimiento debe preocuparse por instituir una cultura colaborativa entre el personal y garantizar la disponibilidad y el acceso a la información. La motivación entre el personal para que comparta sus experiencias y habilidades es fundamental para incrementar el equipo de expertos y ayuda a la descentralización de funciones.

La generación de conocimiento en una organización es muy importante para incrementar sus habilidades en la toma de decisiones y elevar sus ventajas competitivas, sin embargo, es necesario considerar estrategias para la

transferencia de conocimiento con la finalidad que lo utilicen y aprovechen o lo intercambien con otros para mantener un proceso de mejora continua.

Dado el entorno cambiante de los datos se debe contemplar procesos para la integración entre los patrones identificados durante cada proceso de análisis.

La generación de conocimiento es un reto para toda organización; saber cómo tomar las mejores decisiones puede determinar el éxito o fracaso de una organización al igual que el de sus líderes. *Big Data* genera conocimiento especializado que contribuye a la identificación de oportunidades para el negocio y al fortalecimiento de las habilidades en diferentes aristas de la organización que guardan una estrecha relación con las dimensiones establecidas en la definición de *Big Data*.

La investigación ha permitido identificar que los principales puntos de impacto son: el área de mercado, experiencia y habilidades del personal, conocimiento sobre tecnología y procesamiento de datos.

Para el mercado el conocimiento proviene del análisis de grandes cantidades de datos provistos por el cliente y la competencia a través de diferentes medios incluidas las redes sociales y otro tipo de interacciones con medios digitales e Internet. La variedad de los datos analizados proporciona un amplio campo de estudio para el análisis de los segmentos de mercado, tendencias y el desarrollo de ideas innovadoras de alto impacto y con un valor agregado importante.

La colaboración y la participación de nuevas experiencias incrementan las habilidades del personal de la organización para la toma de decisiones y el diseño y creación de ideas innovadoras.

El conocimiento que adquiere la organización sobre tecnología se relaciona con la investigación sobre la solución que se va a implementar, estándares, entornos de aplicación, diseño y construcción de conjuntos de datos, seguridad y privacidad de los datos, procesamiento de datos en tiempo real, costos y rentabilidad de las herramientas y el valor que pueden aportar al negocio.

El procesamiento de datos genera conocimiento a partir del análisis de los conjuntos de datos, la aplicación de técnicas de integración, pruebas, verificación y validación de los datos que aportan valor a la organización para el desarrollo de mejores prácticas y nuevos servicios.

2.6. Experiencias con Big Data

La gestión de los datos ayuda a la transformación de una organización. Las organizaciones han venido capturando y generando una gran cantidad de datos relacionados con los clientes, los productos y los procesos que han permitido generar nuevas oportunidades e identificar ideas innovadoras para el negocio.

Las soluciones proporcionan resultados satisfactorios para la toma de decisiones al igual que las herramientas facilitan el trabajo para alcanzar los objetivos del negocio; sin embargo, el entorno cambiante de los datos, la generación masiva de los mismos y el acceso a la tecnología y al Internet en todos los ámbitos sociales y económicos evidencian la necesidad ampliar el campo de análisis para obtener mayor valor de los datos incluyendo los datos no estructurados. Los factores que generalmente se consideran para la toma de este tipo de decisiones son:

- Alta inversión en el almacenamiento de los datos y poca usabilidad de los mismos
- Limitaciones de las bases de datos relacionales, respecto a la gestión de los datos no estructurados
- Limitaciones de la tecnología para el procesamiento de grandes cantidades de datos en tiempo real o casi real
- Dificultad para integrar los datos estructurados con los no estructurados
- Complejidad en la construcción de los conjuntos de datos para el análisis

Existen dos sectores que a nivel mundial han alcanzado excelentes resultados con la aplicación de *Big Data* en su entorno como es el gobierno (Manyika, y otros, 2011) y el turismo (Ministerio de Turismo del Ecuador, 2012).

2.6.1. Gobierno

La gobernabilidad de un país es una tarea compleja donde, una las principales responsabilidades se enfoca en la gestión y generación de recursos para contribuir al desarrollo de la nación.

Causas como la limitación de recursos, el endeudamiento externo, las excesivas importaciones, la ausencia de inversión y el desarrollo industrial limitan el crecimiento y la productividad del gobierno frente a las exigencias de servicios de calidad.

El acceso a la tecnología, el Internet y las redes sociales han permitido que la población de un país interactúe de forma directa con sus gobiernos, no obstante, los datos que se manejan a través de estos medios no se están aprovechando mayormente para elevar el nivel de gobernabilidad en un país.

Las instituciones gubernamentales tienen el acceso a grandes cantidades de datos, como, datos de texto y numéricos, en algunas áreas como la salud se manejan frecuentemente datos multimedia (imágenes y videos); cuentan con documentos físicos, acceso a redes sociales y la posibilidad de relacionarse con gobiernos de otros países que ya se encuentran aprovechando el valor que los datos les pueden proporcionar y aplicándolos para elevar la productividad y mejorar los servicios.

La salud, la seguridad, la educación y la generación de empleo son servicios que se beneficiarían si el gobierno se preocupa del análisis de los datos generados por la población y a partir de los patrones identificados desarrolla políticas y estrategias gubernamentales que permitan mejorarlos.

Utilidad de *Big Data* en el sector público

El informe presentado por (Manyika, y otros, 2011) describe cinco utilidades de *Big Data* para un gobierno entre las cuales se consideran:

- Transparencia
- Descubrimiento de necesidades a través de la experimentación, pruebas sobre la variabilidad y elevar su rendimiento

- Segmentar la población para personalizar las acciones
- Remplazar o soportar las decisiones humanas con el desarrollo de algoritmos automatizados
- Innovar nuevos modelos de negocios, productos y servicios

Se puede construir un gobierno transparente facilitando el acceso a los datos y a la información el disminuir los trámites burocráticos y los tiempos de espera ayuda además a mejorar su eficiencia; el uso de Internet, mensajes de texto y las redes sociales son una buena opción para disminuir los tiempos y mantener una relación más estrecha con los ciudadanos.

Permitir el acceso en línea a la información gubernamental, transparente la gestión y ayuda a reducir los tiempos que se tardan en buscar la información al interior de una empresa pública o entre las empresas de gobierno. Para lograr la implementación de estas facilidades el gobierno debe establecer una cultura colaborativa y políticas que permitan destruir las barreras existentes.

Experimentar con los datos permite a una empresa gubernamental descubrir variaciones en su rendimiento o productividad al igual que medir y comparar el desempeño de las unidades que la conforman para a partir de esta información establecer estrategias que mejoren el desempeño, eleven la calidad de los servicios y su productividad.

En la actualidad, la segmentación de los grupos sociales ayuda al sector público al estudio para la distribución de los recursos y al desarrollo de servicios; sin embargo, *Big Data* permite profundizar en los patrones obtenidos para trasladar el enfoque a la personalización de las acciones y así cubrir las necesidades de los grupos más pequeños, este es un principio aplicado generalmente por el sector privado, como resultado de estos estudios el Gobierno puede aumentar la eficacia, la eficiencia y satisfacción de los ciudadanos, con la mejora de los servicios, el direccionamiento adecuado de recursos, y la disminución de las brechas sociales y económicas.

La automatización de los servicios y el uso de la tecnología es un apoyo estratégico para un gobierno, pues, disminuye los errores e inconsistencias en

los datos y el porcentaje de incumplimiento al contar con el acceso permanente a información precisa y objetiva para la toma de decisiones.

El análisis e integración de los datos permite a un gobierno identificar oportunidades para innovar o mejorar sus servicios, el acceso a la información facilita una cultura de trabajo compartido y la generación de conocimiento a través del aporte de ideas y sugerencias que pueden ser aprovechadas para beneficio de la sociedad y elevar la calidad de la gestión pública.

Los beneficios que un gobierno puede obtener del análisis e integración de los datos son varias, pues, el acceso a la información, la identificación clara de las necesidades, la transparencia en su gestión, y el cumplimiento de sus compromisos le permitirá mantener a los ciudadanos informados, innovar sus ideas para generar nuevos modelos de negocio, responder oportunamente ante un requerimiento individual o colectivo y el equilibrio social entre el gobierno y su nación.

En Ecuador el uso de Internet está dirigido principalmente a la consulta de información, comunicación como por ejemplo acceso a redes sociales, educación y aprendizaje y en forma general por razones de trabajo; como se muestra en las *Figuras 9a* y *9b* que permiten visualizar que a nivel nacional en el año 2013 el porcentaje más alto de uso de Internet estaba enfocado en la consulta de información alcanzado un 32%.

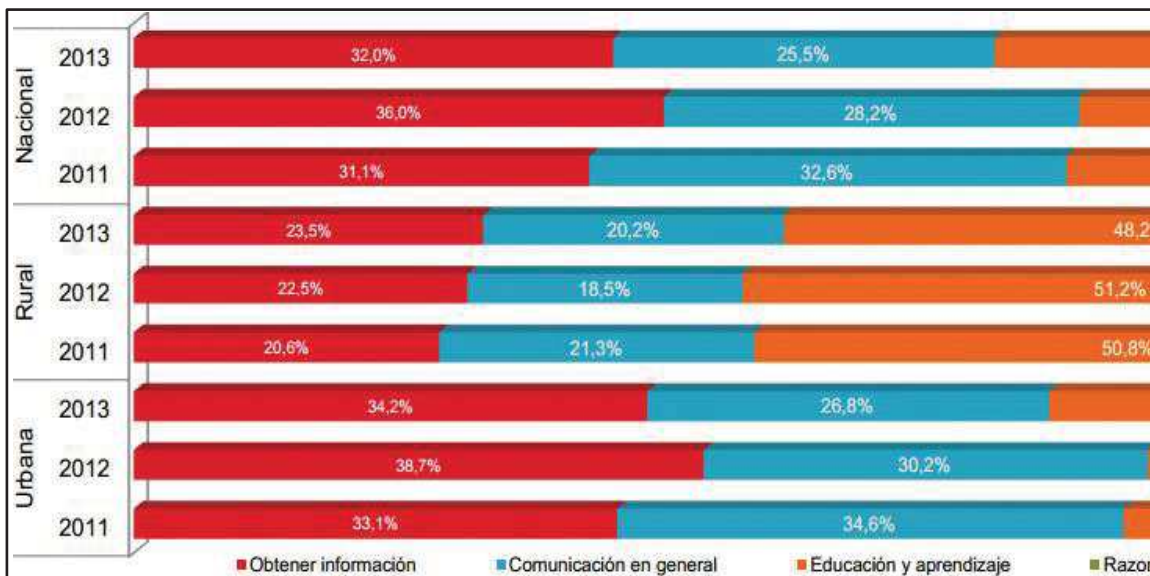


Figura 9a. Razones de uso de Internet por área en el Ecuador

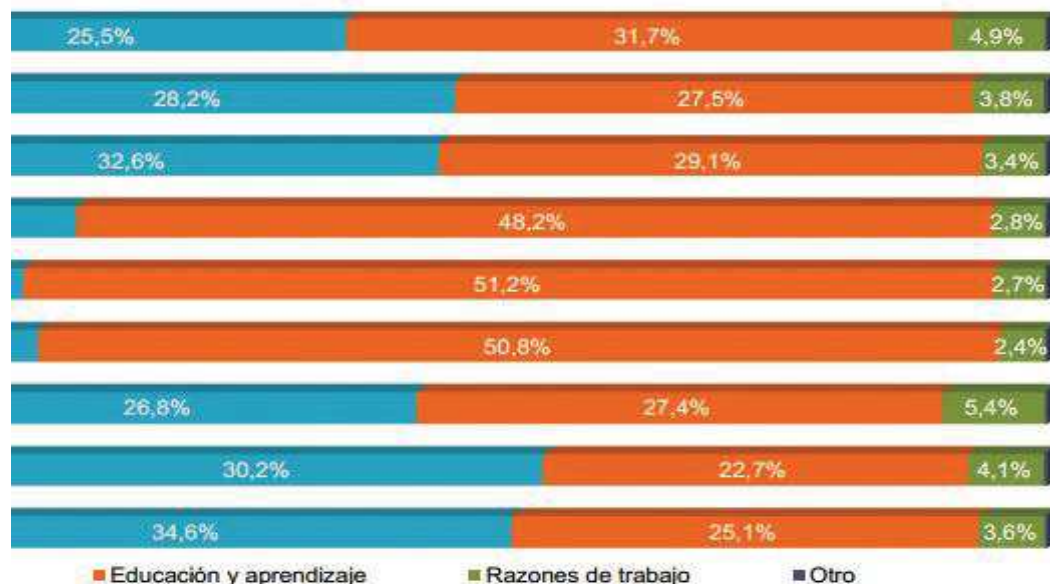


Figura 10. Razones de uso de Internet por área en el Ecuador

Tomado de Instituto Nacional de Estadísticas y Censos, INEC, Tecnologías de la Información y Comunicaciones (TIC'S) 2013, www.ecuadorencifras.gob.ec

- a. Porcentajes de uso del Internet en Ecuador a nivel urbano, rural y nacional de los períodos 2011-2013. El análisis muestra que los ecuatorianos utilizan el Internet para consulta Obtener información y Comunicación en general
- b. Porcentajes de uso del Internet en Ecuador a nivel urbano, rural y nacional de los períodos 2011-2013. El análisis muestra que los ecuatorianos utilizan el Internet para educación y trabajo entre otros

Para el Ecuador la aplicación de *Big Data* será una acción estratégica que permitirá apalancar los objetivos de su gestión (Secretaría Nacional de Planificación y Desarrollo, SENPLADES, 2012):

- “Contar con nuevos esquemas de generación, distribución y redistribución de la riqueza;
- Reducir la vulnerabilidad de la economía ecuatoriana;
- Eliminar las inequidades territoriales;
- Incorporar a los actores que históricamente han sido excluidos del esquema de desarrollo de mercado”.

El impacto que el Gobierno desea fomentar se centra en la generación de recursos, la productividad, el cumplimiento de compromisos y la equidad; acciones que de apoyarse en el análisis de datos ampliarán la visión y optimizarán la toma de decisiones de forma objetiva.

Acciones como las mencionadas han permitido que el acceso a Internet se incremente en los sectores urbanos alcanzando un 37% en el año 2013 y en los sectores rurales hasta un 9%. La *Figura 10* muestra el comparativo de acceso a Internet en el período 2010 al 2013.

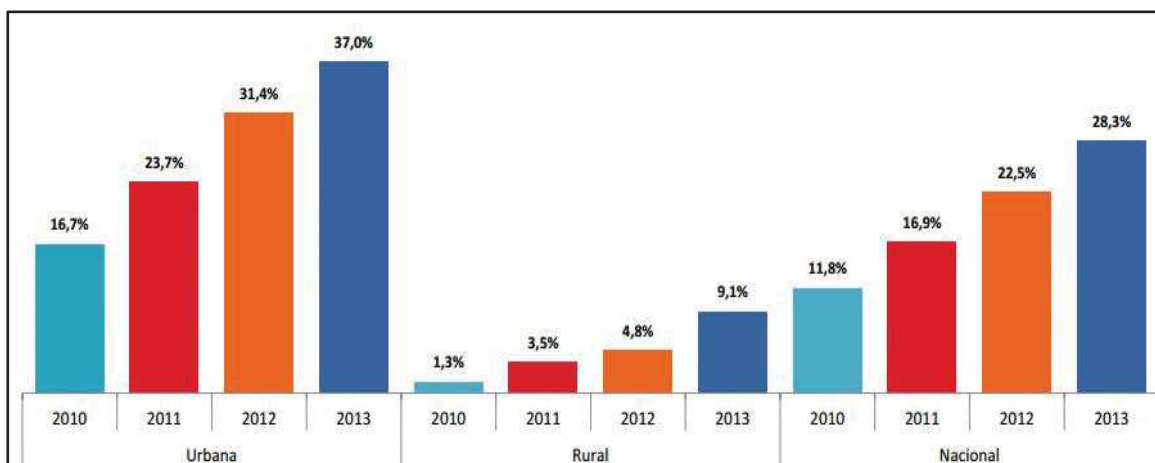


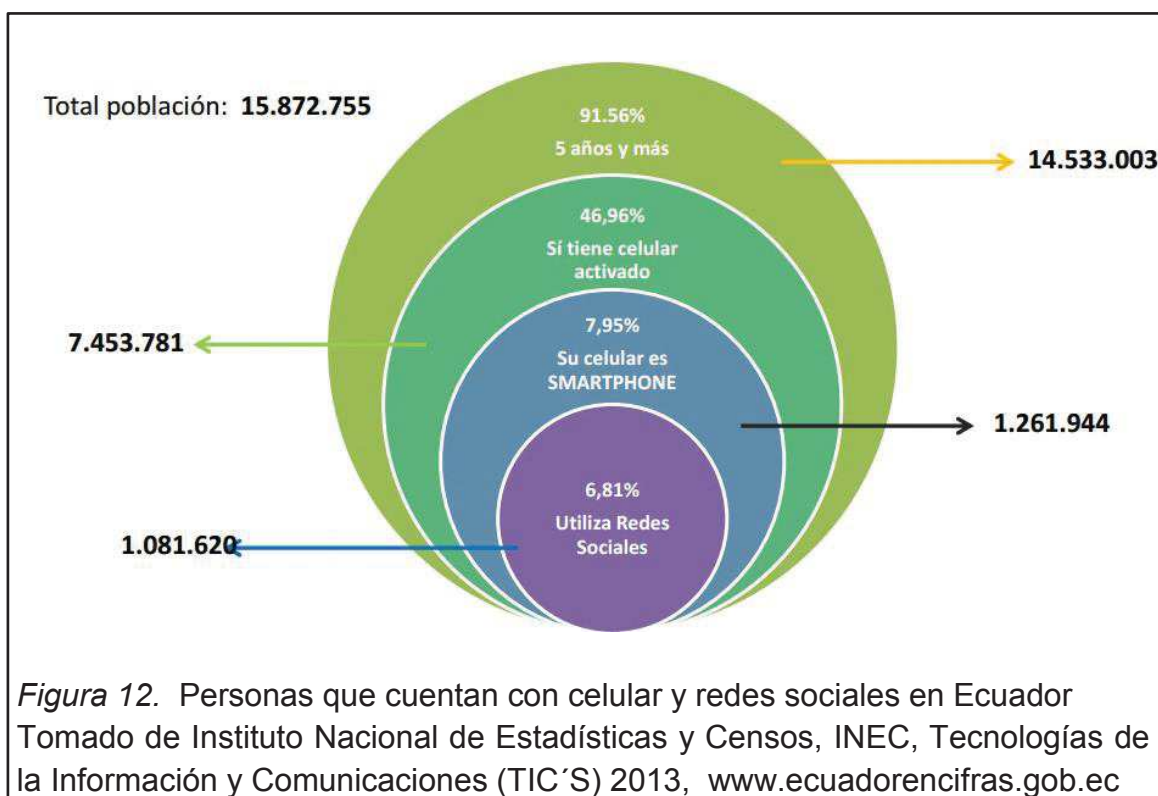
Figura 11. Porcentaje de acceso a Internet en Ecuador

Tomado de Instituto Nacional de Estadísticas y Censos, INEC, Tecnologías de la Información y Comunicaciones (TIC´S) 2013, www.ecuadorencifras.gob.ec

El Gobierno del Ecuador ha instituido acciones y políticas que le han acercado a la población a través del uso de las tecnologías, el Internet, y las redes

sociales; sin embargo, aún no se muestran indicios de que se está trabajando de forma exhaustiva para cambiar el modelo de toma de decisiones en base al análisis de datos y el conocimiento de las empresas públicas.

