

Universidad Nacional  
Sistema de Estudios de Posgrado  
Maestría en Administración de Tecnología de la Información

Propuesta de Proyecto de Aplicación Práctica  
de Tecnología de la Información

Big Data y Analytics, mercadeo y análisis de la información  
no estructurada, experiencias de otros países.

Daniel Umaña Domínguez  
Heredia, Costa Rica, 2019

# Declaración jurada de respeto a derechos de autor

22 de marzo del año 2019

Universidad Nacional  
Facultad de Ciencias Exactas y Naturales  
Escuela de Informática  
Posgrado en Gestión de la Tecnología de Información y Comunicación (ProGesTIC)

**FORMULARIO DE DEPÓSITO LEGAL, AUTORIZACIÓN DE USO DE DERECHOS PATRIMONIALES DE AUTOR E INCORPORACIÓN A REPOSITARIOS INSTITUCIONALES DE INFORMACIÓN DE ACCESO PÚBLICO**

La persona abajo firmante, en condición de estudiante de la maestría

Daniel Umaña Domínguez

y autor del Trabajo final de graduación titulado:

**Big Data y Analytics, mercadeo y análisis de la información no estructurada, experiencias de otros países.**

Para optar al grado académico de Máster en:

**Administración de tecnologías de información**

de conformidad con lo establecido en el documento de "Lineamientos generales para la realización del trabajo final de graduación" y demás normativa universitaria relacionada con estos trabajos de graduación, DECLARO BAJO FE DE JURAMENTO conociendo la responsabilidad civil, penal o administrativa en que podría incurrir al no decir la verdad, lo siguiente:

1. El documento, producto, obra audiovisual, software, resultado del trabajo final de graduación referido anteriormente es original, inédito y ha cumplido con todo el proceso de aprobación académico que confiere el grado académico postulado con esta obra.
2. El trabajo final de graduación referido anteriormente constituye una producción intelectual propia de la persona abajo firmante y a esta fecha no ha sido divulgado a terceros(as) de forma pública, por ningún medio de difusión impreso o digital.
3. Autorizo el depósito de un ejemplar en formato impreso y otro en formato digital (entregado en soporte de disco compacto), en la colección de trabajos finales de graduación del ProGesTIC de la Universidad Nacional, así como la realización de copias electrónicas adicionales para fines exclusivos de seguridad y conservación de la información.
4. En caso de que el trabajo final de graduación haya sido elaborado como obra en colaboración -bien se trate de obras en las que los autores(as) tienen el mismo grado de

participación o aquellas en las que existe una persona autora principal y una o varias personas autoras secundarias-, todos(as) ellos(as) han contribuido intelectualmente en la elaboración del documento y en este acto, libero de responsabilidad a las autoridades del posgrado y a los funcionarios que custodian la colección del ProGesTIC, en relación con el reconocimiento que se realiza respecto de los niveles de participación asignados por el propio autor del proyecto.

5. En caso de que el trabajo final de graduación haya sido elaborado como obras en colaboración (conforme a lo dispuesto en el punto 4), el autor abajo firmante designa a \_\_\_\_\_ como encargado(a) de recibir comunicaciones y representar con autoridad suficiente a los suscritos, en condición de agente autorizado(a) de los demás autores(as).

6. Reconozco que la colección de trabajos finales del ProGesTIC no emite criterios ni valoraciones académicas sobre lo planteado en el producto final del trabajo de graduación y autorizo a esta dependencia para que proceda a poner a disposición del público la obra en mención, a través de los espacios físicos o virtuales que se posea, así como a través del Repositorio Institucional; a partir del cual los usuarios de dichas plataformas puedan acceder al documento y hacer uso de este en el marco de los fines académicos, no lucrativos y de respeto a la integridad del contenido del mismo así como la mención del autor o poseedor de sus derechos.

7. Manifiesto que todos los datos de citas dentro de texto y sus respectivas referencias bibliográficas, así como las tablas y figuras (ilustraciones, fotografías, dibujos, mapas, esquemas u otros) tienen la fuente y el crédito debidamente identificados y se han respetado los derechos de autor.

8. Autorizo la licencia gratuita no exclusiva de los derechos patrimoniales de autor para reproducir, traducir, distribuir y poner a disposición pública en formato electrónico, el documento depositado, para fines académicos, no lucrativos y por plazo indefinido en favor de la Universidad Nacional, que incluye además los siguientes actos:

a. La publicación y reproducción íntegra de la obra o parte de esta, tanto por medios impresos como electrónicos, incluyendo Internet y cualquier otra tecnología conocida o por conocer.

b. La traducción a cualquier idioma o dialecto de la obra o parte de esta.

c. La adaptación de la obra a formatos de lectura, sonido, voz y cualquier otra representación o mecanismo técnico disponible, que posibilite su acceso para personas no videntes parcial o totalmente, o con alguna otra forma de capacidades especiales que les impida su acceso a la lectura convencional del proyecto.

c. La distribución y puesta a disposición de la obra al público, de tal forma que el público pueda tener acceso a ella desde el momento y lugar que cada quien elija, a través de los mecanismos físicos o electrónicos de que disponga.


d. Cualquier otra forma de utilización, proceso o sistema conocido o por conocerse que se relacione con las actividades y fines académicos a los cuales se vincula la

maestría, la colección de trabajos finales del ProGesTIC, la Escuela de Informática y la Universidad Nacional.