La falta de iniciativas gubernamentales, también se refleja en el limitado acceso a las redes sociales y manejo de tecnologías móviles a nivel nacional, como se muestra en la *Figura 11*, solamente un 6,81% de la población cuenta con acceso a ellas mientras un 91,56% todavía no cuenta con posibilidades de acceder o no está interesado.



2.6.2. Turismo

En algunos países los ingresos generados por el turismo son elevados y de gran relevancia en la economía de los mismos, es un motor para el ingreso de recursos económicos y la generación de empleo. En algunos países de Europa se ha creado programas para promover el turismo interno y externo a través de convenios y acceso a facilidades de pago (SEGITTUR, 2012). Para que un país se constituya en una potencia turística es necesario contar con el apoyo

gubernamental y de la sociedad, pero, es fundamental contar con las facilidades turísticas y servicios requeridos por los turistas.

Actualmente existen varios espacios en las redes sociales y páginas como *TripAdvisor*¹ que es una fuente de información muy importante para los turistas que buscan un destino de viaje; los comentarios, las fotos y las recomendaciones realizadas a través de estos medios son críticos al momento en que una persona o un grupo tomen una decisión para visitar un lugar u otro.

Los tipos de datos que se manejan en el sector turístico son texto, números y multimedia para transmitir de forma atractiva la información de los destinos turísticos; también se utiliza redes sociales, sitios web y el Internet.

Los establecimientos turísticos son una importante fuente de datos, pues ellos capturan los datos personales de los visitantes (nivel de educación, país de origen), nivel económico y lo más importante intereses sobre los destinos que desean visitar, los servicios que requieren, las facilidades que consideran fundamentales para garantizar su permanencia en un lugar.

Utilidad de Big Data para el sector Turístico

Big Data es de utilidad para el turismo ya que permite el análisis de las grandes cantidades de datos que se generan en este sector y a partir de su análisis establecer patrones que permitan identificar segmentos de mercado para diferentes áreas de esta industria, hoteles, agencias, transporte, y otro tipo de establecimientos.

Para el sector hotelero la identificación de los servicios que requieren los turistas que los visitan es fundamental para la orientación de sus estrategias de promoción (Alegria, 2014) y (Lancis, 2014); establecer su ubicación es un factor crítico para obtener mayor cantidad de visitantes, pues, facilitan el acceso a diversos destinos en los que está interesado el turista; conocer la cultura, creencias y preferencias de los visitantes eleva la calidad de los servicios e inclusive le permite desarrollar servicios a la medida.

¹ TripAdvisor. (2014). TripAdvisor. Consultado en julio 31, 2014, de TripAdvisor Sitio web: http://www.tripadvisor.es/PressCenter-c6-About_Us.html

El acceso a la información facilita los turistas la planificación de un viaje de forma personal, lo cual, es una desventaja que las agencias de viaje y operadores deben enfrentar; las campañas y ofertas de sus paquetes deben enfocarse en productos atractivos y de interés para el turista, conocer las preferencias, gustos, habilidades, temores y demás información útil permitirá el desarrollo de este tipo de productos e inclusive personalizarlos a las necesidades de una persona. Una agencia de viajes debe preocuparse de aspectos como la salud, la edad, el acceso a seguros, etc., para ofrecer paquetes que cubran emergencias como accidentes u otro tipo de calamidades que se presenten.

Para estructurar estrategias atractivas en el sector turístico es necesario conocer el mercado o el segmento al que se va a impactar y esta es una tarea a largo plazo que requiere de recursos humanos y económicos. Identificar los patrones de comportamiento de los turistas es vital para el desarrollo de los servicios turísticos, personalización y diseño de paquetes y productos, construcción de facilidades, generar oferta de empleo, garantizar el retorno y recomendación de los visitantes, exteriorización de la oferta, incrementar el turismo hacia un destino.

El turismo además es una fuente generadora de empleo; cada uno de los servicios ofertados requiere de personal que contribuya con la calidad del mismo, escoger los profesionales adecuados para prestar el servicio es otra tarea en la que *Big Data* genera oportunidades para determinar la idoneidad de una persona para desempeñar una actividad u otra.

La integración de la información es una gran oportunidad para el sector turístico pues le permite adquirir conocimiento sobre el origen de los visitantes, experiencia sobre la calidad de servicios, seguridad, comportamiento, promedio de gasto, desplazamientos de los turistas preferencias de alojamiento etc.

En el Ecuador el Ministerio de Turismo ha emprendido grandes campañas para identificar al país como una potencia turística, para lo cual, ha trabajado en el

desarrollo de normativa para regular el cuidado y acceso a ciertos destinos turísticos, con la finalidad de garantizar su permanencia.

Como se muestra en la *Figura 12*, en el año 2013 el Ecuador recibió a 1.364.057 visitantes internacionales de donde se pueden captar 1.251,3 millones de dólares sin considerar el turismo interno.



La oferta turística del Ecuador se beneficiaría de forma estratégica con el análisis de los datos, pues, en los estudios que realiza el Ministerio de Turismo se han identificado algunos factores y tendencias respecto al desarrollo del turismo en el País, Big Data permite ampliar este campo de análisis incluyendo datos no estructurados y la información de las redes sociales para el desarrollo de estrategias de inversión, colaboración y emprendimiento.

El Ministerio de Turismo cuenta con algunas iniciativas para empezar con el análisis de datos y la identificación de segmentos de mercado de acuerdo al perfil de los visitantes, la edad es una de las variables con las que se ha iniciado el proceso de segmentación y a partir de estos resultados se espera fortalecer e implementar los servicios atractivos para los turistas. La *Figura 13* muestra que la mayoría de los visitantes se encuentra entre los 18 y 40 años.

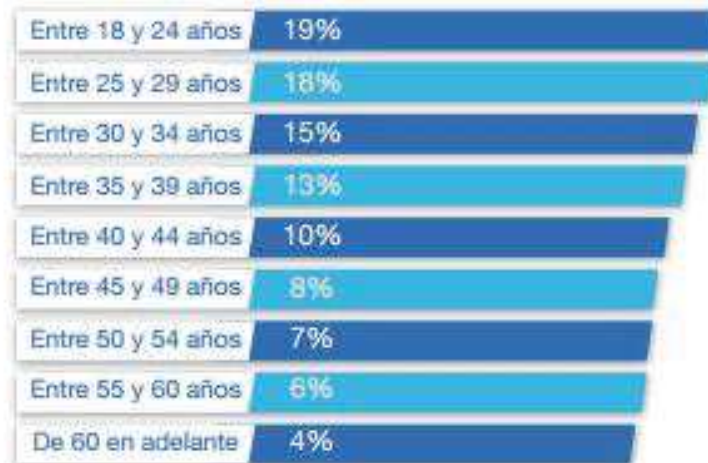


Figura 14. Edad de turistas que visitan el Ecuador
Tomado de Dirección de Investigación, Ministerio de Turismo del Ecuador, La experiencia turística en el Ecuador, 2012, p.34

Detectar las necesidades del turista, apoya una mejor toma de decisiones respecto al direccionamiento de los recursos y el desarrollo de obras que potencien el turismo y logren un impacto de calidad en los visitantes. Igualmente el reconocimiento de los niveles de educación de los turistas que arriban a Ecuador permitirá identificar las facilidades que requieren para elevar la sensación de confort y bienestar mientras dure su visita al país; la *Figura 14* muestra un resumen de los segmentos turísticos de acuerdo a su nivel de educación.

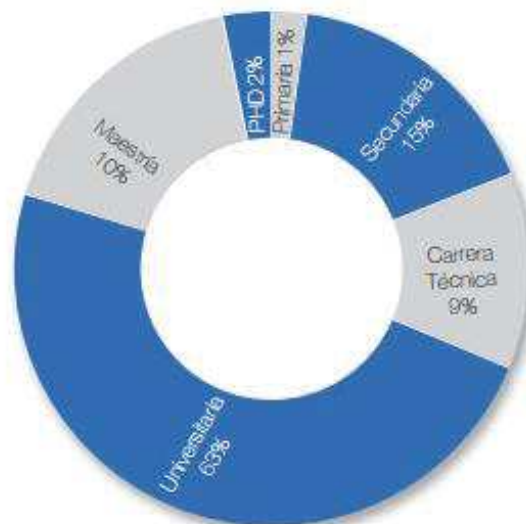


Figura 15. Nivel de educación de turistas que visitan el Ecuador.
Tomado de Dirección de Investigación, Ministerio de Turismo del Ecuador, La experiencia turística en el Ecuador, 2012, p.35

Interés sobre este proyecto

El desarrollo de este proyecto fue promovido inicialmente por el interés en la aplicación de Big Data para el análisis de datos en el Ministerio de Turismo como se detalla mediante una carta de auspicio en el ANEXO 2. Entre los objetivos específicos de este proyecto de investigación se encuentra el desarrollo de una guía básica para la aplicación de Big Data en el sector turístico y dar a conocer los beneficios que la aplicación de esta solución puede ofrecer al sector.

La guía entregada al Ministerio de Turismo, (ver ANEXO 3), comprende un análisis general del problema al que se enfrenta esta Institución, debido a la gran cantidad de datos que debe procesar y que por diversos motivos no pueden analizarse y generar valor al turismo del país.

De la información proporcionada por el Ministerio de Turismo y en base a los recursos bibliográficos consultados en esta investigación se ha podido identificar que existen algunos factores que dificultan el potenciamiento del turismo en el Ecuador y entre los más críticos se encuentran los siguientes:

- Falta de servicios automatizados, no se cuenta con suficientes sitios nacionales especializados en presentación del Ecuador como un destino turístico
- Datos descentralizados
- Calidad y cantidad de datos
- Infraestructura

Por otro lado se encontró varias ventajas de la aplicación de tecnologías como Big Data permitirán al Ministerio de Turismo mayor facilidad de acceso, fiabilidad de los datos, reducción de costos, segmentación del mercado, personalización de productos y servicios, establecimiento de métricas e identificación de tendencias.

Además se realizó un resumen de las técnicas de aplicación de Big Data, se detalló las etapas para el desarrollo adecuado de un proyecto de Big Data y se proporcionó las respectivas recomendaciones.

Finalmente como parte importante de este proyecto el Ministerio de Turismo proporcionó una carta de aceptación y recepción satisfactoria del producto entregado, ver ANEXO 4.

2.7. Generalidades de las Tecnologías

Actualmente existen varias herramientas para el procesamiento de datos; dependiendo del volumen de los datos y de la variabilidad de las fuentes de datos una industria debe seleccionar de forma adecuada las herramientas que va a implementar, considerando la velocidad de procesamiento y el tiempo de respuesta.

Una industria dependiendo de sus necesidades, recursos y habilidades puede seleccionar entre herramientas de distribución open source o herramientas pagadas como en el caso de las proporcionadas por IBM, Oracle, SAP, entre otras.

Big Data cuenta con bases de datos NoSQL que, ayudan a la industria a superar las limitaciones de una base de datos relacional cuando se supera la capacidad de manejo de los datos, al igual que, soportan el almacenamiento de datos no estructurados; en la Tabla 3, se describe las ventajas y desventajas de un modelo NoSQL frente a un modelo SQL.

Tabla 3. Comparación de ventajas y desventajas de un modelo SQL y NoSQL.

SQL		NoSQL	
Ventajas	Desventajas	Ventajas	Desventajas
Velocidad	Escalabilidad	Mayor volumen	Más lento
Facilidad de uso	Hardware especializado	Variabilidad	Complejo
Más variedad	Flexible en la configuración del hardware	Bajo costo de implementación	Requiere Experiencia

Cuando el volumen de los datos es reducido el modelo relacional es la mejor opción, considerando la capacidad de *Big Data*, los datos estructurados son fáciles de relacionar y el resultado de su análisis es predecible o fácil de detectar. La aplicación de un modelo *NoSQL* es aplicable cuando el volumen

de los datos supera la capacidad de los modelos *SQL* y tienen una variabilidad elevada, los costos de implementación son altos, pues, implican capacitación, adquisición de herramientas, personal y mejoras de la infraestructura.

A diferencia de las bases de datos relacionales las *NoSQL* trabajan con el modelo *BASE* que es una alternativa flexible a *ACID*, y se resume de la siguiente manera:

- “*Basic availability*.- El almacén de datos funciona la mayor parte del tiempo incluso ante fallos, esto gracias al almacenamiento distribuido y replicado” (López de Ipiña, 2013).
- “*Soft-sate*.- Los almacenes no tienen por qué ser consistentes ni sus réplicas en todo momento” (López de Ipiña, 2013).
- *Eventual consistency*.- La consistencia se da eventualmente

En el mercado existen una gran variedad de bases de datos para la implementación de *Big Data* se clasifican de acuerdo al tipo de datos que almacenan, las más importantes se detallan en la Tabla 4 que se muestra a continuación:

Tabla 4. 4 tipos de bases de datos *NoSQL*
Tomado de Phil Simon, *Too Big to Ignore*, 2013, p.123

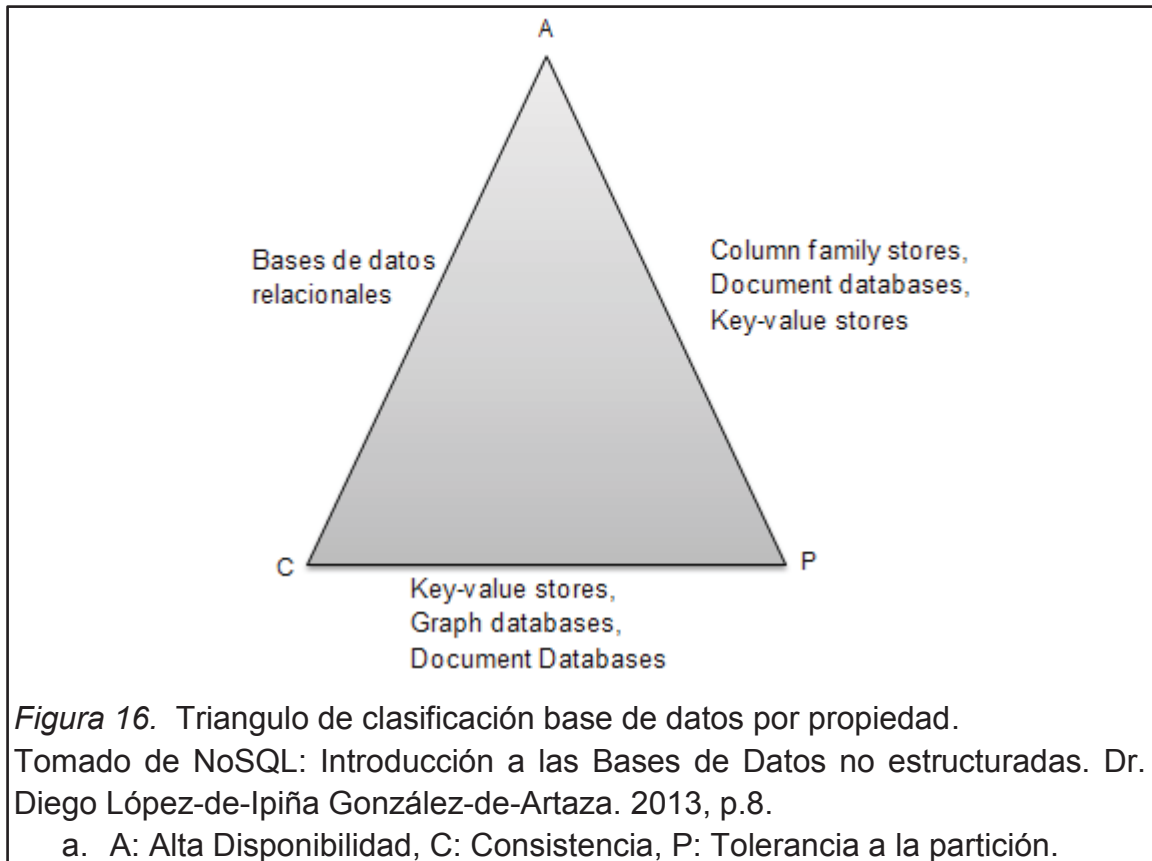
Tipo	Descripción	Ejemplos
<i>Key-value stores</i>	La estructura de este tipo de base de datos consiste en el uso de tablas hash, (tablas asociativas), donde se considera una única clave para cada ítem Son fáciles de implementar Ineficientes cuando se trata de actualizar parte de un dato.	Oracle BDB DynamoDB Redis
<i>Column family stores</i>	Son útiles para el almacenar y procesar grandes cantidades de datos distribuidos sobre varios equipos. Todavía manejan claves, pero para múltiples columnas.	Cassandra HBase