9. Reconozco que la colección de trabajos del ProGesTIC manifiesta actuar con diligencia para evitar la existencia en su sitio web de contenidos ilícitos y en caso de que tenga conocimiento efectivo de la existencia de infracciones a los derechos de propiedad intelectual, se reserva el derecho de proceder a bloquear el acceso durante el trámite del debido proceso para comprobar el incumplimiento y en caso de verificarse la falta, retirar definitivamente el acceso al proyecto depositado.

10. Acepto que la publicación y puesta a disposición del público del trabajo final de graduación, así como la presente autorización de uso de la obra, se regirá por la normativa institucional de la Universidad Nacional y la legislación de la República de Costa Rica. Adicionalmente, en caso de cualquier eventual diferencia de criterio o disputa futura, acepto que esta se dirimirá de acuerdo con los mecanismos de Resolución Alterna de Conflictos y la Jurisdicción Costarricense.

Autor Daniel Umaña Domínguez

Firma:  \_\_\_\_\_

Fecha de entrega: 22 de marzo del 2019

Correo: danielb727@gmail.com

# Índice

Declaración jurada de respeto a derechos de autor .....	ii
Índice .....	v
Dedicatoria .....	viii
Resumen ejecutivo .....	ix
Capítulo I: Introducción .....	1
1.1 Antecedentes .....	1
1.2 Problema .....	2
1.3 Justificación .....	3
Objetivos .....	3
Capítulo II: Marco teórico .....	4
2.1 Marco referencial .....	4
2.2 Marco conceptual .....	4
Definición de Big Data .....	5
Características principales de Big Data – Las 4 Vs .....	7
La quinta V – teorías .....	8
Creación de valor .....	9
Adopción de big data .....	11
Tipos de analytics .....	12
Big data analytics .....	14
Percepción de big data en niveles altos de gerencia .....	15
Big data y la toma de decisiones .....	17
Capacidades necesarias para el análisis de datos .....	18
NoSQL .....	19
Hardware y software para el análisis de datos .....	21
Mejores prácticas .....	23
2.3 Anotaciones finales .....	24
Capítulo III: Marco metodológico .....	26
3.1 Enfoque de la investigación .....	26
3.2 Tipo de investigación .....	27

3.3	Sujetos y fuentes de información.....	27
3.3.1	Sujetos .....	27
3.3.2	Fuentes de Información .....	27
3.4	Población .....	29
3.5	Definición de variables.....	30
3.7	Descripción de instrumentos utilizados .....	30
3.7.1	Entrevistas .....	30
3.7.2	Cuestionarios .....	30
3.7.3	Análisis de contenido .....	31
Capítulo IV: Diagnóstico de la situación actual .....		32
4.1	Situación actual del uso de big data y analytics en diferentes regiones.....	33
	Adopción de big data en Estados Unidos .....	33
	Adopción de big data en América Latina .....	35
	Adopción de big data en Europa .....	37
4.2	Tipo de organizaciones que pueden generar valor mediante el uso de big data y analytics.....	40
	Big data en el sector público .....	46
	Tecnologías emergentes.....	47
4.3	Influencia de big data y analytics en la toma de decisiones .....	48
4.4	Situación actual en Costa Rica .....	51
4.5	Anotaciones finales.....	54
Capítulo V: Solución del problema .....		55
5.1	Desarrollo de la solución .....	55
5.1.1	Manejo del negocio .....	57
5.1.2	Manejo de los datos .....	58
5.1.3	Análisis de datos .....	64
5.2	Procedimiento de implementación.....	65
5.2.1	Manejo del negocio .....	65
5.2.2	Manejo de los datos .....	67
5.2.3	Análisis de datos .....	75
5.3	Pruebas y resultados .....	76

5.3.1 Sobre la organización .....	77
5.3.2 Taller de capacitación .....	78
5.3.3 Alinear requerimientos .....	80
5.3.4 Evaluación de requerimientos de big data .....	82
5.3.5 Definición de estrategia de big data .....	84
5.3.7 Taller de implementación iterativa .....	87
5.3.8 Impacto de big data.....	88
5.4 Anotaciones finales.....	89
Capítulo VI: Análisis financiero.....	91
6.1 Inversión inicial .....	91
6.2 Recuperación de la inversión inicial .....	92
6.3 Proyección del flujo de efectivo .....	93
6.4 Análisis final de la inversión.....	93
Capítulo VII: Conclusiones y recomendaciones .....	95
7.1 Conclusiones .....	95
7.2 Recomendaciones .....	98
Capítulo VIII: Análisis retrospectivo.....	99
Capítulo VIII: Bibliografía.....	102
Glosario .....	109
Anexos .....	111
Anexo 1 – Cuestionario sobre la organización .....	111
Sección 1 – Información sobre la organización .....	111
Sección 2 – Generalidades de big data .....	112
Sección 3 – Big data en la organización .....	113
Sección 4 – Cultura organizacional.....	114
Sección 5 – Infraestructura .....	115
Sección 5 – Innovación con Big Data .....	116
Anexo 2 – Presentación: Conceptos y definiciones: Big Data & Analytics .....	117
Anexo 3 – Presentación: Conceptos de implementación iterativa.....	120
Anexo 4 – Cuestionario de valoración de aplicación del piloto.....	121
Índice de tablas.....	123

Índice de figuras .....	124
-------------------------	-----

## Dedicatoria

A Dios por darme vida, salud, sabiduría, las condiciones necesarias para desarrollarme como profesional y por hacer parte de mi vida a aquellas personas que han sido mi apoyo y compañía durante todo el periodo de estudio.

A mi esposa, Mar, pieza esencial en mi vida, ejemplo de superación y motivación para ser mejor cada día, por su apoyo incondicional, paciencia y muestras de amor durante todo el proceso de la maestría.