<p><i>Document Databases</i></p>	<p>Inspirada en lotus notes y silimar a <i>Key-value stores</i>. Versiona documentos que son colecciones de otras colecciones de acuerdo a un valor clave Los documentos semiestructurados son almacenados en formatos como JSON Soporta las consultas de base de datos de una forma más eficiente</p>	<p>MongoDb CouchDB</p>
<p><i>Graph databases</i></p>	<p>En lugar de tablas de filas y columnas este modelo de bases de datos puede escalar a través de múltiples equipos Trabaja con lenguaje <i>NoSQL</i> para evitar el tiempo de procesamiento</p>	<p>Infinite Graph Neo4J</p>

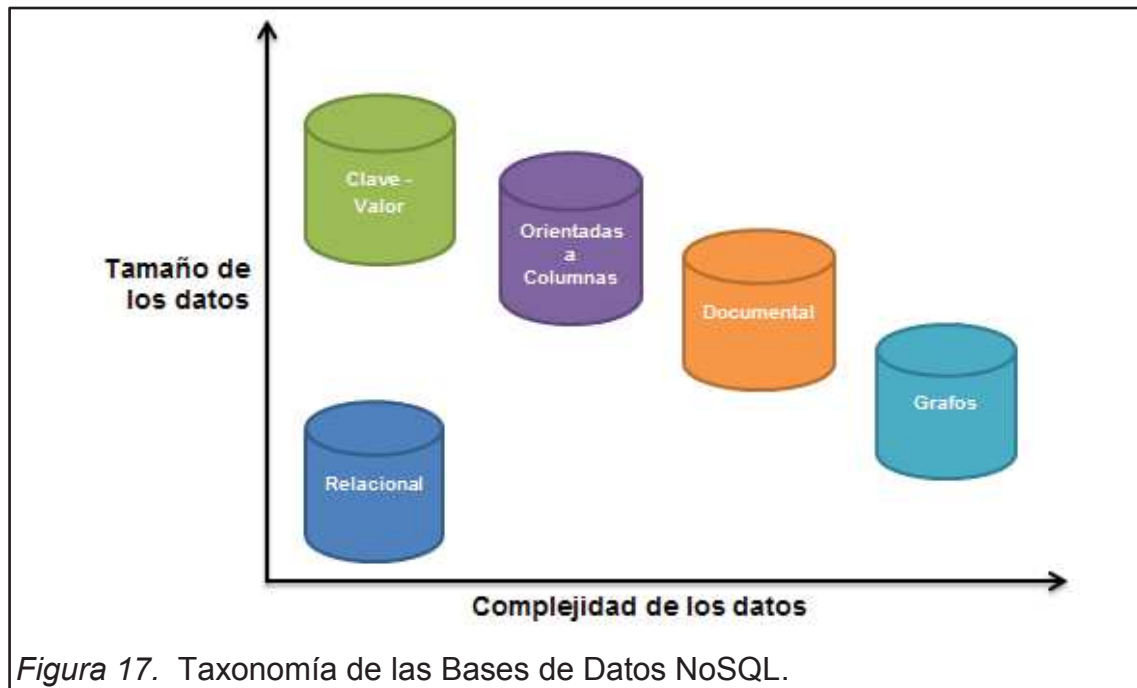
El Teorema de CAP es una afirmación que asegura que en un sistema de datos distribuido, dos, de las siguientes propiedades pueden cumplirse (López de Ipiña, 2013):

- Fuerte Consistencia.- La visualización de los datos es igual para todos los clientes
- Alta Disponibilidad.- Toda petición debe recibir al menos una respuesta
- Tolerancia a Particiones.- Procesamiento de mensajes incluso si un nodo se separa de la red.

Esto es equivalente al triangulo de gestión de proyectos donde se establece lo siguiente: “*You can have it good, you can have it fast, you can have it cheap: pick two.*” (Swisher, 2014); Como se muestra en la *Figura 15*, los diferentes tipos de base de datos se clasifican entre 2 de las propiedades mencionadas:



Existen varios criterios que una organización debe considerar al momento de elegir una base de datos *NoSQL*, para comenzar, será necesario conocer el tamaño de los datos que se desea procesar y por otro lado la complejidad de los datos seleccionados para el análisis (Soumendra, Madhu, & Harsh, 2013); con estas dos variables, como se ve en la *Figura 17*, se podrá determinar qué tipo de base de datos cubre los requerimientos del negocio.



Según el ranking de las bases de datos NoSQL más utilizadas se encuentra MongoDB puesto 4, Cassandra puesto 8, Redis puesto 10, Hbase puesto 14, CouchDB puesto 22, Neo4j puesto 23 (solid IT, 2015); las características generales de estas bases de datos se detallan en la Tabla 5:

Tabla 5. Descripción de las principales características de las bases de datos NoSQL

Características generales

Nombre	Licencia	Sistema Operativo	Modelo	SQL
MongoDB	Open Source	Win/Lin/OS	Documental Store	No soporta
Cassandra	Open Source	Win/Lin/OS	Wide column store	No soporta
Redis	Open Source	Win/Lin/OS	Key-Value store	No soporta
Hbase	Open Source	Win/Lin/Unix	Wide column store	No soporta
CouchDB	Open Source	Win/Lin/OS/Andr	Documental Store	No soporta
Neo4j	Open Source	Win/Lin/OS	Graph database	No soporta

Características Técnicas

Nombre	Tipo de Datos	Particiones	Réplicas	MapReduce	Concurrencia	Consistencia	Control de Acceso
MongoDB	Varios	Soporta	Soporta	SI	Soporta	Eventual/ Inmediata	Usuarios y roles

Cassandra	Varios	Soporta	Soporta	SI	Soporta	Eventual/ Inmediata	Usuarios por objeto
Redis	Datos especiales	No soporta	Soporta	NO	Soporta	No definido	Acceso basado en clave
Hbase	No definido	Soporta	Factores de selección	SI	Soporta	Inmediata	ACL
CouchDB	No definido	Soporta	Soporta	SI	Soporta	Eventual	Por base de datos
Neo4j	Varios	No soporta	Soporta	NO	Soporta	Eventual	No definido

La *Figura 17* muestra el ranking general de las bases de datos con mayor relevancia en el mercado:

256 systems in ranking, February 2015					
Rank	Last Month	DBMS	Database Model	Score	Changes
1.	1.	Oracle	Relational DBMS	1439.72	+0.56
2.	2.	MySQL	Relational DBMS	1272.45	-5.06
3.	3.	Microsoft SQL Server	Relational DBMS	1177.48	-21.13
4.	 5.	MongoDB	Document store	267.24	+16.35
5.	 4.	PostgreSQL	Relational DBMS	262.34	+7.85
6.	6.	DB2	Relational DBMS	202.42	+2.29
7.	7.	Microsoft Access	Relational DBMS	140.54	+1.41
8.	8.	Cassandra	Wide column store	107.08	+8.34
9.	9.	SQLite	Relational DBMS	99.56	+3.37
10.	10.	Redis	Key-value store	99.21	+4.97
11.	11.	Sybase ASE	Relational DBMS	86.34	+2.55
12.	12.	Solr	Search engine	81.48	+4.74
13.	13.	Teradata	Relational DBMS	69.45	+2.40
14.	14.	HBase	Wide column store	57.15	+3.56
15.	15.	FileMaker	Relational DBMS	53.44	+1.74
16.	16.	Elasticsearch	Search engine	52.84	+3.80

Figura 18. Ranking bases de datos
Tomado de <http://db-engines.com/en/ranking>, by solid IT

Hadoop

Según (Russom, 2013) Apache *Hadoop* es una potente y muy popular herramienta en el desarrollo de proyectos de *Big Data*, consiste en un ecosistema constituido por varios productos y tecnologías de fuente abierta supervisadas por *Apache Software Foundation* entre las que se destacan HDFS (*Hadoop Distributed File System*), *MapReduce*, *Hbase*, que constituyen

la base para aplicaciones Inteligencia de Negocios y *Datawarehouse*. Las principales características de esta solución son:

- El tipo de licencia es de código abierto, pero, dependiendo de la necesidad del negocio y los recursos se puede implementar a través de otros fabricantes que cuentan con soluciones personalizadas basadas en *Hadoop*.
- Esta solución es principalmente un sistema de archivo por lo cual carece de funciones que se asocian a los sistemas de gestión de base de datos.
- Entre las características de *Hadoop* se encuentra la gestión y procesamiento de altos volúmenes de datos no estructurados basados en archivos.
- Respecto a *MapReduce* se debe indicar que existe relación en el funcionamiento de estas tecnologías pero no son dependientes, es decir, *Hadoop* puede desplegarse sin *MapReduce*.
- La gestión y almacenamiento de los datos solo se encuentra limitada por la facilidad para depositar los datos en archivos y copiarlos en HDFS.

Considerando este antecedente es evidente que dentro del análisis de base de datos NoSQL es imposible considerar una comparación con *Hadoop*.

Como parte de este estudio es necesario realizar una recomendación respecto al uso de bases de datos NoSQL considerando las características más relevantes del modelo, la solución y la importancia en el mercado respecto a su usabilidad (Solid IT, 2015), por lo cual, solamente se considerará a MongoDB y Cassandra pues son las más conocidas y utilizadas en la actualidad.

MongoDB es una base de datos orientada a documentos, gestionan principalmente datos semi estructurados y almacenan los datos en formatos estándar como lo es JSON. Está centrada en la consistencia y la tolerancia a particiones por lo tanto es CP dentro del Teorema de CAP, se recomienda su uso en los siguientes casos:

- Almacenamiento y registro de eventos
- Sistemas de manejo de documentos
- Comercio electrónico

- Móviles
- Páginas web
- Almacenamiento de Comentarios
- Gestión de estadísticas

No se recomienda su uso si los requerimientos de análisis consideran lo siguiente:

- Tablas de base de datos
- Joins
- Transacciones
- Esquema de base de datos

Cassandra es una base de datos orientado al almacenamiento en columnas, altamente escalable y está centrada en la disponibilidad y tolerancia a fallos, con consistencia eventual (milisegundos) por tanto es AP dentro del Teorema de CAP, se recomienda su uso en los siguientes casos:

- Replicación de datos
- Análisis en tiempo real
- Marketing
- Redes sociales
- Marketing en redes sociales

No se recomienda su uso si los requerimientos de análisis consideran lo siguiente:

- Joins, para generar mayor velocidad
- Orden de datos en tiempos de consulta
- SQL

Es difícil realizar una recomendación específica respecto a la implementación y uso de una base de datos *NoSQL*, pues, se requiere conocer las necesidades específicas del negocio, al igual que, los recursos con los que cuenta, considerando en este punto la experiencia del equipo encargado de la implementación, tiempos de respuesta esperados y variabilidad de los datos como se muestra en las *Figuras 15 y 16*; No obstante, si es posible establecer

algunas consideraciones para reducir el nivel de complejidad durante el análisis:

- Establecimiento de un caso de aplicación y su alcance
- Conformación del equipo de trabajo
- Identificar la cantidad de datos y la variedad de las fuentes
- Establecer los tiempos de respuesta en las consultas de datos
- Valorar la necesidad de la organización respecto a la Alta Disponibilidad, la Consistencia y la Tolerancia a particiones
- Ubicar la taxonomía más próxima al objetivo del caso de aplicación

Entre las soluciones que puede aplicar una industria para el procesamiento de datos desde la perspectiva de *Big Data*, se encuentran las soluciones *Open Source*, aplicaciones pagadas y soluciones *Cloud*. Los criterios para seleccionar una herramienta dependerán de las necesidades del negocio.

En la Tabla 6 se muestra un cuadro comparativo entre estos tres tipos de soluciones:

Tabla 6. Comparación de las soluciones para el procesamiento de datos desde la perspectiva de *Big Data*

Criterio	Open Source	Soluciones Pagadas	Cloud
Costo	Precios accesibles, dependen de la necesidad del proyecto	El precio es elevado	El precio es flexible, depende de los requerimiento del negocio
Flexibilidad	Configuración abierta, y capacidad de integración	Soporta algunas configuraciones del usuario y admite la integración con algunas herramientas	Acceso mínimo a la configuración, la integración depende de la infraestructura contratada, limitación de permisos
Escalabilidad	Soporta la escalabilidad sin importar el hardware, requiere de	Soporta la escalabilidad pero está limitada por el hardware	La escalabilidad es una de las principales características de estas soluciones, completa

	conocimiento y experiencia en la solución		independencia del hardware
Experimentación	Admite la experimentación y la configuración de varios entornos de pruebas	Los entornos de experimentación y pruebas agregan un costo a la implementación	Los entornos de experimentación y pruebas agregan un costo a la implementación debido al consumo de recursos
Conocimiento	Requiere de conocimientos y habilidades para la configuración y mantenimiento	No requiere conocimientos específicos pues cuenta con el soporte del vendedor	No es necesario contar con expertos para el soporte y mantenimiento pero se recomienda mantener un especialista para el control de los niveles de servicio

Una industria debe considerar que la implementación de cualquier tipo de solución para apoyar el procesamiento de datos requiere de la inversión de recurso económico y humano, y no debe considerarse como el objetivo principal de un proyecto de *Big Data* pues este debe ser visto por la Gerencia como una Estrategia de Negocio.

3. Estudio Comparativo

3.1. Análisis de técnicas

La integración de los datos es una estrategia de negocio que se ha aplicado durante años en el análisis de los datos estructurados para generar oportunidades de mejora e innovación de los productos y servicios de una industria; sin embargo, existen limitaciones en el análisis de los datos no estructurados debido a la complejidad y variedad de los mismos.

El uso de técnicas de integración desde la perspectiva de *Big Data* se diferencia de la integración de datos tradicional debido a las siguientes consideraciones:

- *Big Data* contempla la integración de numerosas fuentes de datos
- El dinamismo de las fuentes de datos estudiadas generalmente es elevado
- Las fuentes de datos estudiadas son heterogéneas, difieren en la estructura de los datos inclusive entre entidades similares
- Las fuentes de datos presentan diferentes cualidades, se diferencian en el tipo de datos, la calidad, la dificultad para su recopilación y nivel de accesibilidad.
- La integración de los datos se logra a partir de las transacciones de grandes cantidades de datos y su interacción para el procesamiento de los datos.

El análisis de los datos no estructurados requiere de la aplicación de técnicas y el uso de tecnologías que permitan recopilar, procesar, gestionar, consumir, almacenar y gobernar estos datos de forma que aporten valor al negocio. La integración de los datos ayuda a una industria a construir conjuntos de datos que se ajusten a sus necesidades y que le permitan alcanzar sus objetivos estratégicos a partir de datos recopilados de diversas fuentes.

El uso de las técnicas de análisis e integración de los datos disminuye la dependencia directa de la industria hacia TI, pues, el éxito de las decisiones tomadas en ese contexto requiere del involucramiento de todas las áreas del negocio. Esta cercanía del negocio con el estudio de los datos permite establecer un modelo para la toma de decisiones más confidencial y disminuye el riesgo del fracaso de proyectos de integración a largo plazo.

La aplicación de *Big Data* para el análisis de datos se realiza a través de la aplicación de técnicas y métodos que apoyan al procesamiento adecuado de datos estructurados y no estructurados en base a algoritmos matemáticos parametrizados de acuerdo a los objetivos del análisis. La identificación de patrones de comportamiento, tendencias, segmentos de mercado, requerimiento de productos y servicios son algunos de los resultados que pueden beneficiar a la industria si cuenta con el conocimiento y habilidades para seleccionar una técnica de análisis de datos que se ajuste a sus necesidades.

Los niveles de madurez del proceso de análisis serán de utilidad para que una organización pueda seleccionar las técnicas que cumplan con los requerimientos del negocio; Los niveles de madurez contemplan el nivel Básico, Anticipado y Predictivo (Krishnan, 2013).

- Nivel Básico.- La organización se encuentra en una etapa de evaluación respecto a su funcionamiento a través del tiempo
- Nivel Anticipado.- La organización ha identificado razones y causas respecto a los éxitos y fracasos del negocio frente a la gestión de los datos
- Nivel Predictivo.- La organización ha establecido un modelo de análisis y simulación, además, se preocupa por estudiar las predicciones para conocer lo que sucederá en el futuro inmediato.

En la Tabla 7, se observa un resumen de los niveles de madurez del proceso de análisis y algunos ejemplos de técnicas de *Big Data* aplicables según el nivel en el que se encuentre la organización.

Tabla 7. Niveles de madurez del proceso de análisis

Madurez del proceso de análisis	Características	Técnicas
Básico	Provee una vista estática e histórica del funcionamiento del negocio Se generan gráficos en tableros básicos y reportes estáticos	<ul style="list-style-type: none"> • Query y drill down • Ad hoc reporting • Standard reporting
Anticipado	Crea transparencia respecto al pasado y sobre el potencial rendimiento del futuro Uso de sistemas y procesos para realizar una serie de análisis descriptivos	<ul style="list-style-type: none"> • Segmentación • Análisis Estadístico • Análisis Sensitivo
Predictivo	Presentan ofertas dinámicas e ideas progresistas con compensaciones cuantificadas Requiere de alta calidad en la integración de datos y habilidades matemáticas complejas	<ul style="list-style-type: none"> • Optimización • Simulación • Modelos predictivos

Big data cuenta con técnicas de análisis estadístico, semántico, predictivo, de visualización y de automatización que brindan diversas facilidades para el estudio de la información y generación de oportunidades y conocimiento para el negocio.

Entre las principales técnicas basadas en un modelo de análisis estadístico se encuentran, minería de datos, clasificación, reglas de asociación, análisis de cluster, redes neuronales, *a/b testing*, aprendizaje simultaneo(ensemble), análisis de redes, reconocimiento de patrones, análisis espacial, aprendizaje supervisado y aprendizaje no supervisado. El análisis de series de tiempo y los mapas de calor son técnicas de visualización de datos.

Las técnicas de automatización se basan principalmente en el aprendizaje de máquina; y las técnicas que trabajan en base a un modelo de análisis semántico son: procesamiento de lenguaje natural, análisis de textos, análisis de sentimientos, procesamiento de señales y fusión e integración de datos.

Las técnicas de análisis predictivo y las más utilizadas en la industria son: aprendizaje predictivo, optimización, simulación, modelos predictivos, crowdsourcing y algoritmos genéticos.

La industria debe seleccionar de forma adecuada una o varias técnicas de análisis para aplicarlas en el caso de estudio establecido; conocer las características de cada una de las técnicas aplicables desde la perspectiva de *Big Data* es importante para valorar si cumplen con los requerimientos del negocio y de esta manera faciliten el trabajo del equipo dedicado al análisis de los datos y estudio e interpretación de información generada a partir de los datos. Factores como el volumen de los datos, la variedad de las fuentes y la velocidad de respuesta son un buen punto de partida para la selección de una técnica

La Tabla 8, muestra en resumen las principales características de las técnicas de análisis estadístico y las recomendaciones para su aplicación dependiendo de las necesidades del negocio:

Tabla 8. Técnicas de análisis estadístico

Técnica	Características	Usabilidad
Minería de Datos	Diseño de modelos matemáticos para el descubrimiento de eventos inusuales, que no pueden apreciados fácilmente	<ul style="list-style-type: none"> • Segmentación de clientes • Detectar patrones de comportamiento y tendencias • Predicción de comportamientos futuros

Clasificación	<ul style="list-style-type: none"> • Identificación de categorías • Se requiere de habilidades para la identificación de las categorías en base a observaciones de pertenencia y relacionamiento • Se incluye en los conjuntos de datos a los datos históricos. 	<ul style="list-style-type: none"> • Asignación automática de objetos • Categorización • Desarrollo de perfiles
Reglas de Asociación	<ul style="list-style-type: none"> • Descubrimiento de correlaciones entre variables de las fuentes de datos 	<ul style="list-style-type: none"> • Estrategias para incrementar ventas de productos • Extraer información de diversas fuentes de datos para detectar patrones de comportamiento en diferentes áreas
Análisis de Cluster	<ul style="list-style-type: none"> • Agrupamiento de objetos en clases en base a sus similitudes • Se trabaja con varias dimensiones para el análisis • Trabaja con patrones para establecer la similitudes entre objetos 	<ul style="list-style-type: none"> • Definición de modelos de negocio • Descubrimiento de patrones
Redes neuronales	<ul style="list-style-type: none"> • Procesamiento de una gran cantidad de variables • Capacidad para detectar relaciones no lineales • Análisis de información en bruto, es decir sin procesos de simplificación y en tiempo real • Capacidad para auto-aprender y ajustarse de forma independiente para análisis futuros. • Requieren más trabajo de exploración que otras técnicas 	<ul style="list-style-type: none"> • Identificar patrones ocultos • Construcción de modelos analíticos • Procesamiento de nuevos conjuntos de datos en tiempo real

A/B Testing	Comparación entre un grupo de datos de control y un grupo de datos de prueba para identificar diferencias y semejanzas ante diferentes variaciones establecidas en el análisis	<ul style="list-style-type: none"> • Aceptación de cambios en productos y servicios • Introducción de nuevos productos • Diseño de campañas de marketing • Diseño de marcas y páginas web
Regresión	Manipulación de variables dependientes para visualizar la influencia sobre las variables dependientes	<ul style="list-style-type: none"> • Niveles de satisfacción y lealtad sobre los productos y servicios • Identificación de comportamientos en base a la variación de variables independientes • Establecimiento de precios en base a factores externos • Identificar relaciones en base a características similares o dependientes

Estas técnicas se basan en el análisis de variables; las variables corresponden a un conjunto de datos de prueba que han sido previamente ordenados y clasificados de acuerdo al caso de estudio en el que se van a emplear, estos datos servirán como punto de comparación para identificar similitudes y diferencias entre todos los conjuntos de datos seleccionados para el estudio.

La aplicación de estas técnicas no implica un gran esfuerzo para la industria que tiene conocimientos en Inteligencia de Negocios o en análisis de datos estructurados, los cambios para su uso en *Big Data* son principalmente en los algoritmos para la integración y procesamiento de grandes volúmenes de datos.

Requieren de la participación de expertos para la selección de los datos y la construcción de los conjuntos de datos para el análisis. Entre las habilidades del equipo de trabajo se debe considerar conocimiento en el manejo de algoritmos matemáticos y probabilidades

Se utilizan principalmente para identificar patrones de comportamiento, medir los niveles de impacto de un producto, servicio o campaña de marketing, descubrir segmentos de mercado, modelos de negocio, tomar decisiones estratégicas en base a dos o más opciones y clasificar al mercado y la competencia.

La Tabla 9, muestra las principales características de las técnicas de análisis de visualización de datos y las recomendaciones para su aplicación dependiendo de las necesidades del negocio:

Tabla 9. Técnicas de análisis de visualización de datos

Técnica	Características	Usabilidad
Mapas de calor	Representación visual de los datos en 2 dimensiones, en el que los valores están representados por colores	<ul style="list-style-type: none"> • Proporcionar un resumen visual de la información • Comprensión de conjuntos de datos complejos

Las técnicas de visualización de datos se basan en el uso de software especializado para ayudar a la organización a ordenar los datos y representarlos de forma visual de modo que se facilite su comprensión.

Permiten el manejo de varias fuentes e integración de las mismas para construir los reportes.

Se utilizan generalmente para la toma de decisiones, identificación de tendencias, y métricas de negocio además de la estacionalidad en algunos segmentos de mercado.

La Tabla 10, muestra las principales características de las técnicas de análisis de automatización y las recomendaciones para su aplicación dependiendo de las necesidades del negocio:

Tabla 10. Técnicas de análisis de automatización

Técnica	Características	Usabilidad
Aprendizaje de Máquina	<ul style="list-style-type: none"> • Uso de software para facilitar el aprendizaje de la maquina desde los datos. • Análisis autónomo debido al entrenamiento continuo proporcionado por los datos 	<ul style="list-style-type: none"> • Identificar diferencias entre objetos • Identificar preferencias y recomendaciones de selección de productos • Predicción de probabilidades

Las técnicas de automatización se basan en el uso de software que permite el análisis de grandes cantidades de datos de diversas fuentes en diferentes niveles de profundidad.

Estas técnicas son de gran utilidad para el análisis de datos no estructurados, en especial ayudan a la comprensión de los datos provenientes de sensores.

Se utilizan para identificar segmentos de clientes, recomendar productos, detección de fraude y elevar la calidad del análisis predictivo. Estas técnicas enfatizan su uso en la generación de resultados en tiempo real.

La Tabla 11, muestra las principales características de las técnicas de análisis de semántico y las recomendaciones para su aplicación dependiendo de las necesidades del negocio:

Tabla 11. Técnicas de análisis semántico

Técnica	Características	Usabilidad
Procesamiento de lenguaje natural	<ul style="list-style-type: none"> • Requiere de alta precisión para el análisis • Análisis de grandes cantidades de textos • Análisis sintáctico y morfológico • Uso de patrones lingüísticos 	<ul style="list-style-type: none"> • Extracción de entidades y clasificación de los datos • Extraer hechos o conocimiento a partir de texto libre

Análisis de textos	<ul style="list-style-type: none"> • Identificación de cualidades de los textos • Brinda acceso a la información plasmada en diferentes tipos de textos • Análisis de palabras y documentos 	<ul style="list-style-type: none"> • Identificación de emociones • Mapeo de conductas, perfiles, etc. • Extracción de índices numéricos • Automatización en el procesamiento de correos
Análisis de sentimientos	<ul style="list-style-type: none"> • Uso de técnicas analíticas para extraer información subjetiva de los datos • Identificación de sentimientos respecto a un tópico 	<ul style="list-style-type: none"> • Clasificación de opiniones • Mejoramiento de servicios • Personalización de productos y servicios • Identificación de comportamientos
Fusión e integración de datos	Fusión e integración de varias fuentes de datos para ampliar el campo de análisis	<ul style="list-style-type: none"> • Determinar niveles de confianza • Identificar patrones de comportamiento

Estas técnicas se basan en el análisis de datos no estructurados y semiestructurados, considerando entre las fuentes más utilizadas en este tipo de análisis a los textos largos y complejos.

Las técnicas de análisis semántico facilitan la transformación de los datos no estructurados y semiestructurados en datos a los que se puede tratar con técnicas de análisis tradicional.

Las técnicas de análisis semántico aplican técnicas lingüísticas y estadísticas para identificar patrones comportamiento, categorizar y clasificar documentos, audio, video, e imágenes, personalización de productos y servicios, análisis de sentimientos y relaciones.

La Tabla 12, muestra las principales características de las técnicas de análisis predictivo y las recomendaciones para su aplicación dependiendo de las necesidades del negocio:

Tabla 12. Técnicas de análisis predictivo

Técnica	Características	Usabilidad
Análisis predictivo	Análisis de datos actuales e históricos para hacer predicciones de eventos futuros	<ul style="list-style-type: none"> • Identificación de patrones de comportamiento • Identificación de potenciales riesgos y consecuencias considerando su aplicación en la toma de decisiones • Identificación de relaciones entre clientes y productos • Identificación de tendencias
Optimización	<ul style="list-style-type: none"> • Se basa en modelos de decisión para describir la relación entre todos los elementos de una decisión. • Considera variables, valores, decisiones previas para optimizar la toma de nuevas decisiones 	<ul style="list-style-type: none"> • Mejoramiento de procesos • Análisis de inversión • Mejoramiento de la toma de decisiones
Simulación	Permite modelar el funcionamiento de sistemas complejos	<ul style="list-style-type: none"> • Predicciones sobre resultados • Diseño de escenarios • Predicciones de eventos futuros

Las técnicas de análisis predictivo emplean técnicas de minería de datos para reconocer patrones y predicciones sobre eventos futuros.

Estas técnicas basan su análisis en datos históricos y comparaciones con datos actuales para identificar diferencias y semejanzas en un período de tiempo.

Se utilizan para reconocer predicciones sobre diferentes eventos, identificar patrones de comportamiento, mejoramiento de procesos, diseño de escenarios y mejorar la toma de decisiones.

A continuación en la Tabla 13, muestra un resumen de las técnicas de análisis aplicables en *Big Data* y una comparación entre las características más significativas y requeridas por la industria:

Tabla 13. Comparación de técnicas de análisis de datos aplicables en Big Data

Características	Técnicas				
	Análisis Estadístico	Visualización de Datos	Automatización	Análisis Semántico	Análisis Predictivo
Volumen de Datos	Medio	Medio	Alto	Alto	Alto
Velocidad	Tiempo casi real	Tiempo real y casi real	Tiempo real y casi real	Tiempo real	Tiempo real y casi real
Variedad de fuentes	Requiere integración de los datos	Requiere integración de los datos	Soporta variedad de fuentes	Soporta variedad de fuentes	Requiere integración de los datos
Esfuerzo de aplicación	Bajo	Bajo	Medio	Alto	Alto
Implementación de software	Bajo impacto	Bajo impacto	Alto impacto	Alto impacto	Alto impacto
Equipo de trabajo	Analistas y Especialistas	Analistas	Especialistas y Expertos	Especialistas y Expertos	Analistas, Especialistas y Expertos
Identificación de patrones	Bajo	Medio	Alto	Alto	Medio
Predicciones	Medio	Bajo	Medio	Medio	Alto
Segmentación	Alto	Medio	Medio	Alto	Medio
Clasificación	Alto	Alto	Alto	Alto	Alto
Tendencias	Medio	Alto	Medio	Alto	Alto
Integración	Medio	Alto	Alto	Alto	Alto

De acuerdo a datos consultados en diversas fuentes las técnicas más utilizadas son las que se basan en modelos estadísticos, sin embargo, la generación de

datos a partir de las redes sociales ha incrementado el uso de las técnicas de análisis semántico y de automatización.

Las técnicas de visualización son útiles para industrias que desean trabajar de forma práctica con los datos y cuentan con los recursos para la adquisición de herramientas y el personal dedicado a la selección, clasificación e integración de los datos.

Las técnicas de análisis predictivo ayudan especialmente en la toma de decisiones estratégicas de una industria; las predicciones ayudan a la gerencia a identificar posibles riesgos y el impacto de una decisión u otra. Estas técnicas ayudan a la generación de conocimiento y al desarrollo de habilidades para identificar oportunidades de negocio.

La aplicación de las técnicas de análisis puede realizarse de forma individual o se puede seleccionar un grupo de técnicas que se ajusten a las necesidades de análisis del negocio.

Considerando las características mencionadas se realizará una valoración de las mismas en base al estudio de diversos análisis publicados por consultoras como Gartner Research Inc y McKinsey Global Institute para determinar cuáles de las técnicas analizadas permitirá generar mayor valor para el negocio y entrega más prestaciones al área de TI (Soumendra, Madhu, & Harsh, 2013); (Manyika, y otros, 2011); (Zikopoulos, Eaton, deRoos, Deutsch, & Lapis, 2012); (Krishnan, 2013).

La Tabla 14, presenta el cuadro con la valoración asignada a las características más representativas de las técnicas de *Big Data* identificadas durante la investigación:

Tabla 14. Valoración de características de las técnicas de Big Data

Características	Valoración (puntos)	Descripción de la valoración
Volumen de Datos	9	<ul style="list-style-type: none"> Fundamental para el análisis de Big Data Soporte de altos volúmenes de datos. Se considera un rango aceptable entre 9 y 7 puntos
Velocidad	7	<ul style="list-style-type: none"> La velocidad en el procesamiento de los datos es crítica dependiendo del tipo de resultado esperado. No siempre el negocio requiere niveles de velocidad altos. Se considera un rango aceptable entre 7 y 6 puntos
Variedad de fuentes	8	<ul style="list-style-type: none"> El análisis de datos considera el soporte de fuentes de datos variadas. Esta característica es crítica dependiendo de la vertical de negocio, aunque se puede restringir el tipo de fuente Se considera un rango aceptable entre 8 y 6 puntos
Esfuerzo de aplicación	5	<ul style="list-style-type: none"> La implementación de una técnica es una de las preocupaciones más importantes del negocio. El tiempo de aprendizaje y socialización es de gran impacto para un proyecto. Se considera un rango aceptable entre 5 y 4 puntos
Identificación de patrones	6	<ul style="list-style-type: none"> Entre los principales objetivos de Big Data está la identificación de Patrones por lo cual es importante que la técnica seleccionada cuente con esta facilidad. Se considera un rango aceptable entre 6 y 4 puntos
Predicciones	5	<ul style="list-style-type: none"> Objetivo de negocio Soporte análisis orientado a las predicciones Se considera un rango aceptable entre 5 y 3 puntos
Segmentación	5	<ul style="list-style-type: none"> Objetivo de negocio Soporte análisis orientado a la segmentación Se considera un rango aceptable entre 5 y 3 puntos
Clasificación	5	<ul style="list-style-type: none"> Objetivo de negocio Soporte análisis orientado a la clasificación Se considera un rango aceptable entre 5 y 3 puntos
Tendencias	4	<ul style="list-style-type: none"> Objetivo de negocio no primario y se apoya en la identificación de patrones y segmentación Soporte análisis orientado a la identificación de tendencias Se considera un rango aceptable entre 4 y 2 puntos

Integración	10	<ul style="list-style-type: none"> • La integración de las fuentes, modelos de datos, procesamiento y gestión son vitales para el éxito de un proyecto de Big Data • Se considera un rango aceptable entre 10 y 8 puntos
-------------	----	--

El puntaje máximo que una técnica puede obtener es de 64 puntos. La valoración de las características se ha asignado considerando nivel de importancia para la industria, requerimientos de negocio, relevancia de las industrias que han aplicado *Big Data* y estudio de investigaciones desarrolladas por consultoras reconocidas a nivel mundial. Características de peso similar cuentan con un puntaje idéntico para disminuir el margen de error en la calificación.

A través de un cuadro comparativo entre las técnicas de *Big Data* se pretende seleccionar una de ellas considerando la valoración establecida en la Tabla 14, es necesario aclarar que una organización no está obligada a aceptar esta valoración como una regla de aplicación ya que este proyecto de investigación se considera como una guía para la implementación de las técnicas de *Big Data* y a partir de este proceso gestionar de forma adecuada el conocimiento generado.