A mis papás, por toda su dedicación en mi formación, por inculcarme valores que hoy me hacen sentir orgulloso, por su apoyo incondicional y por siempre querer lo mejor para mí.

A mi hermano por siempre estar ahí, por ser guía cuando lo he necesitado, por sus palabras de apoyo y su amor incondicional.

A todos los profesores de la maestría, quienes con vocación compartieron su conocimiento y siempre tuvieron la mejor disposición de ayudar.

A los compañeros de la maestría, un grupo de excelentes profesionales, con quienes fue un gusto compartir clases.

## Resumen ejecutivo

En el campo de las tecnologías de información los cambios son constantes, algunos inclusive representan cambios que trascienden a otros sectores no relacionados con la tecnología y revolucionan la manera en la que se hacen los negocios. Este documento presenta la experiencia de big data y analytics en diversos países, demostrando ser un agente de cambio a nivel mundial.

La cantidad de datos que se genera diariamente crece de manera constante y se duplica cada dos años. Esta gran variedad y volumen de datos tiene un gran potencial para brindar una ventaja competitiva, sin embargo, esta enorme cantidad de datos no representa ningún valor si no son debidamente analizados.

Esta investigación hace una revisión de literatura para brindar una perspectiva de la adopción de big data en diferentes regiones del mundo, el impacto que tiene sobre la toma de decisiones a nivel de distintas organizaciones, así como los planes a futuro para mantener un liderazgo o mejorar la adopción de big data. Además, comprende un análisis de los diferentes sectores que pueden beneficiarse del uso de análisis de datos, tanto a nivel privado como público.

Se plantea como solución, una guía de mejores prácticas con el objetivo de apoyar a las organizaciones interesadas en aplicar big data como una herramienta en la toma de decisiones. La guía se divide en diferentes secciones que pueden ser aplicadas según los requerimientos de la organización. Adicionalmente se sugiere una implementación iterativa, que le permite a la organización una revisión según cambien los requerimientos del negocio.

Con el fin de poder evaluar la solución planteada, se puso en práctica un plan piloto que tomó varias de las secciones de la guía de mejores prácticas y fueron aplicadas en la empresa *Commercial Solutions*. Se logró un mejor entendimiento a nivel organizacional sobre big data y analytics, además se definieron los objetivos de big data basados en los objetivos de negocio.

Debido al resultado positivo de la implementación del piloto, se le recomendó a la empresa poner en práctica el resto de la guía de mejores prácticas enfocándose en

primera instancia en solo un objetivo de negocio, para posteriormente medir la efectividad de la estrategia planteada y hacer las mejoras necesarias en iteraciones futuras.

# Capítulo I: Introducción

## 1.1 Antecedentes

Un mundo sin almacenamiento de datos es algo prácticamente impensable, un lugar donde cada detalle de una persona u organización que pudiera ser documentado se pierda inmediatamente después de su uso. Las organizaciones perderían la habilidad de extraer valiosa información y conocimiento, no podrían hacer análisis detallados ni brindar nuevas oportunidades y ventajas (F. Bagci).

Los datos se han convertido en una parte esencial de nuestras vidas, y la habilidad de almacenarlos, extraerlos y utilizarlos se ha vuelto esencial para la operación de muchas organizaciones.

En el pasado cuando no existían sistemas interconectados, los datos se almacenaban y se consumían en un solo lugar. Con el avance de la tecnología y la llegada de Internet, la posibilidad de compartir y transformar datos en algo útil se ha convertido en una necesidad.

El concepto de Big Data se creó mucho antes de su definición y del gran auge que tiene en la actualidad, motivado en gran parte por el enorme avance en las tecnologías de información y la aparición de nuevos dispositivos inteligentes con la capacidad de generar enormes cantidades de datos cada día. Avances en la creación de sensores digitales con capacidad de comunicarse, aumenta aún más las posibilidades para capturar datos sin necesidad de interacción humana.

En un futuro no muy lejano es probable que casi cualquier objeto que se cree va a tener la capacidad de generar datos, inclusive nuestros cuerpos podrían ser capaces de generar datos con los dispositivos y sensores adecuados.

A esta gran cantidad de datos no estructurados no se les puede organizar de la misma manera que datos convencionales, y requieren una enorme capacidad de procesamiento en paralelo. Se necesitan tecnologías capaces de manejar y procesar hasta petabytes (**PB** equivale a  $10^{15}$  bytes), de datos a un costo que no sea prohibitivo.

La capacidad de generar valor de esos datos masivos es lo que les da a las organizaciones el potencial de transformar desde el mercado, hasta la manera en la que vivimos, trabajamos y pensamos.

## 1.2 Problema

La cantidad de datos que se genera a nivel mundial crece considerablemente cada día. Desde el constante uso de redes sociales donde cualquier individuo puede generar contenido hasta los diferentes componentes de la internet de las cosas hace que avance inclusive con más rapidez.

Diariamente aumenta no solo el volumen de datos, sino también la variedad. Datos que pueden contener información que lleve a una organización a obtener una ventaja competitiva sobre los demás. La habilidad para analizar esta gran cantidad de datos y tomar decisiones basadas en ese análisis puede ser el gran diferenciador.

¿Cómo impacta el uso de Big Data y Analytics en la toma de decisiones en las organizaciones de diferentes partes del mundo?

## 1.3 Justificación

### Objetivos

#### **1.3.1 General**

Desarrollar una guía de mejores prácticas para la implementación de Big Data y Analytics, mediante la revisión de literatura y análisis de experiencia en otros países, para brindar una herramienta que genere ventaja competitiva en las organizaciones.