En la Tabla 15, se muestra la calificación de las técnicas de *Big Data* respecto a las características más relevantes identificadas en este estudio

Tabla 15. Calificación de las técnicas de Big Data

Características	Técnicas				
	Análisis Estadístico	Visualización de Datos	Automatización	Análisis Semántico	Análisis Predictivo
Volumen de Datos	7	8	8	9	9
Velocidad	5	5	6	7	6
Variedad de fuentes	6	5	8	8	7
Identificación de patrones	3	4	4	6	6
Predicciones	3	2	4	4	5
Segmentación	5	4	3	5	3

Clasificación	5	4	3	5	5
Tendencias	2	3	3	4	4
Integración	7	8	9	10	9
Total	43 puntos	43 puntos	48 puntos	58 puntos	54 puntos

De la calificación realizada se ha obtenido el siguiente orden de importancia frente a las características consideradas en el estudio:

- 1) Análisis Semántico con 58 puntos de un total de 64 puntos
- 2) Análisis Predictivo con 54 puntos de un total de 64 puntos
- 3) Automatización con 48 puntos de un total de 64 puntos
- 4) Análisis Estadístico con 43 puntos de un total de 64 puntos
- 5) Visualización de Datos con 43 puntos de un total de 64 puntos

El ranking obtenido permite reconocer que las técnicas de análisis semántico y predictivo son las que mayores prestaciones pueden ofrecer a una organización, por otro lado, si bien las técnicas de automatización no son las más utilizadas son bastante robustas en lo que respecta al análisis basado en el aprendizaje de máquina.

Como lo señalan (Krishnan, 2013) y (Manyika, y otros, 2011), las técnicas de análisis estadístico y de visualización de datos son las más fáciles de implementar y por tanto la mayoría de negocios las han aplicado en algún punto de la evolución para la aplicación de *Big Data*.

A continuación se muestra un análisis más detallado de los resultados obtenidos en la calificación de las técnicas de análisis de *Big Data*, considerando los 3 primeros puestos:

1) Análisis Semántico

Dentro de las técnicas de análisis semántico se encuentran técnicas como el procesamiento de lenguaje natural, análisis de textos, análisis de sentimiento, fusión e integración de datos; estas técnicas permiten procesar un alto volumen de datos, generando resultados en tiempo real, soporta una gran variedad de

fuentes, característica que hace atractivas a este tipo de técnicas para actividades como el marketing. Presenta facilidades para la identificación de patrones y tendencias, funcionan de mejor manera en estudios predictivos; se aplican en proyectos de segmentación y clasificación basados principalmente en el análisis de sentimientos y emociones, soporta la integración de datos sin necesidad de procesamiento previo ni estructuración de datos no estructurados.

Por otro lado es necesario reconocer que este tipo de técnicas requiere de un gran esfuerzo de la organización para su implementación al igual que la del software que soporte el análisis, por tanto es necesario contar con un equipo dedicado específicamente al análisis de datos y expertos en la obtención, valoración, gestión y procesamiento de los datos.

Dentro del cuadro de niveles de madurez del proceso de análisis este tipo de técnicas es aplicable a organizaciones que se encuentren en un nivel de madurez anticipado con un serio proceso de evolución a un nivel de madurez predictivo.

2) Análisis Predictivo

Entre las técnicas de análisis predictivo se encuentran las técnicas de optimización y simulación, estas técnicas se especializan en la identificación de patrones y tendencias a través de la predicción de escenarios en base a las variables incluidas en el estudio. Se caracterizan por el manejo de gran cantidad de datos recolectados de fuentes variadas que deben pasar por un proceso de depuración e integración para su análisis y generan resultados en tiempo real y casi real dependiendo de las necesidades del negocio. Los proyectos de clasificación y segmentación pueden basarse en predicciones considerando el tipo de variables establecidas en el análisis.

Cuentan con una alta capacidad para la integración de diversas fuentes de datos, pero aquí se consideran actividades de depuración y en ciertos casos de estructuración.

La implementación de técnicas basadas en análisis predictivo requiere de un caso de estudio bien elaborado con conocimiento claro de los productos esperados, el equipo de trabajo debe encontrarse bien estructurado y capacitado, de modo que, tengan las habilidades necesarias para la gestión del software de apoyo y el procesamiento de los datos.

Considerando los niveles de madurez del proceso de análisis este tipo de técnicas deben aplicarse en organizaciones ubicadas en el nivel predictivo.

3) Automatización

Las técnicas de automatización se especializan en el aprendizaje de máquina considerando el constante entrenamiento que pueden proporcionar los datos. Se utilizan para el análisis predictivo, y de segmentación considerando la identificación de patrones que faciliten el establecimiento de las rutas en los casos indicados. Soportan grandes cantidades de datos, provenientes de diversas fuentes que pueden ser procesadas en tiempo real y casi real dependiendo de las necesidades del negocio. Permiten el desarrollo de análisis basados en clasificación y cuentan con alta capacidad de integración.

Estas técnicas requieren de un equipo capacitado y con habilidades en la clasificación y selección de los datos, el software de apoyo para el análisis requiere de un manejo y gestión adecuada por lo que se recomienda contar con al menos un experto en el tema.

De acuerdo a los niveles de madurez del proceso de análisis este tipo de técnicas pueden aplicarse en organizaciones en el nivel predictivo.

Para el caso de este estudio se puede recomendar el uso de técnicas de análisis semántico para la implementación de *Big Data*; sin embargo, es necesario aclarar que las necesidades de una organización pueden solventarse con la aplicación de técnicas más o menos complejas dependiendo del nivel de madurez del proceso de análisis y otros factores críticos para el negocio.

El aporte de la integración y el análisis de los datos desde la perspectiva de *Big Data* a la generación de conocimiento es muy amplia y requiere de una gestión

adecuada para evitar el caos al interior de la industria; mantener una base de conocimiento, documentación actualizada y una cultura colaborativa elevará las posibilidades de generar valor a partir del conocimiento obtenido del análisis de los datos.

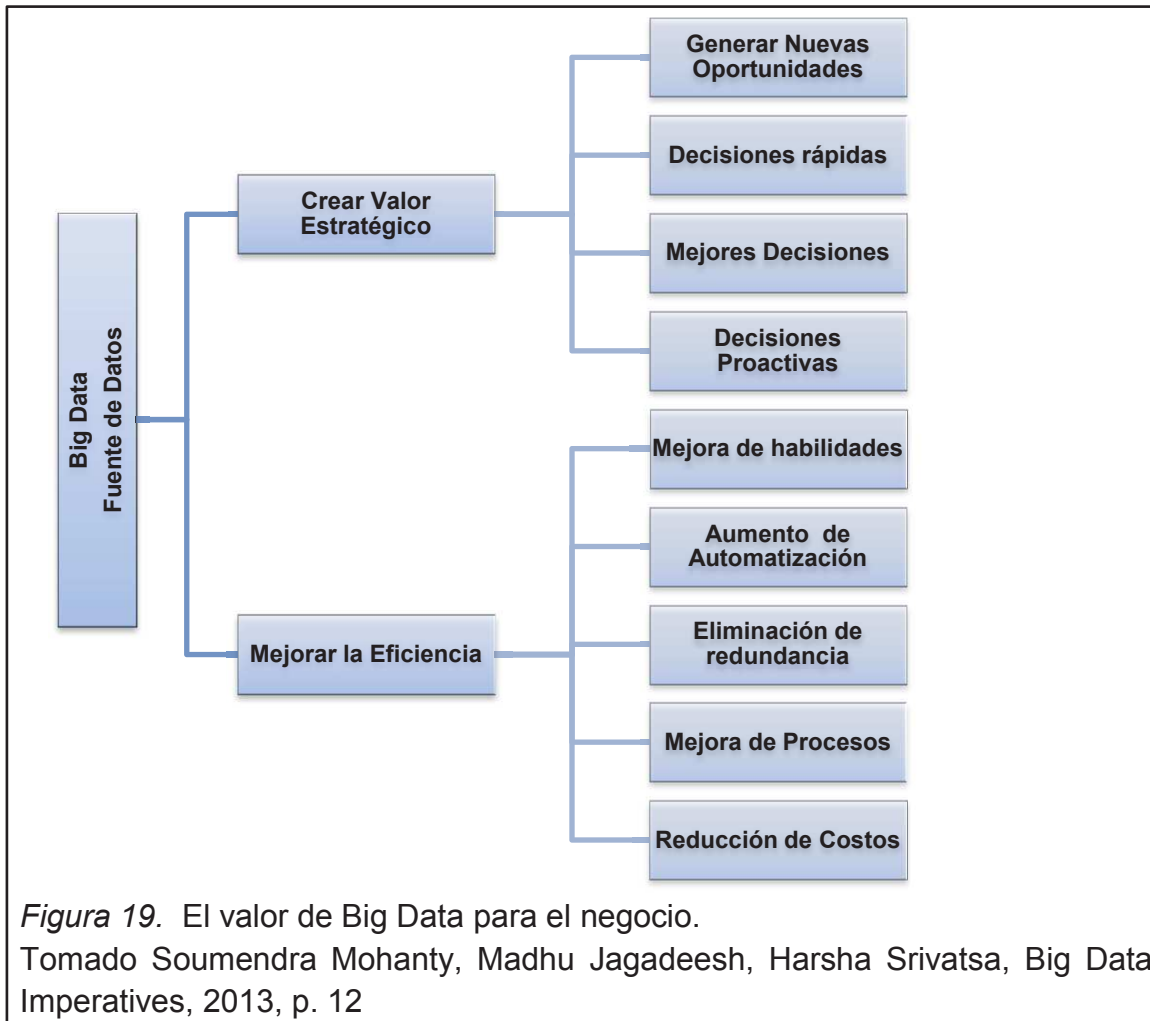
La integración apropiada de los datos requiere agilidad para la recolección y clasificación de los datos, además de habilidades para construir una arquitectura de integración escalable, soporte de especialistas, ayuda para crear conjunto de datos confiables y flexibilidad para la entrega de datos.

Finalmente la integración de los datos brinda algunos beneficios a la industria que si se gestionan de forma adecuada permitirán aportar valor al negocio a partir de los datos:

- Gestión de grandes cantidades de datos provenientes de fuentes variadas
- Información confiable
- Generación de valor para el negocio a través de la implementación de técnicas para el análisis de datos, desarrollo de aplicaciones eficientes, visualización de los datos y generación de conocimiento
- Incremento de las oportunidades de negocio
- Reducción de costos en el mantenimiento de los datos
- Retención de clientes y captación de nuevos clientes
- Identificación de riesgos y fraudes
- Gestión eficiente de la inversión
- Desarrollo de nuevos productos y servicios

3.2. Resumen y Análisis de resultados

Como resultado de la investigación se ha logrado identificar factores críticos que una industria debe considerar antes y después de emprender en un proyecto de análisis de datos desde la perspectiva de Big Data. La *Figura 18*, muestra un resumen del valor generado a partir de Big Data:



El análisis de datos es una gran oportunidad para generar valor al negocio, puede enfocarse en el mejoramiento e innovación en los procesos, productos y servicios de una industria, además de apoyar de forma estratégica la toma de decisiones al conocer el entorno sobre el que se va a trabajar.

La generación de conocimiento es uno de los principales objetivos del análisis de datos, la identificación de patrones, segmentos y tendencias permiten a una industria obtener las habilidades y experiencia para capturar oportunidades y enfrentar los requerimientos del mercado y a la competencia. Igualmente la implementación de herramientas proporciona velocidad y seguridad en la gestión y procesamiento de los datos al mismo tiempo que genera conocimiento para el equipo que administra la infraestructura.

Big Data es una solución orientada al negocio, sin embargo, no es la solución a todos los problemas de una industria; existen algunas consideraciones que se deben analizar antes de iniciar un proyecto de este tipo y entre las principales se encuentran las siguientes (Soumendra, Madhu, & Harsh, 2013):

- Conocer las necesidades y prioridades de la industria es fundamental antes de iniciar un proyecto de *Big Data*; si no existe interés al interior de la organización en la implementación de soluciones de Big Data será difícil conseguir recursos para el proyecto.
- Establecer un caso de aplicación; el esfuerzo de la organización debe tener objetivos claros y alcanzables, una industria no conseguirá ningún valor para el negocio si se dedica al análisis de datos sin una meta relacionada.
- Alinear los objetivos del proyecto a los objetivos estratégicos de la industria, el alcance el proyecto debe proponer resultados que generen valor al negocio, promoviendo el cumplimiento de los objetivos de la industria.
- La industria debe contar con un equipo de trabajo conformado por especialistas y expertos en la selección y análisis de los datos, entre sus habilidades se debe considerar el manejo de herramientas de procesamiento de datos, conocimiento sobre integración de datos estructurados y no estructurados, aplicación de técnicas para el análisis e integración de datos; este equipo debe trabajar directamente con el equipo del negocio en el estudio de la información y debe ser independiente del área de TI.
- Implementar una cultura de colaboración es muy importante para lograr que una industria consolide y expanda el conocimiento generado a partir del análisis de datos; conseguir un equipo de trabajo basado en la colaboración ayuda a incrementar sus habilidades y reproducir el conocimiento hacia el resto de la organización.
- La comunicación es una de las claves para lograr el éxito de un proyecto de *Big Data*, se debe transmitir los avances y los productos generados, al igual que las limitaciones identificadas y las dificultades que se

presenten para mantener el interés de la gerencia y mantener el control sobre la gestión de cambios.

- Transmitir los resultados a toda la organización ayudará a establecer un marco de colaboración para la identificación de las oportunidades de mercado y mejorar la toma de decisiones. Otro beneficio es la identificación de nuevo proyectos de implementación.

Un resumen de los requerimientos para el desarrollo del proyecto se muestra en la figura siguiente:

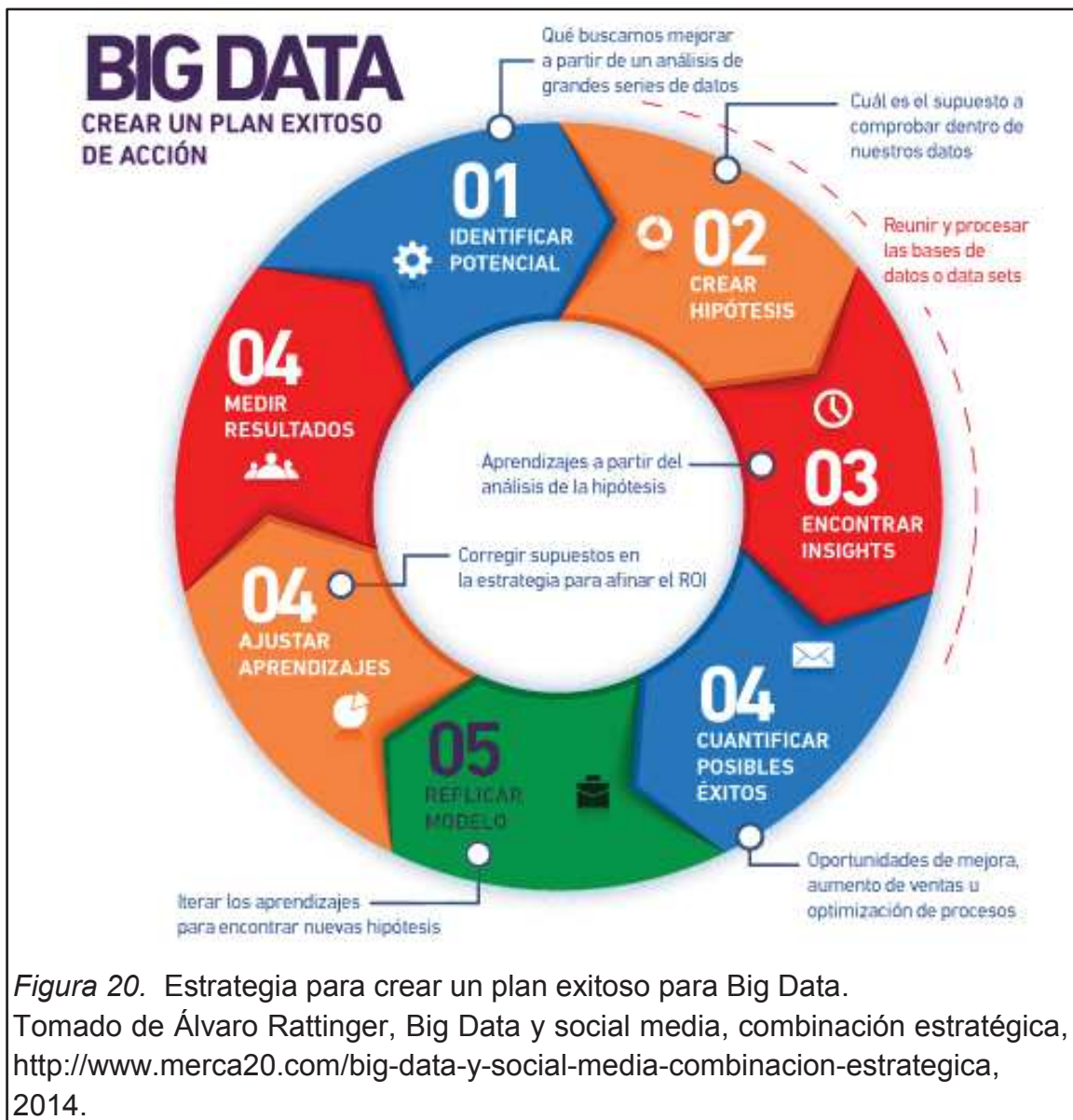


Figura 20. Estrategia para crear un plan exitoso para Big Data.

Tomado de Álvaro Rattinger, Big Data y social media, combinación estratégica, <http://www.merca20.com/big-data-y-social-media-combinacion-estrategica>, 2014.

Existen varios escenarios de aplicación de Big Data, en el estudio realizado se mencionaron al Gobierno y al Turismo como sectores donde Big Data puede generar valor a partir del análisis de los datos. El direccionamiento de recursos hacia la construcción de obras, la identificación de servicios, la automatización y el mejoramiento de los procesos son algunos de los beneficios que el análisis de datos puede proporcionar a estas industrias en cada uno de sus entornos.