#### **1.3.2 Específicos**

1. Registrar la situación actual del uso de big data y analytics en diferentes regiones del mundo por medio de una revisión de literatura y así determinar los aspectos relevantes.
2. Identificar por medio de investigación qué tipo de organizaciones pueden generar valor con el uso de big data y analytics para la toma de decisiones.
3. Determinar por medio de revisión de literatura la influencia que tiene el uso de Big Data y Analytics en la toma de decisiones de las organizaciones para identificar los factores claves a tomar en cuenta.
4. Analizar mediante una revisión de literatura el mercado de tecnologías de información en Costa Rica para tener un panorama claro de la situación.
5. Desarrollar una guía de mejores prácticas los datos obtenidos de la investigación para promover el uso de Big Data y Analytics para la toma de decisiones.
6. Aplicar mediante un plan piloto parte de la guía de mejores prácticas desarrolladas para verificar su efectividad.

## Capítulo II: Marco teórico

### 2.1 Marco referencial

Por ser una investigación de revisión de literatura, se considera que el marco referencial no aplica ya que éste formará parte del capítulo IV, descripción de la situación actual.

### 2.2 Marco conceptual

El volumen y la cantidad de datos que están siendo generados utilizando computadoras u otros “dispositivos inteligentes” se duplica cada dos años. Se estima que, en el año 2015, se generaron 8 Zettabytes de datos (**ZB** equivale a  $10^{21}$  bytes), de los cuales la mayoría son datos no estructurados como correos electrónicos, publicaciones en blog, redes sociales, fotografías y videos (V. Rajaraman, 2016).

Además de los datos generados por los humanos, también existen datos generados por comunicación entre máquinas (M2M - machine to machine) que contribuyen al enorme crecimiento. Por ejemplo, un avión moderno puede generar hasta medio terabyte (**TB** equivale a  $10^{12}$  bytes) de datos por vuelo (Airbus, 2013) sin la necesidad de interacción humana, todo por medio de sensores, que van desde los parámetros de vuelo básicos como la altitud y la velocidad, hasta la presión de aceite de cada motor o el flujo de combustible durante las diferentes etapas del vuelo.

No solo los aparatos tan complejos como los aviones cuentan con sensores que generan grandes cantidades de datos, algunos más comunes como las nuevas generaciones de automóviles y camiones están equipados con cada vez más sensores. Otro concepto que ha impactado la conectividad es la Internet de las cosas (IoT), permitiendo la integración del mundo físico con el digital, donde prácticamente cualquier objeto puede llegar a tener la capacidad de comunicarse y generar datos.

## Definición de Big Data

El término Big Data en el mundo de la computación fue acuñado por Roger Magoulas en el año 2005, refiriéndose a una gran variedad de grandes conjuntos de datos casi imposibles de manejar y procesar utilizando herramientas convencionales para el manejo de datos, tanto por su tamaño como por su complejidad (Halevi, 2012).

Posteriormente han aparecido otras definiciones con enfoques distintos que cubren las implicaciones que tiene en diferentes campos, estos son algunos ejemplos:

*El término Big Data ha sido aplicado recientemente a conjunto de datos que crecen tanto que se vuelve incómodo trabajar con ellos utilizando herramientas de manejo de bases de datos tradicionales. Estos conjuntos de datos tienen un tamaño que supera la capacidad de las herramientas de software y de los sistemas de almacenaje para capturar, manejar y procesar los datos dentro de un margen de tiempo razonable.*

*Big Data también se refiere a las bases de datos que son medidas en terabytes y de ahí en adelante y que son muy complejas y grandes como para ser usadas efectivamente en sistemas convencionales. (W.R. Kubick, 2012, pp 1)*

*Big Data es alto volumen, alta velocidad y/o alta variedad de activos de información que requieren procesamiento a un costo efectivo, formas innovadoras de procesamiento de información que permitan una visión mejorada de los datos, toma de decisiones y automatización de procesos. (Gartner)*

*Big Data, un concepto que significa muchas cosas para muchas personas, ha dejado de estar limitado al mundo de la tecnología. Hoy en día se trata de una prioridad empresarial dada su capacidad para influir profundamente en el comercio de una economía integrada a escala global. Además de proporcionar soluciones a antiguos retos empresariales, big data inspira nuevas formas de transformar procesos, empresas, sectores enteros e incluso la propia sociedad. (IBM Institute for Business Value, 2012)*

*Big data se refiere a grandes volúmenes de datos con altos niveles de complejidad y los métodos analíticos que serán necesarios sobre esos datos, que además requieren tecnologías y técnicas avanzadas para lograr obtener información significativa y perspectiva en tiempo real. (Ricardo Energy & Environment, 2017)*

En estas definiciones queda claro que los datos masivos necesitan un gran poder de procesamiento y además necesitan hacer uso de nuevas técnicas y tecnologías. Usualmente esto recibe el nombre de “analytics” o análisis de datos, cuyo principal objetivo es hacer utilizables datos que por sí solos no tienen mucho sentido. Big data está tan relacionado con el análisis de datos, que algunas personas inclusive consideran que big data es tan solo otra manera decir análisis de datos (McAfee y Brynjolfsson, 2012).

El cuadro a continuación es el resultado de un estudio realizado por la firma Transforming Data with Intelligence (TDWI) en el 2011, donde 92 organizaciones de diferentes regiones respondieron a la interrogante de cuál es el término que utilizan para analizar los grandes sets de datos.