En este estudio se ha realizado una breve revisión de las herramientas existentes en el mercado para la aplicación de Big Data y se ha descrito de forma general su utilidad y las principales características de su implementación

La investigación además permitió identificar las técnicas de análisis e integración de los datos estructurados y no estructurados desde la perspectiva de Big Data y las prácticas comúnmente aplicadas por la industria; además se ha logrado en base a un estudio comparativo reconocer la importancia de su aplicación para la generación de conocimiento dependiendo del tipo de análisis que se realice.

La selección de una técnica de análisis debe realizarse considerando las características y beneficios que esta puede brindar al negocio. Características como el análisis de grandes cantidades de datos, soporte de variedad de fuentes y tiempo de respuesta son la base para identificar si una técnica cumple con los requerimientos de la industria.

El tipo de análisis que se va a realizar es otro factor a considerar, pues, a partir de los requerimientos establecidos por el negocio se puede seleccionar una técnica o un conjunto de técnicas que permitan realizar el análisis de los datos y proporcionar los resultados esperados.

Durante la investigación realizada se pudo identificar las técnicas que presentan mayor cantidad de cualidades para la aplicación de Big Data, considerando características como, volumen, diversidad de fuentes, velocidad, integración de datos, e implementación de proyectos orientados a la

identificación de patrones, la segmentación, la clasificación y la predicción de escenarios son las técnicas de análisis semántico.

En el mercado existe una diversidad de técnicas y tecnologías aplicables para el análisis de datos desde la perspectiva de Big Data; existen soluciones privadas y cloud que prometen entregar el máximo valor de los datos al negocio y la generación de conocimiento así como el desarrollo de habilidades en la industria para la toma de decisiones, la industria debe ser capaz de identificar qué es lo mejor para su negocio y dedicar su esfuerzo para alcanzar los objetivos planteados con el apoyo de todos los recursos a su alcance.

CONCLUSIONES Y RECOMENDACIONES

- En conclusión la investigación ha mostrado que el análisis de los datos estructurados y no estructurados es una de las mejores opciones para la generación de conocimiento, al orientar la toma de decisiones sobre la optimización e innovación de los productos y servicios al cumplimiento de los requerimientos esperados por el cliente. La industria puede beneficiarse de los datos capturados desde diferentes medios de comunicación e inclusive de los datos publicados en sus portales, blogs y otros, para conocer lo que los clientes esperan recibir.
- Big data debe transmitirse al interior del negocio como una solución accesible que incorpora técnicas y herramientas para el análisis e integración de diversas fuentes de datos para mejorar e incrementar las habilidades para la toma de decisiones en entornos críticos con un enfoque estratégico y generar conocimiento a partir de las experiencias adquiridas.
- Un adecuado ecosistema para Big Data debe considerar los servicios, la integración, el análisis de grandes cantidades de datos, los datos estructurados y no estructurados, la colaboración y el almacenamiento para mantener el equilibrio en los procesos de análisis de datos, inversión de recursos y generación de conocimiento en la industria.
- La participación del Área de TI en la implementación de Big Data debe ser una estrategia de negocio para apoyar el trabajo de los expertos en el análisis de los datos y en la administración de las herramientas; no se debe incluir al área de TI como el administrador o gestor de este tipo de proyectos pues es importante que su dirección este a cargo de un área de negocio o del negocio en general.
- En el caso en que el Área de TI deba encargarse de la gestión de un proyecto de Big Data, está, debe preocuparse de que la organización reconozca el análisis de los datos como un servicio útil para la toma de decisiones, pues, al identificar el valor que el estudio de la información entrega al negocio será más fácil conseguir el apoyo y los recursos para dar continuidad al proyecto.

- La generación de valor a partir del análisis de los datos dependerá de la selección adecuada de las técnicas y plataforma que la industria va a implementar, al igual que de la calidad de los datos con los que se cuente para la construcción de los conjuntos de datos para el análisis. El tiempo que se tarde el negocio en obtener el valor del análisis de los datos dependerá de factores como el costo y la complejidad del análisis.
- Una industria debe considerar dentro de la estrategia de Big Data la recolección, la integración, el procesamiento, el análisis, la generación de conocimiento, la anticipación a la competencia, y las acciones para alcanzar los objetivos planteados por la alta gerencia.
- Para lograr la definición adecuada de un caso de aplicación para Big Data la organización debe considerar aspectos como la innovación, experiencia, excelencia operacional, optimización de las actividades de marketing, mejorar la experiencia del consumidor o la generación de ingresos con la finalidad de lograr resultados que impacten positivamente en el negocio.
- Es necesario concientizar a la organización que ha decidido implementar Big Data sobre el desarrollo de este tipo de proyectos y el impacto de los resultados generados, donde, generalmente se requiere el rediseño de los procesos de negocio o la innovación del modelo de negocio.
- Al igual que otros proyectos de innovación, la organización que emprende en Big Data debe enfrentar problemas como la resistencia al cambio, cambio de roles, movimiento temporal o permanente de personal, entre otros, por lo que es necesario considerar programas de capacitación continua y mantener los canales de comunicación abiertos entre el personal del negocio y el personal encargado del proyecto de Big Data.
- Antes de emprender en una iniciativa de Big Data la organización debe ser consciente que existe un proceso de evolución por el que debe pasar antes de lograr una implementación exitosa, por lo que es recomendable una análisis previo donde se establezca en qué nivel de madurez del proceso de análisis se encuentra la misma.
- La industria debe ser consciente que el análisis de datos sin una arquitectura y las normas de gobierno adecuadas, no podrá generar valor

para el negocio debido, a las limitaciones para realizar las validaciones de los conjuntos de datos seleccionados a causa de la falta de calidad que existirá en ellos. La selección de los conjuntos de datos solo se logrará estableciendo una cultura de datos y el gobierno de los mismo para que la industria conozca los datos que le son de utilidad y que datos debe desechar.

- La calidad de los datos es una característica que la industria no debe olvidar, pues, el valor que los datos entreguen al negocio dependerá de las técnicas que se aplique para la adquisición, clasificación e integración de los mismos.
- Una adecuada integración y gobierno de los datos se logra incluyendo el descubrimiento automático, identificación y comprensión de los conjuntos de datos para proveer al negocio información que permita tomar decisiones que mejoren las oportunidades del mismo.
- No todas las técnicas son aplicables para todos los casos de estudio, es decir, cada organización debe establecer de forma clara sus objetivos con la finalidad de garantizar el alineamiento de las estrategias de negocio con las de la implementación de Big Data; de esta manera la organización podrá seleccionar una técnica apropiada para el análisis de los datos orientado al tipo de productos esperados.
- La valoración de las características requeridas por el negocio puede proporcionar información relevante que permita seleccionar una técnica de Big Data alineada al nivel de madurez de la organización respecto a sus procesos y al análisis de datos.
- Un buen punto de partida para la selección de una técnica de Big Data es la investigación de las recomendaciones de empresas especializadas como Gartner que cuentan con variedad de documentos que permitirán ampliar el conocimiento del negocio para orientarlo a tomar una decisión
- En base al estudio realizado sobre las técnicas de análisis para Big Data se debe recomendar a las técnicas de análisis semántico ya que presentan una gran cantidad de cualidades para la integración, procesamiento y

análisis de los datos alcanzado 58 puntos de los 64 puntos establecidos en la valoración de las características seleccionadas para la calificación.

- Si bien las técnicas de análisis estadístico y de visualización alcanzaron 43 puntos en la calificación realizada, se debe recomendar su utilización en los casos que la organización se encuentre empezando con una iniciativa de Big Data, pues, los costos de implementación para el negocio reflejarán menor impacto frente al costo de aplicación de una técnica con nivel de complejidad más elevado, en los casos que el proyecto fracase.
- La selección de las herramientas para el análisis de los datos orientado a la aplicación de Big Data debe realizar de forma crítica y evaluando características como la Disponibilidad, Consistencia y Tolerancia a fallos.
- Para seleccionar una base de datos NoSQL en un proyecto de Big Data la organización debe considerar las ventajas y desventajas de las soluciones con mayor importancia en el mercado, pues, de esta manera garantizará la existencia de información para la investigación y el mantenimiento de la herramienta.
- Herramientas como Hadoop se pueden utilizar en organizaciones que desean trabajar de forma más estructurada respecto al procesamiento y almacenamiento de los datos, al igual que en la gestión de los paquetes incluidos en este ecosistema.
- Para la selección de las herramientas que apoyen la implementación de Big Data y sus técnicas es necesario que la organización evalúe las opciones presentes en el mercado y considere factores como el tipo de licencia, soporte en diferentes sistemas operativos, o al menos que utiliza la organización, capacitación, acceso a información entre otros.
- Como resultado de esta investigación se han establecido recomendaciones para el uso de bases de datos NoSQL, entre las cuales se han identificado a MongoDB y Cassandra como las soluciones que presentan mayores beneficios para el almacenamiento y acceso a los datos.
- Se recomienda el uso de MongoDB en casos donde se requiera el almacenamiento de datos provenientes de aplicativos móviles y en general de páginas web

- Se recomienda el uso de Cassandra para el almacenamiento de datos provenientes principalmente de redes sociales y campaña de marketing.
- Las organizaciones que emprenden en Big Data deben ser conscientes que los proyectos orientados a este tipo de soluciones no siempre presentan resultados agradables para el negocio y por tanto se debe preparar para manejar este tipo de escenarios e identificar los aspectos positivos de la implementación para procesos de mejora continua.

GLOSARIO

Big Data.- Aplicada para el análisis de grandes cantidades de datos provenientes de fuentes diversas a través de la implementación de aplicaciones que permiten gestionar y manipular la data para convertirla en información útil para la organización (QUE ES BIG DATA, 2015).

Cloud Computing.- Consiste en la posibilidad de ofrecer servicios a través de Internet.

Facebook.- Es el sitio web de redes sociales más popular en la actualidad, fue lanzado en el año 2004.

Flickr.- Sitio de internet que permite mantener ordenados fotos y videos, brinda el servicio de almacenamiento, búsqueda y venta en línea de los mismos (Calle, 2015).

Forrester.- Es una empresa dedicada a la investigación de mercados ofreciendo servicios de asesoría para visualizar el impacto existente y potencial de las tecnologías sobre sus clientes.

Foursquare.- Servicio aplicado en las redes sociales para la geolocalización de un dispositivo en base a su ubicación geográfica (González, 2011).

Gartner.- Es una de las consultoras más importantes del mundo, se dedica a la investigación de las tendencias en el uso de tecnologías y herramientas informáticas (Brito, 2014).

Google.- Es una empresa multinacional que se dedica a la prestación de servicios y el desarrollo de productos relacionados principalmente con software, motores de búsqueda, dispositivos móviles entre otros.

IBM.- “*International Business Machines*” por sus siglas en inglés, es una de las empresas más importantes del mundo, dedicada a la fabricación y comercialización de todo tipo de productos informáticos (IBM, 2013)

INEC.- Instituto Nacional de Estadísticas y Censos.

Instagram.- Es la red social que permite publicar y compartir fotografías entre diferentes usuarios, brinda algunas opciones entre las que se destacan sus facilidades para agregar efectos a las imágenes (Ortega, 2014).

Java.- Es un lenguaje de programación utilizado para la codificación de aplicaciones de entorno web que soportan alta disponibilidad y acceso múltiple en línea (Oracle, 2015).

KPI.- “*Key Performance Indicator*”, por sus siglas en inglés se define como la métrica utilizada para cuantificar el nivel de cumplimiento de las metas establecidas en una organización” (Project Management Institute, Inc. , 2008).

Map Reduce.- Es una técnica de programación aplicada que permite realizar dos acciones: la primera es mapear los conjuntos de datos en otros conjuntos compuestos por un valor clave y el valor relacionado del conjunto de datos principal y la segunda corresponde al ordenamiento y clasificación de los conjuntos para extraer los más representativos.(Trejo, 2014).

Ministerio de Turismo.- Institución pública encargada de regular al sector turístico en el Ecuador.

NoSql.- “Es una amplia clase de sistemas de gestión de bases de datos que difieren del modelo clásico del sistema de gestión de bases de datos relacionales (RDBMS) en aspectos importantes, el más destacado es que no usan SQL como el principal lenguaje de consultas. Los datos almacenados no requieren estructuras fijas como tablas” (wikipedia, 2015).

Pmbok.- Es una guía enfocada en la gestión y administración de proyectos donde se establecen parámetros para estructurar de forma adecuada y rentable la planificación de un proyecto y garantizar el éxito en el desarrollo y entrega del mismo (Project Management Institute, Inc. , 2008).

SCRUM.- Metodología compuesta de buenas prácticas enfocadas en el desarrollo ágil y colaborativo principalmente de proyectos de software, actualmente es uno de los movimientos con mayor acogida en este entorno (Crawford, 2013).

Semántica.- Término que permite interpretar el significado de los signos identificando las relaciones de este con los demás (Semántica, 2013).

SQL.- “*Structured Query Language*”, por sus siglas en inglés, es el lenguaje utilizado para gestionar y administrar las bases de datos relacionales y los datos contenidos en diferentes estructuras (Leiva, 2014).

Stakeholders.- Generalmente se refiere al miembro o miembros internos o externos de una organización que pueden verse afectados por una decisión de la misma (Lalinde, 2014).

Teorema de CAP.- Asegura que un sistema de datos distribuidos puede ofrecer al menos dos de las siguientes propiedades: Fuerte Consistencia, Alta Disponibilidad o Tolerancia a Particiones.

Tumblr.- Plataforma de microblogs que soporta la creación de blogs visualmente atractivos, manejo de plantillas, integración con redes sociales y publicación de enlaces (Ortega, 2014).

Twitter.- Es una red social que sirve para establecer diferentes estados, publicar información o hacer comentarios cortos sobre diferentes eventos de una persona u organización (Butron, 2014).

WordPress.- Es un gestor de contenido o CMS (por sus siglas en inglés, Content Management System) utilizado generalmente para la creación de blogs (sitios web periódicamente actualizados).

Youtube. - Es un portal de Internet para la carga, descarga y visualización de todo tipo de videos en línea.

REFERENCIAS

- Agrawal, R., Imielinsk, T., & Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. SIGMOD, (págs. 207-216).
- Alegria, A. (2014). Big Data y su aplicación al turismo para ofrecer una mejor experiencia. . Obtenido de vuelo digital: <http://www.vuelodigital.com/big-data-y-su-aplicacion-al-turismo-para-ofrecer-una-mejor-experiencia>.
- Australia: Departamento de Finanzas y Desregularización. (2013). Big Data Strategy. Commonwealth of Australia 2013
- Barranco Fragoso, R. (18 de Junio de 2012). Qué es Big Data. Recuperado el 9 de Enero de 2014, Obtenido de IBM developerWorks: <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>
- Barlow, M. (2013).Real Time Big Data Analytics: Emerging Architecture. United Sates of America: O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472
- Brito, D. (10 de Diciembre de 2014). Inteligencia de Negocios. Obtenido de Inteligencia de Negocios: <http://inteligenciadenegociosdiegobrito.blogspot.com/>
- Butron, B. (21 de Octubre de 2014). Trabajo de asignación de sitios web. Obtenido de slideshare: <http://es.slideshare.net/andybkmamani1/trabajo-de-informatica-brayen-butron>
- Buytendijk, F., & Laney, D. (12 de Marzo de 2014). Information 2020: Big Data and Beyond. Recuperado el 20 de Marzo de 2014, Obtenido de Gartner Webinars: <https://www.gartner.com/doc/2681316/information--big-data>
- Calle, K. (22 de Enero de 2015). Obtenido de slideshare: <http://es.slideshare.net/michucalle1997/slides-hare-karen-calle>
- Chui, M., Mille, A., & Roberts, R. (2009). Six ways to make Web 2.0 work. McKinsey Quarterly. Obtenido de Michael Chui, Andy Miller, and Roger Roberts, "Six ways to make Web 2.0 work," McKinsey .