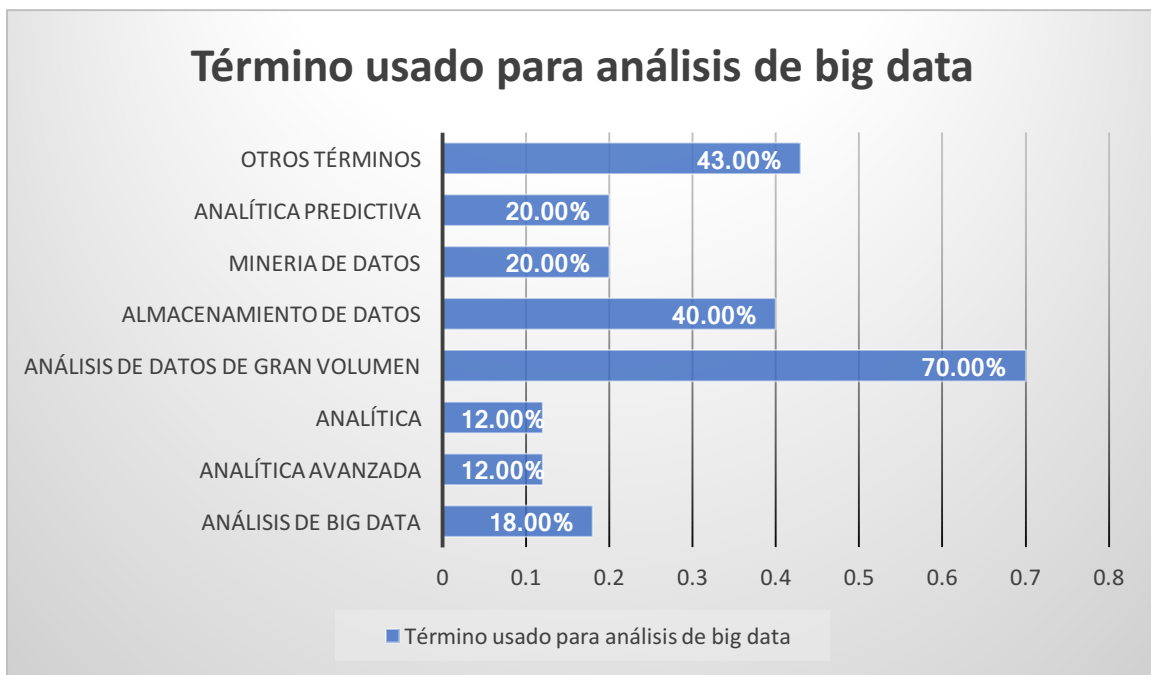


Figura 1 - Diferentes términos usados para describir lo que sería el análisis de big data. Fuente: TDWI, 2011

## Características principales de Big Data – Las 4 Vs

Gartner en el año 2001 tres características principales de Big Data. Con el pasar de los años estas características principales han evolucionado y frecuentemente se expresan en término de las 4 Vs (IBM, 2017), las cuales listan a continuación:

- ***Volumen – la dimensión de los datos***

Una de las características que mejor definen Big Data es su inmenso volumen. Una enorme cantidad de datos generados por máquinas, redes y la interacción humana con distintos sistemas. Generalmente se habla de Big Data cuando los conjuntos de datos tienen un tamaño que se expresa en terabytes o petabytes.

Se estima que para el año 2020 se crearán 40 Zettabytes (43 trillones de Gigabytes), un incremento de un 300% de lo que se creó en el año 2005 (IBM). Para dar una perspectiva, el gigante comercial *Walmart* genera más de 2.5 Petabytes de datos por hora, esto tomando solo con las transacciones de sus clientes (McAfee y Brynjolfsson, 2012).

- ***Variedad – las diferentes formas de datos***

Las fuentes de datos cada día son más diversas, la variedad se refiere a estas fuentes de datos y a los tipos de datos que son generados, tanto estructurados (bases de datos convencionales) como no estructurados (audios, fotografías, videos y mensajes de texto por mencionar algunos). Siendo los datos no estructurados los que representan un mayor desafío para almacenar y analizar.

- ***Velocidad – el análisis de la transmisión de datos***

Se refiere a la gran rapidez con la que se generan nuevos datos, y también a la velocidad con la que debe hacerse el análisis de los datos para que puedan ser utilizados.

La información en bases de datos tradicionales solía cambiar a un paso más lento, ahora la mayoría de los datos se generan en tiempo real. Debido a esta gran velocidad a la que se generan los datos, si no se usan a tiempo, se llegan a volver irrelevantes rápidamente. (V. Rajaraman, 2016). En el 2016 el número de conexiones a la red llegó a superar los 18.9 mil millones, casi 2.5 conexiones por cada persona en la tierra (IBM).

- ***Veracidad – la incertidumbre de los datos***

Los datos son relevantes cuando son confiables, la veracidad se refiere a que no deben existir sesgos, ruido o alguna anomalía en los datos. Para asegurarse de la veracidad, los datos deben ser consolidados, limpiados y lo más actuales posibles para que faciliten una toma de decisiones correcta.

Sin embargo, muchos de los datos generados son ruidosos, como por ejemplo: los datos obtenidos por sensores. Los datos pueden ser incorrectos. Es difícil tener absoluta seguridad de la veracidad de Big Data. (V. Rajaraman, 2016).

## **La quinta V – teorías**

Además de las cuatro características mencionadas anteriormente, algunos autores agregan otras características que también consideran relevantes (V. Rajaraman, 2016) (K. Normandeau, 2013).