- clcanela. (s.f.). ¿Qué es JSON y dónde se usa? Obtenido de canela.me:
<http://canela.me/articulo/%C2%BFqu%C3%A9-es-json-y-d%C3%B3nde-se-usa>
- Crawford, T. (2013). Big Data Analytics Project Management. United States of America.
- Deloitte. (2013). Que hacer para evolucionar hacia Big Data en México. México: Deloitte Consulting Group
- Desouza, K. (8 de Julio de 2014). Como implementar un proyecto de Big Data. Recuperado el 24 de septiembre de 2014. Obtenido de CIO Peru:
<http://cioperu.pe/articulo/16330/como-implementar-un-proyecto-de-big-data>
- Dijcks Jean-Pierre. (2013). Oracle: Big Data for the Enterprise. USA: Oracle Corporation
- EMC Corporation. (2011). Big Data: Big Opportunities to Create Business Value. Hopkinton Massachusetts. Obtenido de EMC.
- González, R. (7 de Noviembre de 2011). Posiciona tu empresa a través de las Redes Sociales. Recupedado el 20 de febrero de 2015. Obtenido de slideshare:
<http://www.slideshare.net/roberdevigo/posiciona-tu-empresa-a-travs-de-las-redes-sociales>
- Hopkins, B. (11 de Septiembre de 2013). Don't have a big data strategy yet? Good. Obtenido de Forrester: http://blogs.forrester.com/brian_hopkins/13-09-11-dont_have_big_data_strategy_yet_good
- Ibarra Eduardo. (2014). Minería de datos en supermercados. Sitio web:
<http://nanouam.tripod.com/sitebuildercontent/sitebuilderfiles/supermineria.pdf>. Consultado en noviembre 10, 2014, de ND.
- IBM. (20 de Octubre de 2013). Obtenido de Diccionario de Informática:
<http://eldiccionariodeinformatica.blogspot.com/2010/09/con-la-letra-i.html>
- IBM. (2014). Mejore la tomada de decisiones con la analítica de contenido. Recuperado el 20 de diciembre de 2014. Obtenido de IBM:
http://www.ibm.com/smarterplanet/ec/es/business_analytics/article/it_business_intelligence.html

- IBM Institute for Business Value, Escuela de Negocios Saïd en la Universidad de Oxford. (2012). Analytics: el uso de big data en el mundo real. España: IBM Corporation
- IDC Customer Spotlight. (2011). Whirlpool corporation's digital detectives: Attensity provides the lens.
- Inc. TechRepublic. (2013). Executive's Guide to Big Data strategies and best practices. CBS Interactive Inc.
- Instituto Nacional de Estadísticas y Censos, INEC. (2013). Tecnologías de la Información y Comunicaciones (TIC'S) 2013. www.ecuadorencifras.gob.ec
- Krishnan, K. (2013). Data Warehousing in the Age of Big Data. United States of America: Morgan Kaufmann.
- Lalinde, S. (23 de Febrero de 2014). EL COMPROMISO CON LOS STAKEHOLDERS COMO OBJETIVO DE LA COMUNICACIÓN ERRORES PARA NO COMETER. Recuperado el 20 de diciembre de 2014. Obtenido de Club Ensayos: <http://clubensayos.com/buscar/STAKEHOLDERS /pagina1.html>
- Lancis, E. (2014). Big Data aplicada al sector turístico. Obtenido de SEGITTUR: <http://www.segittur.es/es/sala-de-prensa/detalle-documento/Presentacin-Big-Data-aplicada-al-sector-turstico-/#.VSIBrfmG9FU>
- Laney, D. (2013). Big Data and Analytics Strategy Essentials. Gartner Research Inc.
- Lapkin, A. (2012). Hype Cycle for Big Data, 2012. USA: United States of America: Gartner, Inc. and/or its Affiliates.
- Leiva, J. (11 de Abril de 2014). HISTORIA DE LOS COMPUTADORES. Obtenido de Prezi: <https://prezi.com/xiumljoessbx/untitled-prezi/>
- López de Ipiña, D. (17 de Julio de 2013). NoSQL: Introducción a las bases de Datos no estructuradas. Recupedado el 25 de noviembre de 2014. Obtenido de slideshare: <http://eventos.citius.usc.es/bigdata/>
- Loshin, D. (2013). Integrating structured and unstructured data. Obtenido de TDWI (The Data Warehousing Institute).

- Manyika, J., Chui Michael, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh Charles, C., & Hung Byers, A. (Mayo de 2011). Big data: The next frontier for innovation, competition, and productivity. Estados Unidos de Norteamérica: McKinsey&Company. Recuperado el 20 de Enero de 2014, de McKinsey Global Institute.
- Merino, M. (2014). ¿Qué son los mapas de calor? Obtenido de ticbeat: <http://www.ticbeat.com/tecnologias/que-son-los-mapas-de-calor/>
- Metodología de la Investigación. (2014). Metodología de la Investigación. Recuperado el 20 de Enero de 2014. Obtenido de http://catarina.udlap.mx/u_dl_a/tales/documentos/lad/arenas_m_a/capitulo3.pdf
- Ministerio de Turismo del Ecuador. (2012). La experiencia turística en el Ecuador. Quito: Dirección de Investigación.
- Ministerio de Turismo del Ecuador. (2014). Principales Indicadores de turismo. Coordinación General de Estadística e Investigación. www.turismo.gob.ec.
- Mitchell Ian, Locke Mark, Wilson Mark, Fuller Andy. (2012). The White Book of Big Data. Fujitsu Services Ltd
- MySQL. (2015). Executive Guide: Enabling Digital Transformation with MySQL. Obtenido de MySQL: <https://www.mysql.com/why-mysql/white-papers/executive-guide-enabling-digital-transformation-with-mysql/>
- O'Reilly Media, Inc. (2012). Big Data Now: 2012 Edition. United States of America: Gravenstein Highway North: Sebastopol, CA95472.
- O'Reilly Media. (2011). Big Data Now. United States of America: Gravenstein Highway North, Sebastopol, CA 95472
- O'Reilly Radar Team. (2012). Planning for Big Data. United States of America: O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472
- Oracle. (2015). ¿Qué es Java?. Recuperado en enero de 2015. Obtenido de Java: https://www.java.com/es/download/whatis_java.jsp

- Ortega, J. (16 de Febrero de 2014). Las redes sociales. Obtenido de slideshare: <http://www.slideshare.net/ortegaj/las-redes-sociales-jose-ortega>
- Project Management Institute, Inc. . (2008). GUÍA DE LOS FUNDAMENTOS PARA LA DIRECCIÓN DE PROYECTOS. EEUU: 14 Campus Boulevard Newtown Square, Pennsylvania 19073-3299.
- QUE ES BIG DATA. (2015). Obtenido de Big Data el nuevo paradigma de la información: <http://www.25imagenes.es/>
- Rajpathak, T., & Narsingpurkar, A. (2013). Managing Knowledge from Big Data Analytics in Product Development. Obtenido de Tata Consultancy Services Limited.
- Revista LÍDERES. (2013). El big data le hace un zoom al consumidor. Recuperado el 15 de Septiembre de 2014. Obtenido de Revista LÍDERES: <http://www.revistalideres.ec/lideres/big-data-le-zoom-consumidor.html>
- Rivera, J., Van der Meulen, R. (2012). Gartner's Hype Cycle for Emerging Technologies Maps the Journey to Digital Business. STAMFORD, Conn.
- Rivera, J., Van der Meulen R. (2013). Gartner's Hype Cycle for Emerging Technologies Maps the Journey to Digital Business. STAMFORD, Conn.
- Rivera, J., Van der Meulen R. (2014). Gartner's Hype Cycle for Emerging Technologies Maps the Journey to Digital Business. STAMFORD, Conn.
- Rodríguez, A., Contreras, R., López, C., Mandianes, Á., Minaya, H. (2012). En la Era del Big Data. Computing. España: BPS
- Russom, P. (2013). Aclarando 10 mitos sobre Hadoop. . Obtenido de <http://www.sas.com/offices/europe/spain/sascom/Q42013/Hadoop.pdf>
- Secretaría Nacional de Planificación y Desarrollo, SENPLADES. (2012). Transformación de la Matriz productiva Revolución productiva a través del conocimiento y el talento humano. Quito: EDIECUATORIAL.

- SEGITTUR. (2012). Turismo sostenible. Obtenido de SEGITTUR:
<http://www.segittur.es/es/areas-conocimiento/>
- Selvage, M. (24 de Julio de 2013). Patterns for Integrating Structured and Unstructured Data. Recuperado el 20 de Enero de 2014, de Gartner Webinars: <http://www.gartner.com/webinar/2588915>
- Semántica. (23 de Junio de 2013). Obtenido de Definiciones: <http://definicion.de/semantica/>
- Simon, P. (2013). Too Big to Ignore. United States: New Jersey: John Wiley & Sons.
- solid IT. (Febrero de 2015). DB-Engines Ranking. Obtenido de db-engines: <http://db-engines.com/en/ranking>
- Stephenson Debbie. (2013). 7 Big Data Techniques That Create Business Value. <http://www.firmex.com/blog/7-big-data-techniques-that-create-business-value/>. Consultado el 24 de septiembre de 2014.
- Sun Helen, Heller Peter. (2012). Oracle Information Architecture: An Architect's Guide to Big Data. USA: Oracle Corporation
- Soumendra, M., Madhu, J., & Harsh, S. (2013). Big Data Imperatives. Estados Unidos de América: Nueva York: Apress.
- Swisher, R. (2014). Fast, Good or Cheap. Pick Three? Obtenido de business.com: <http://www.business.com/management/fast-good-cheap-pick-three>
- TechAmerica Foundation. (2012). DEMYSTIFYING BIG DATA. Washington, D.C: 601 Pennsylvania Avenue, N.W. North Building, Suite 600.
- Tecnologías de la Información. (2009). Minería de Datos. Obtenido de Tecnologías de la Información: <http://www.tecnologias-informacion.com/mineria-de-datos.html>
- Torres iViñals Jordi. (2012). Del cloud computing al big data Barcelona: Eureka Media, SL
- Trejo, A. (21 de Agosto de 2014). Open Smart Cities: Tecnologías de fuentes abiertas para ciudades inteligentes. Obtenido de slideshare: <http://www.slideshare.net/anatrejo/open-smart-cities-tecnologas-de-fuentes-abiertas-para-ciudades-inteligentes>

- TripAdvisor. (2014). TripAdvisor. Obtenido de TripAdvisor:
http://www.tripadvisor.es/PressCenter-c6-About_Us.html
- Urquizu, P. (21 de Junio de 2009). Breve historia del Business Intelligence. Recuperado el 9 de Enero de 2014., de business Intelligence Fácil:
<http://www.businessintelligence.info/definiciones/historia-business-intelligence.html>
- usun, o. (5 de Septiembre de 2011). El tratamiento masivo de datos (Big Data): La próxima frontera para la innovación, la competencia, y la productividad. Obtenido de Ibermática: <http://rtdibermatica.com/?tag=mineria-de-datos>
- wikipedia. (2015). NoSQL. Obtenido de wikipedia:
<http://es.wikipedia.org/wiki/NoSQL>
- Williams, P. (2012). Unstructured data and the enterprise. Estados Unidos de América: California: Smartlogic US.
- Zikopoulos, P. C., Eaton, C., deRoos, D., Deutsch, T., & Lapis , G. (2012). Understanding Big Data. United States of America: McGraw-Hil

ANEXOS

1. Caso de aplicación de minería de datos²

(Ibarra, 2014) “En supermercados Wal-Mart, decidieron iniciar un proyecto de *basket analysis* utilizando la ingente cantidad de información contenida en su Teradata datawarehouse. Inicialmente, los resultados no parecieron muy espectaculares, ya que vieron que quien compraba pasta dentífrica también compraba cepillos de dientes; tampoco sorprendió que quien comprase whisky DYC inevitablemente adquiriese Coca cola. La sorpresa fue observar una correlación estadísticamente significativa entre la compra de pañales y cerveza.

Incorporaron la información de sus tarjetas de fidelización para profundizar en el estudio, y vieron que los compradores de cerveza y pañales eran varones de entre 25 y 35 años, que solían comprar estos productos conjuntamente los viernes por la tarde.

En base a estos datos, fue fácil dar una explicación al fenómeno, y tomar medidas comerciales para aumentar las ventas, y justificar así el sueldo de todos estos analistas. Como los pañales son voluminosos, las madres enviaban a los padres de las criaturas a comprar los pañales, estos esperaban a último momento para hacerlo, y aprovechaban para comprar cervezas que se tomarían en casa, ya que en la sociedad americana está mal visto que un padre de familia frecuente los bares hasta tarde, y vuelva a casa mareado. Con este antecedente tomaron la decisión de colocar las cervezas cerca de los pañales, con la intención de que los padres que compraban pañales y que no solían comprar cerveza, se acordasen que faltaba cerveza en casa.

Los resultados fueron espectaculares y aumentaron entre un diez y un quince por ciento tanto las ventas de cervezas como de pañales”.

² Ibarra Eduardo. (2014). Minería de datos en supermercados. Consultado en noviembre 10, 2014, de ND Sitio web: <http://nanouam.tripod.com/sitebuildercontent/sitebuilderfiles/supermineria.pdf>

2. Carta de auspicio

Quito, 18 de Diciembre de 2013

Señor Ingeniero
Germán Pancho
Director de la Maestría en Gerencia de Sistemas y Tecnologías de la Información
Universidad de las Américas

Presente.-

De mi consideración:

Por medio de la presente, quiero comunicarle en mi calidad de Director de Tecnologías de la Información y Comunicaciones del Ministerio de Turismo, que me comprometo a otorgar el auspicio a la señorita ingeniera Sonia Carolina Guama Morales, para el desarrollo del proyecto de tesis denominado **“Estudio comparativo de métodos existentes para integrar la información estructurada y no estructurada de una industria enfocado en la generación de conocimiento, desde la perspectiva de una solución integral de Big Data”**, bajo las siguientes condiciones:

- La Dirección de Tecnologías de la Información y Comunicaciones del Ministerio de Turismo se compromete, de ser el caso, a entregar el apoyo logístico necesario para que el mencionado proyecto pueda aplicarse a la realidad de una institución pública.
- La información que se entregue será clasificada por la Dirección de Tecnologías de la Información y Comunicaciones del Ministerio de Turismo y podrá ser divulgada solo con autorización expresa.
- Se llevará seguimiento al estudio efectuado, comprometiéndose a entregar una carta de conformidad al finalizar el proyecto.
- Aceptamos que la propiedad intelectual es de la UDLA, sin embargo el trabajo puntual podrá ser utilizado por la Dirección de Tecnologías de la Información y Comunicaciones del Ministerio de Turismo sin requerimiento de autorización alguna.

Atentamente,

Ing. Germán Maroto B.
Director de Tecnologías de la Información y Comunicaciones
Ministerio de Turismo

3. Informe Ministerio de Turismo

Big Data: Aplicación y oportunidades para el Turismo del Ecuador

Sobre el Autor

Autor

Sonia Carolina Guama Morales

Profesor Guía

Jaime Vinueza

Universidad de las Américas

Maestría en Gerencia de Sistemas y Tecnologías de la Información

2014

El presente informe fue desarrollado para el uso exclusivo del Ministerio de Turismo del Ecuador. No se permite el uso comercial de la obra original ni de las obras derivadas. No se permite la copia o difusión sin la autorización expresa del autor o del Ministerio de Turismo.

ÍNDICE

1. Generalidades	1
2. Propósito del Informe.....	5
3. Descripción del Problema	5
4. Ventajas de la aplicación de Big Data en el Turismo	6
5. Análisis Detallado	7
6. Recomendaciones	12

1. Generalidades

Para iniciar el análisis de la aplicación de Big Data en el sector turístico y la identificación de los beneficios para el sector es necesario dar a conocer algunas definiciones generales respecto a este tipo de soluciones. La información proporcionada en esta sección no pretende ser una guía de Big Data y se recomienda el estudio de bibliografía específica para ampliar el conocimiento sobre el tema.

Big Data es una de las soluciones más innovadoras y atractivas del mercado para la generación de oportunidades y mejorar la toma de decisiones basados en el análisis de los datos. Big Data se define como el análisis de grandes volúmenes de datos provenientes de una variedad de fuentes que comprenden datos estructurados y no estructurados a gran velocidad considerando resultados en tiempo real o casi real y que requiere del análisis de la veracidad de los datos analizados.

Los datos estructurados provienen de fuentes relacionales como bases de datos SQL, la principal diferencia con los datos no estructurados es el orden y extensión de los mismos. Los datos no estructurados son datos provenientes de redes sociales, documentos, imágenes y videos que requieren de la aplicación de técnicas y herramientas para su integración y análisis.

Existen varias técnicas de análisis que se clasifican en estadísticas, visualización de datos, automatización, predictivas, y semánticas. La tabla a continuación muestra un resumen de las técnicas mencionadas:

Tabla 1. Resumen de técnicas de análisis aplicables en Big Data

Técnica	Tipo	Características	Usabilidad
Minería de Datos	Estadística	Diseño de modelos matemáticos para el descubrimiento de eventos inusuales, que no pueden apreciados fácilmente	<ul style="list-style-type: none"> • Segmentación de clientes • Detectar patrones de comportamiento y tendencias • Predicción de comportamientos futuros
Clasificación	Estadística	<ul style="list-style-type: none"> • Identificación de categorías • Se requiere de habilidades para la identificación de las categorías en base a observaciones de pertenencia y relacionamiento • Se incluye en los conjuntos de datos a los datos históricos. 	<ul style="list-style-type: none"> • Asignación automática de objetos • Categorización • Desarrollo de perfiles
Reglas de Asociación	Estadística	Descubrimiento de correlaciones entre variables de las fuentes de datos	<ul style="list-style-type: none"> • Estrategias para incrementar ventas de productos • Extraer información de diversas fuentes de datos para detectar patrones de comportamiento en diferentes áreas
Análisis de Clúster	Estadística	<ul style="list-style-type: none"> • Agrupamiento de objetos en clases en base a sus similitudes • Se trabaja con varias dimensiones para el análisis • Trabaja con patrones para establecer la similitudes entre objetos 	<ul style="list-style-type: none"> • Definición de modelos de negocio • Descubrimiento de patrones
Redes neuronales	Estadística	<ul style="list-style-type: none"> • Procesamiento de una gran cantidad de variables • Capacidad para detectar relaciones no lineales • Análisis de información en bruto, es decir sin procesos de simplificación y en tiempo real • Capacidad para auto-aprender y ajustarse de forma independiente para análisis futuros. • Requieren más trabajo de exploración que otras técnicas 	<ul style="list-style-type: none"> • Identificar patrones ocultos • Construcción de modelos analíticos • Procesamiento de nuevos conjuntos de datos en tiempo real

A/B Testing	Estadística	Comparación entre un grupo de datos de control y un grupo de datos de prueba para identificar diferencias y semejanzas ante diferentes variaciones establecidas en el análisis	<ul style="list-style-type: none"> • Aceptación de cambios en productos y servicios • Introducción nuevos de productos • Diseño de campañas de marketing • Diseño de marcas y páginas web
Regresión	Estadística	Manipulación de variables dependientes para visualizar la influencia sobre las variables dependientes	<ul style="list-style-type: none"> • Niveles de satisfacción y lealtad sobre los productos y servicios • Identificación de comportamientos en base a la variación de variables independientes • Establecimiento de precios en base a factores externos • Identificar relaciones en base a características similares o dependientes
Mapas de calor	Visualización de Datos	Representación visual de los datos en 2 dimensiones, en el que los valores están representados por colores	<ul style="list-style-type: none"> • Proporcionar un resumen visual de la información • Comprensión de conjuntos de datos complejos
Aprendizaje de Máquina	Automatización	<ul style="list-style-type: none"> • Uso de software para facilitar el aprendizaje de la maquina desde los datos. • Análisis autónomo debido al entrenamiento continuo proporcionado por los datos 	<ul style="list-style-type: none"> • Identificar diferencias entre objetos • Identificar preferencias y recomendaciones de selección de productos • Predicción de probabilidades
Procesamiento de lenguaje natural	Semántica	<ul style="list-style-type: none"> • Requiere de alta precisión para el análisis • Análisis de grandes cantidades de textos • Análisis sintáctico y morfológico • Uso de patrones lingüísticos 	<ul style="list-style-type: none"> • Extracción de entidades y clasificación de los datos • Extraer hechos o conocimiento a partir de texto libre

Análisis de textos	Semántica	<ul style="list-style-type: none"> • Identificación de cualidades de los textos • Brinda acceso a la información plasmada en diferentes tipos de textos • Análisis de palabras y documentos 	<ul style="list-style-type: none"> • Identificación de emociones • Mapeo de conductas, perfiles, etc. • Extracción de índices numéricos • Automatización en el procesamiento de correos
Análisis de sentimientos	Semántica	<ul style="list-style-type: none"> • Uso de técnicas analíticas para extraer información subjetiva de los datos • Identificación de sentimientos respecto a un tópico 	<ul style="list-style-type: none"> • Clasificación de opiniones • Mejoramiento de servicios • Personalización de productos y servicios • Identificación de comportamientos
Fusión e integración de datos	Semántica	Fusión e integración de varias fuentes de datos para ampliar el campo de análisis	<ul style="list-style-type: none"> • Determinar niveles de confianza • Identificar patrones de comportamiento
Análisis predictivo	Predictiva	Análisis de datos actuales e históricos para hacer predicciones de eventos futuros	<ul style="list-style-type: none"> • Identificación de patrones de comportamiento • Identificación de potenciales riesgos y consecuencias considerando su aplicación en la toma de decisiones • Identificación de relaciones entre clientes y productos • Identificación de tendencias
Optimización	Predictiva	<ul style="list-style-type: none"> • Se basa en modelos de decisión para describir la relación entre todos los elementos de una decisión. • Considera variables, valores, decisiones previas para optimizar la toma de nuevas decisiones 	<ul style="list-style-type: none"> • Mejoramiento de procesos • Análisis de inversión • Mejoramiento de la toma de decisiones
Simulación	Predictiva	Permite modelar el funcionamiento de sistemas complejos	<ul style="list-style-type: none"> • Predicciones sobre resultados • Diseño de escenarios • Predicciones de eventos futuros

La selección de una de estas técnicas dependerá de las necesidades análisis que requiera el caso de aplicación establecido por el sector turístico.