- **Valor**

El valor está ligado a los beneficios que se pueden obtener de los conjuntos de datos una vez que están listos para ser utilizados. Dicho valor no es absoluto, esto quiere decir que el valor de los datos crudos puede ser diferente del valor de esos datos una vez que han sido analizados y procesados. Estas acciones pueden aumentar el valor de los datos al hacerlos más accesibles y fáciles de comprender.

Según Rajaraman, los datos por sí solos no tienen ningún valor, a menos de que hayan sido procesados para obtener información con la que se pueda tomar alguna acción.

- **Validez**

Al igual que la veracidad, la validez también puede ser un problema de Big Data. Los datos deben ser correctos y precisos para el uso que se les pretende dar. Claramente los datos válidos van a ser claves para la toma de decisiones (K. Normandeau, 2013).

- **Volatilidad**

Ligado a la velocidad, la volatilidad quiere decir que los datos únicamente son válidos por cierto periodo de tiempo y define por cuanto tiempo deberían ser almacenados. En un mundo donde los datos son generados en tiempo real, se debe determinar en qué punto los datos ya no son relevantes para su análisis (K. Normandeau, 2013).

## **Creación de valor**

El inmenso volumen de datos es un fenómeno global, sin embargo, no todas las personas ni las organizaciones lo ven como algo positivo. Existe gran preocupación por los problemas de privacidad que pueden generarse cuando cada acción que se realiza queda registrada digitalmente.

Por otra parte, existe evidencia que demuestra que big data puede tener un papel importante en la economía con el potencial de beneficiar no solo a organizaciones privadas, sino también al público en general y la economía de cada nación (MGI, 2011).

McKensey Global Institute (2011) define cinco formas en las que Big Data puede ofrecer un potencial de transformación para generar valor, lo que al mismo tiempo tendría un impacto en como las organizaciones son diseñadas, organizadas y manejadas.

- Generar transparencia
- Hacer que big data sea más accesible para las partes interesadas de manera oportuna puede generar gran valor.
- Habilitar la experimentación para descubrir necesidades, exponer la variabilidad y mejorar el rendimiento.
- Con la generación de más datos de manera digital, las organizaciones pueden recolectar y almacenar mayor cantidad de datos que posteriormente pueden ser analizados en experimentos controlados.
- Segmentar población para personalizar acciones
- Este tipo de práctica es común en mercadeo, entre mayor sea la recolección de datos mayor será la precisión de la segmentación.
- Reemplazar o dar soporte a las decisiones tomadas por humano con algoritmos automáticos.
- El análisis de datos puede mejorar el proceso de toma de decisiones drásticamente, minimiza los riesgos y logra obtener conclusiones que de otra manera no sería posible.
- Innovar nuevos modelos de negocio, productos y servicios

## Adopción de big data

La adopción de big data en organizaciones a nivel mundial está en auge, pasó de un 17% en el 2015 a un 53% en el 2017, con compañías de telecomunicaciones y servicios financieros encabezando la lista (Dresner Advisory Services, 2017).

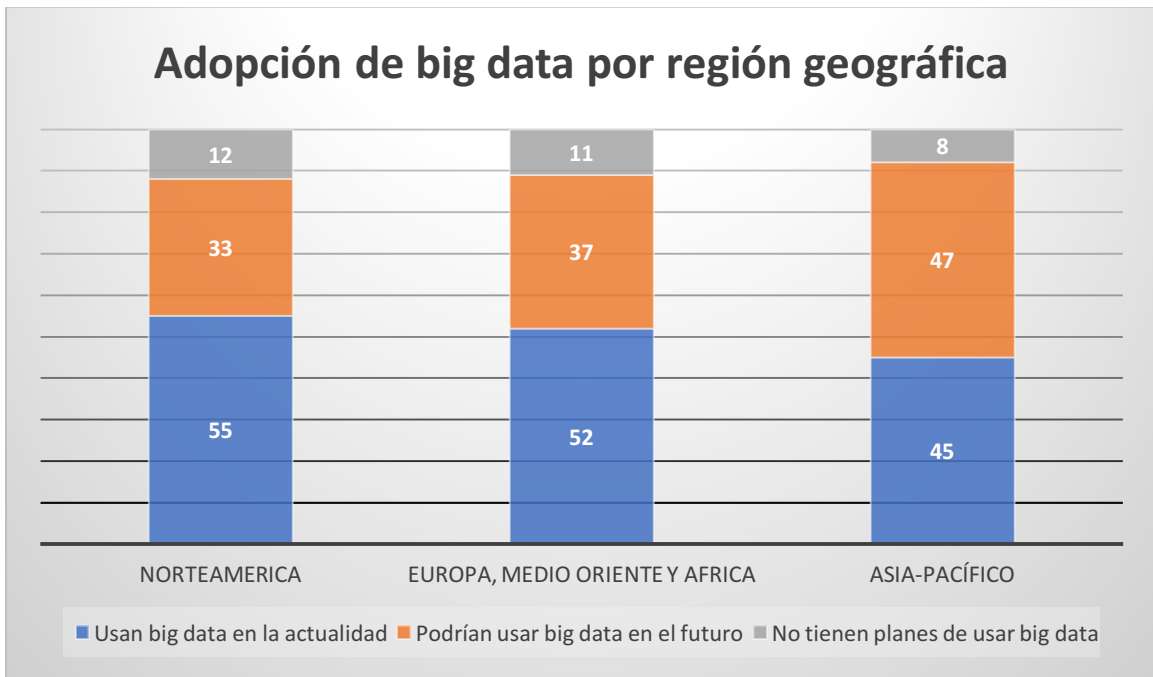


Figura 2 - Porcentaje de adopción en organizaciones según región geográfica. Fuente: Dresner Advisory Services, 2017

En el gráfico se puede apreciar Norteamérica (55%) lidera la adopción de big data, seguido de cerca por la región que comprende Europa, Medio Oriente y África (53%) en adopción actual. Asia-Pacífico por su parte son los que tienen más alta probabilidad de usar big data en el futuro, con un menor porcentaje de organizaciones que no piensan hacer uso de big data en el futuro.

## ***El caso particular de Europa***

Según el informe del año 2016 del Parlamento Europeo, Big Data Analytics tiene la capacidad de identificar ineficiencias en una gran variedad de sectores, que pueden dar lugar a la innovación de nuevos procesos y productos, con una mayor competitividad y crecimiento económico. Los estudios realizados por el Parlamento Europeo concluyeron que el uso de big data analytics incrementa la productividad entre un 5% a un 10% y que el producto interno bruto de la región podría aumentar en un 1.9% en el periodo de tiempo comprendido entre el 2014 y el 2020.

Tanto es el compromiso de Europa con el Big Data, que el gobierno de El Reino Unido lo incluyó en la lista de las “ocho grandes tecnologías” que van a propulsar a la región hacia el crecimiento del futuro y los ayudará a mantenerse a la vanguardia en la competencia global (UK Intellectual Property Office Informatics Team, 2014)

## **Tipos de analytics**

El análisis de datos se refiere a la extracción de conocimiento del cual se puede tomar acciones y de las perspectivas de big data. Esto se logra mediante la formulación de hipótesis que normalmente están basadas en conjeturas formadas por experiencia y descubrimiento (algunas veces por casualidad) de correlaciones entre variables (Rajaraman, 2016 pp 701).

Con un mismo conjunto de datos se pueden obtener diferentes resultados según el enfoque que se les dé. Principalmente se definen cuatro tipos de enfoques diferentes que responden a preguntas en específico. A continuación, se listan los diferentes tipos de análisis de datos, ordenados de menos a más complejos.

- **Descriptivo - ¿Qué está pasando?**

Es el análisis de datos más común y suele ser el primero que se realiza sobre un set de datos. Se utiliza para dar una perspectiva de lo que está pasando de una manera fácil de entender, tomando los datos “crudos” y convirtiéndolos en algo que pueda ser explicado, usualmente basado en métricas.

Muy útil para describir y analizar acontecimientos pasados para conocer cómo podrían tener un impacto en el futuro cercano, sin embargo, no da una razón del por qué está pasando (A. Fernando, 2018). Otra manera de verlo, es que ayudan a las organizaciones a aprender.

Los resultados de un censo de población son un claro ejemplo de este tipo de análisis de datos, donde la población se clasifica por sexo, grupos de edad, educación, ingresos y densidad de población entre otros parámetros (Rajaraman, 2016).

- **Descubrimiento o diagnóstico - ¿Por qué está pasando?**

Es un análisis más profundo, datos históricos pueden ser utilizados en conjunto a otros datos para crear conexiones que antes no se conocían y determinar por qué está ocurriendo algo. Gracias a este tipo de análisis se pueden descubrir dependencias e identificar patrones, en especial cuando se utilizan conjunto de datos de distintas fuentes (A. Bekker, 2017) (V. Rajaraman, 2016)

Se utiliza ampliamente para descubrir patrones de comportamiento de clientes en especial para el mercadeo.

- **Predictivo - ¿Qué es lo que podría pasar?**

Este análisis se utiliza para pronosticar el futuro, se extrapola información con datos existentes para predecir qué se espera que ocurra en el futuro. Las herramientas

utilizadas para la extrapolación son análisis de series temporales que utilizan métodos estadísticos, redes neurales y algoritmos de aprendizaje automático. (V. Rajaraman, 2016)

- **Prescriptivo - ¿Qué necesito hacer?**

La analítica prescriptiva utiliza el conocimiento resultante del análisis de lo que ha pasado, por qué ha pasado y qué podría pasar para determinar cuál sería la mejor acción a tomar a continuación. Por decirlo de otra manera, el análisis dice lo que se debe hacer para alcanzar una meta.

Las aerolíneas suelen a usar mucho este tipo de análisis para establecer el precio de los asientos basados en un histórico de datos de patrones de viajes, con puntos de origen y destinos populares, eventos importantes y días festivos entre otros, para maximizar sus ganancias. (V. Rajaraman, 2016)

## **Big data analytics**

Big data analytics se puede definir como el uso técnicas avanzadas de análisis para trabajar sobre grandes conjuntos de datos. Estos por sí solos no representan conocimiento hasta que se acomodan de la manera adecuada para que tenga sentido.

Entre las herramientas utilizadas para el análisis de big data, están minería de datos, analíticas avanzadas, visualización de datos y analíticas en base de datos.

## ***Estado de la adopción de big data analytics***

En mayo del 2011, la firma TDWI realizó un estudio donde participaron alrededor de 325 organizaciones, de las cuales un 56% están ubicadas en Estados Unidos, un 17% en Europa, un 3% en Centroamérica y Sudamérica y el restante 24% distribuido en el resto del mundo.

Se logró determinar que un 74% de las organizaciones para el momento del estudio ya habían adoptado algún tipo de analítica en su estrategia, aunque un 40% del total de estas organizaciones no lo aplicaba sobre big data, sino más bien en bases de datos tradicionales. El restante 34% si aplicaban analíticas sobre grandes conjuntos de datos.