2. Propósito del Informe

Establecer una guía básica para la aplicación de Big Data en el sector turístico y dar a conocer los beneficios que la aplicación de esta solución puede ofrecer al sector. El análisis de los datos y la gestión adecuada del conocimiento generado es la clave para identificar las oportunidades y plasmarlas en decisiones adecuadas y rentables para el turismo.

Explorar las oportunidades que puede ofrecer al sector turístico e incorporar el análisis de datos como una estrategia para promover su desarrollo e innovación.

El análisis de grandes cantidades de datos provenientes de diversas fuentes brinda una nueva opción para que las empresas turísticas puedan mejorar e innovar los servicios que ofrecen y la gestión de en la toma de decisiones para el negocio.

En el caso de este estudio se pretende promover el desarrollo del turismo a través del mejoramiento de los servicios y el impulso a proyectos basados en el análisis de datos.

3. Descripción del Problema

En la actualidad el sector turístico genera una gran cantidad de datos que podrían generar valor para promover el desarrollo del turismo si se promoviera la toma de decisiones basadas en el análisis de datos.

Los datos provienen de la web, redes sociales y sistemas transaccionales que trabajan de forma independiente, lo cual, limita el campo de análisis y por tanto, no permite visualizar oportunidades integrales que se puedan aplicar de forma general para beneficiar el desarrollo del turismo en el país.

Otro problema identificado es la ausencia de servicios automatizados para el sector turístico; Ecuador cuenta con algunos sitios en la web que proporcionan información turística del país como <http://ecuador.travel> que muestra al mundo

la oferta de servicios y productos entregados por los diferentes sectores del país. Estos sitios han sido de gran ayuda para presentar un Ecuador turístico al mundo; sin embargo aún no se han implementado opciones para facilitar el acceso a los servicios y productos antes de visitar el país.

La ausencia de aplicaciones web y móviles para el turista, desaprovecha la oportunidad de capturar datos respecto a la experiencia percibida sobre la calidad de los servicios recibidos y su estancia en el país.

La falta de una plataforma para integrar los datos de todas las instituciones públicas es otra limitante para el desarrollo del turismo, ya que, el acceso a la información facilita la identificación de oportunidades y la generación de conocimiento en este caso para impulsar el turismo en el país.

Finalmente, no existe en el país una cultura de trabajo colaborativa ni basada en el análisis de los datos, por lo cual se limita la generación de conocimiento y el desarrollo de habilidades para identificar oportunidades y tomar decisiones que aporten valor al sector turístico.

4. Ventajas de la aplicación de Big Data en el Turismo

Las ventajas más representativas que puede obtener el turismo con la aplicación de Big Data son las siguientes:

Reducción de costos.- Al establecer procesos de selección de datos y una cultura de colaboración para facilitar el acceso a la información compartida, pueden reducir costos en el mantenimiento y gestión de los datos, igualmente el ahorro de los recursos se puede generar a través del enfoque en el desarrollo de proyectos que sean atractivos para el turista y rentables para el turismo.

Fiabilidad de los datos.- Una adecuada selección de datos permitirá realizar un análisis con información fiable y proveer un entorno apegado a la realidad para la toma de decisiones.

Acceso a la información.- Mantener un entorno de colaboración facilita el acceso a la información y la generación de conocimiento.

Segmentación de mercado.- La aplicación de técnicas de análisis de datos permite identificar segmentos de mercado en los que se puede promover campañas y ofertar servicios existentes o desarrollar nuevos servicios en los que están interesados.

Personalización de productos y servicios.- El conocimiento generado a partir del análisis de los datos, permite la construcción de servicios y productos personalizados para cubrir las necesidades y requerimientos de los segmentos de mercado que visitan el país y que podrán llevarse una mejor experiencia turística del país al contar con este tipo de facilidades.

Métricas para el sector.- Medir los niveles de satisfacción, calidad, accesibilidad y atención, son algunas de las métricas que se pueden obtener a partir del análisis de los datos generados por los turistas.

Identificación de Tendencias.- Identificar tendencias es una ventaja que puede aprovechar el sector turístico para el desarrollo de oportunidades e innovación de los servicios y productos turísticos.

5. Análisis Detallado

La aplicación de Big Data es un proceso complejo, y requiere del apoyo de toda la industria para alcanzar buenos resultados. Las oportunidades de su aplicación son varias y dependiendo de las estrategias con las que se decida emprender en este tipo de proyectos se pueden alcanzar objetivos sociales y económicos de gran impacto.

La aplicación de Big Data debe alinearse a las estrategias de negocio en este caso para promover el desarrollo del turismo, considerando que puede aplicarse desde el sector público con el estudio de los datos para promover políticas que impacten sobre el sector y direccionen los recursos a facilidades atractivas y necesarias para acceder a los sitios turísticos.

Es importante que antes de emprender en el análisis de datos desde la perspectiva de Big Data se consideren los siguientes aspectos:

- Reconocer la importancia del análisis de los datos para la toma de decisiones
- Contar con procesos que permitan identificar las áreas donde el análisis de datos genere mayor impacto
- Identificar y contratar el personal necesario para dedicarse al proceso de selección, clasificación y análisis de los datos; este personal debe ir desde analistas a expertos de datos y contar con el apoyo del negocio
- Mantener una mentalidad abierta a los cambios y conocer que los resultados de un proyecto de Big Data pueden ser en algunos casos solamente el desarrollo de habilidades, de experiencia y de conocimiento.
- Analizar la capacidad de la infraestructura y las posibilidades de mejorarla o invertir en una nueva plataforma e inclusive la implementación de soluciones de cloud.

Gestión de un proyecto de Big Data

Una gestión adecuada de un proyecto de Big Data es una clave para alcanzar el éxito del proyecto, se fundamentan en el análisis de grandes cantidades de datos, por lo cual, el entorno de análisis es cambiante y esto exige flexibilidad ante los cambios en las fases del proyecto y en los plazos para la entrega de resultados.

Scrum es una de las metodologías ágiles más importantes y más utilizadas en la actualidad que se basa en el trabajo iterativo y colaborativo para el desarrollo de aplicaciones y la gestión de proyectos.

Los proyectos de Big Data pueden gestionarse de mejor manera con esta metodología ya que admiten el trabajo iterativo, es flexible ante el manejo de los cambios, gestión del factor humano, colaboración con el cliente y el desarrollo iterativo para la generación de resultados en proyectos con requerimientos cambiantes y plazos cortos. Además que permite gestionar el conocimiento entre los equipos y la organización para apalancar el proyecto y fortalecer sus bases.

La coordinación de actividades muestra una adecuada gestión sobre las reuniones diarias y los resultados reportados por cada miembro del equipo; la distribución adecuada del equipo para cada tarea y trabajo colaborativo; auditorias diarias del avance del proyecto para facilitar el cierres de tareas y lo más importante no se preocupa por temas irrelevantes o de mínima importancia.

Los proyectos de Big Data se gestionan de mejor manera con Scrum de acuerdo a las siguientes consideraciones:

- El éxito del proyecto requiere de comunicación entre los grupos participantes del proyecto
- El ciclo de vida de un proyecto de Big Data es más corto que el de un proyecto en cualquier otra área y se puede retroalimentar de la experiencia adquirida durante la investigación, es iterativo y exploratorio
- Requiere del apoyo incondicional de la alta gerencia
- Tiempos de implementación flexibles pero no ilimitados
- Requiere de personal con habilidades específicas en el análisis de datos pero que pueda colaborar en otras actividades del proyecto
- El alcance del proyecto puede modificarse de acuerdo al avance en las investigaciones y a los resultados generados a partir del análisis de los datos

Etapas de la aplicación de Big Data

Para iniciar la implementación de un proyecto de Big Data a diferencia de otro tipo de proyectos de negocio y tecnología, la organización debe considerar algunos requerimientos clave para el análisis de datos desde la perspectiva de Big Data.

Antes de describir las etapas para la aplicación de Big Data se debe establecer que la comunicación es un factor clave para este tipo de proyectos, pues, disminuye la pérdida de tiempo y recursos manteniendo una ambiente participativo de todos los niveles del negocio retroalimentando el avance del mismo.

La comunicación es esencial durante todo el proyecto, es así que, durante la fase de planificación se debe comunicar la visión que tiene el negocio respecto a la aplicación de Big Data y los objetivos estratégicos y su alineación con las metas del negocio, con la finalidad de establecer de forma adecuada los recursos necesarios para el desarrollo del proyecto.

Durante la fase de ejecución es necesario comunicar los avances del proyecto para garantizar el interés de la organización en el mismo.

Al finalizar el proyecto se debe comunicar a toda la organización de los éxitos logrados y/o los objetivos incumplidos (en los casos que se presenten situaciones de incumplimiento de objetivos por hipótesis mal elaboradas o limitaciones de recursos y tiempo), con la finalidad de alentar a la organización a dar continuidad a la implementación de otros proyectos de Big Data y fomentar en la organización una cultura orientada a la colaboración y la toma de decisiones basada en el análisis de datos.

Las etapas identificadas son la planificación, la ejecución y el seguimiento (Post-implementación) que se apoyaran en una estrategia de comunicación para mantener informada a la organización de lo que sucede en el proyecto.

Etapa 1: Planificación.

En la etapa de planificación se debe establecer los objetivos del proyecto, el alcance y la inversión en tiempo y recursos que realizará la organización para lograr un proyecto de calidad.

Como estrategia para desarrollar de forma adecuada la planificación se debe considerar lo siguiente:

- **Definir un caso de aplicación**
 - Establecer el alcance del proyecto
 - Identificar riesgos para el proyecto y el nivel de complejidad del análisis
 - Determinar los datos necesarios para el análisis

- Establecer niveles de seguridad para el acceso y manipulación de los datos seleccionados
- **Big Data como estrategia de negocio**
 - Contar con el auspicio de la alta gerencia o de una persona con poder de decisión y relación con la alta gerencia
 - Incluir en los objetivos del proyecto el mejoramiento o innovación de los procesos, estructura organizacional, infraestructura, servicios y productos es clave para alinear el proyecto a las estrategias del negocio.
 - Establecer indicadores clave de rendimiento para los objetivos del proyecto

Etapa 2: Ejecución

Como parte de la ejecución del proyecto de Big Data se debe contemplar algunos factores de los que dependerá el éxito o fracaso del mismo.

Entre los factores a considerar están los siguientes:

- **Equipo de Trabajo**
 - Contar con un equipo de analistas, especialistas y expertos en el análisis de datos, representación de resultados, habilidades en la selección de datos, conocimientos estadísticos y matemáticos es clave para el proyecto
- **Trabajo colaborativo**
 - Implantar una cultura de colaboración en toda la organización para desarrollar habilidades a través de las experiencias compartidas.
 - Promover la generación de conocimiento a través de la colaboración entre los miembros del equipo del proyecto de Big Data
- **Gobierno de los datos**
 - Desarrollar políticas para el acceso, modificación y manipulación de los datos.
 - Establecer mecanismos de procesamiento de los datos a través del uso de herramientas

- Establecer niveles de seguridad para los datos
- **Infraestructura**
 - Analizar las características de la infraestructura actual y los requerimientos para procesar los datos del caso en estudio.
 - Incluir propuestas de infraestructura que incluyan opciones como la adquisición de una infraestructura privada, la combinación de infraestructura de la organización con integración a una plataforma cloud para el procesamiento y almacenamiento, o definitivamente una infraestructura completamente en cloud dependiendo de la capacidad económica y las necesidades del negocio y los niveles de seguridad que se desee proporcionar a los datos.

Etapa 3: Post – implementación

Al finalizar la implementación de un proyecto de Big Data inicia una etapa de evaluación y seguimiento para evidenciar los aciertos y fracasos en la toma de decisiones al igual que nuevas oportunidad para la aplicación de Big Data.

El impacto del proyecto sobre los procesos de la organización, y el cumplimiento de los KPI es otro factor de análisis en esta etapa, pues determinarán el nivel de afectación que este tuvo sobre el negocio.

- **Identificación de nuevos proyectos de Big Data**
 - Identificar nuevas oportunidades para la aplicación de Big Data.
 - Promover la aplicación de Big Data para la toma de decisiones en otras áreas de negocio
- **Conocimiento generado**
 - Desarrollo de habilidades en todo el personal de la organización
 - Mejoramiento en la toma de decisiones
 - Identificación de nuevas ideas para el desarrollo de productos y servicios

6. Recomendaciones

En relación al análisis realizado respecto a la aplicación de Big Data en el Turismo del Ecuador y la generación de beneficios y oportunidades para el

sector, se recomienda el potenciamiento y desarrollo de proyectos basados en el análisis de datos para la toma de decisiones y el enfoque en el mejoramiento de los productos y servicios ofrecidos por el país.

Es importante que las instituciones a cargo del desarrollo del sector turístico del Ecuador dediquen su esfuerzo para identificar servicios y productos esperados por los turistas a través del análisis de los datos recopilados de las diversas fuentes relacionadas al sector.

Las instituciones públicas y privadas deben desarrollar una cultura basada en la colaboración para facilitar el acceso a la información, con la finalidad de compartir el conocimiento e identificar ideas y proyectos que impulsen el mejoramiento y construcción de servicios atractivos para los turistas.

Es necesario impulsar el turismo a través del uso de la tecnología a través de la web y las redes sociales, se debe desarrollar aplicaciones web y móviles que permitan a los turistas y visitantes acceder a los servicios ofertados por los establecimientos turísticos.

El análisis e integración de los datos recopilados por las entidades públicas y privadas, permitirá la identificación de patrones de comportamiento de los turistas al igual que predecir el impacto de la oferta turística sobre un segmento de mercado o sobre la población total de visitantes en un período de tiempo determinado.

Se debe establecer objetivos para el desarrollo de destinos turísticos inteligentes con la finalidad de generar experiencias satisfactorias en los turistas y así recopilar datos de su estancia y percepción de la calidad de los servicios recibidos.

El Gobierno debe establecer apoyar el relacionamiento con otros países para promover e impulsar el desarrollo del turismo en el Ecuador. El intercambio de datos y conocimiento promoverá el desarrollo y mejoramiento de las habilidades de los prestadores de servicios turísticos.

4. Carta de aprobación

Quito, 21 de Noviembre de 2014

**Señor Ingeniero
Germán Pancho
Director de la Maestría en Gerencia de Sistemas y Tecnologías de la Información
Universidad de las Américas**

Presente.-

De mi consideración:

Por medio de la presente, quiero comunicarle en mi calidad de Director de Tecnologías de la Información y Comunicaciones del Ministerio de Turismo, que, la señorita ingeniera Sonia Carolina Guama Morales ha presentado el informe denominado “Big Data: Aplicación y Oportunidades para el Turismo del Ecuador” como parte del proyecto de tesis denominado **“Estudio comparativo de métodos existentes para integrar la información estructurada y no estructurada de una industria enfocado en la generación de conocimiento, desde la perspectiva de una solución integral de Big Data”**, el cual ha sido revisado y recibo a conformidad con lo establecido en la carta de auspicio entregada para el desarrollo del proyecto de tesis mencionado.

Atentamente,

**Ing. Germán Maroto B.
Director de Tecnologías de la Información y Comunicaciones
Ministerio de Turismo**