### **Percepción de big data en niveles altos de gerencia**

*The Economist Intelligence Unit (EIU)* de La revista *The Economist* realizó estudio en el 2014 con directivos de alto nivel en diversas organizaciones a nivel global. El estudio comprendía una encuesta en la que participaron 18 directivos de diferentes industrias distribuidas en diferentes regiones:

- Asia – 34%
- América del norte – 27%
- Europa oriental – 26%
- Resto del mundo – 13%

Entre los resultados del estudio, destacan los siguientes:

- Actitud positiva hacia Big Data

La mayoría de los directivos consideran que Big Data va a jugar un papel importante en el futuro de la organización. Un 48% considera que es una herramienta muy útil, 23% piensa que va a revolucionar la manera en la que los negocios son manejados y un 21% creen que es esencial para lidiar con el gran volumen de datos que se

generan a gran velocidad. Tan solo un 3% consideran que el Big Data es solo una moda pasajera.

- Motivos que frenan la implementación

La falta de entendimiento de cómo aplicar el big data es el principal motivo de las organizaciones para no implementarlo. Seguido muy de cerca por la falta de recursos financieros orientados hacia nuevas tecnologías.

Otro motivo que vale resaltar, es la falta de un acuerdo sobre el valor real del valor de big data, respaldado por un 25% de los ejecutivos que formaron parte del estudio, quienes coinciden que también afecta la toma de decisiones.

- Optimización de uso

Se necesitan destrezas técnicas especializadas para optimizar el uso de big data, con puestos como científicos de datos que pueden brindar el soporte necesario para apoyar la toma de decisiones.

A nivel de directores individualmente, también se ha identificado una brecha en las capacidades cognitivas que las organizaciones poseen para hacerle frente a big data. Capacidades que son cruciales para evitar sesgos cognitivos en los que big data puede ser de gran utilidad. En lo que se refiere a las juntas de directivos, la cohesión de la junta también es trastornada por big data, pues los directores tienen dificultades para tomar decisiones al lidiar con los datos cambiantes en un periodo de tiempo tan corto. (Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L. y Canhoto, A., 2018).

## **Big data y la toma de decisiones**

Las organizaciones tienen la costumbre de analizar datos internos, ya sea inventario, envíos, cantidad de clientes o un sinnúmero de variables según sea el tipo de negocio. Big Data representa un cambio fundamental en la toma de decisiones de negocios, abriendo la posibilidad de tener acceso a datos externos que pueden dar una nueva perspectiva del mercado, los clientes y el ambiente.

La aparición de big data ha significado una nueva fuente de conocimiento en expansión constante, lo que ha estimulado a que los encargados de la toma de decisiones en las organizaciones actualicen sus competencias para abordar proactivamente los cambios ambientales y ser capaces de tomar decisiones más rápidas (S. Fosso Wamba, S. Akter, A. Edwards, G. Chopin y D. Gnanzou, 2015).

Según Capgemini (2017) Big Data y la perspectiva ampliada que brinda va a generar las más grandes oportunidades de diferenciación en los próximos cinco a diez años. Un estudio realizado por esta firma encontró evidencia de que las organizaciones ya están viendo los beneficios que Big Data puede ofrecer, donde en promedio se ha observado un 26% de mejora en rendimiento en un periodo de tres años y se espera que llegue a un 41% en los próximos tres.

En otro estudio realizado por McAfee y Brynjolfsson (2012) se tenía el objetivo de demostrar el valor que obtienen las compañías al hacer uso de los datos para tomar decisiones. Como parte del estudio se tomaron en cuenta ejecutivos de 330 compañías norteamericanas, junto con datos anuales sobre el rendimiento de las mismas compañías obtenidos de fuentes independientes.

McAfee y Brynjolfsson llegaron a la conclusión de que no todas las compañías que usan datos realmente los toman en cuenta para la toma de decisiones. Sin embargo, descubrieron que las compañías que se caracterizan por hacer más uso de los datos tenían un mejor rendimiento en sus objetivos financieros y operacionales. Hacer uso de

los datos debidamente analizados para la toma de decisiones, resulta en un aumento de productividad de un 5% en promedio, y un aumento de rentabilidad de un 6% en comparación a sus competidores.

También existe evidencia que big data está cambiando el proceso de decisión estratégica, generando cierto grado de ansiedad sobre el impacto que esto pueda tener en la organización, motivado en gran parte por la falta de preparación a niveles de gerencia (Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L. y Canhoto, A., 2018).

### **Capacidades necesarias para el análisis de datos**

El análisis de big data no depende únicamente de tecnología y algoritmos. Muchas compañías en la actualidad pueden costear computadores con capacidad de almacenamiento y procesamiento que hasta hace poco tiempo hubieran estado solo al alcance de las más grandes. Estas computadoras junto con científicos de datos capacitados pueden correr análisis complejos con el potencial de influir sobre la eficiencia de la compañía (The Economist, 2012).

Sobre este último término, IBM define la ciencia de datos como el proceso de descubrir información oculta a partir de cantidades masivas de datos estructurados y no estructurados, utilizando métodos tales como estadísticas, aprendizaje automático, extracción de datos y análisis predictivo. Esta área multidisciplinaria está cambiando la manera en que las organizaciones resuelven los problemas y obtienen una ventaja competitiva.

Los científicos de datos son los que precisamente deben saber cómo hacer uso de herramientas y al mismo tiempo tener habilidades para comunicar claramente los resultados mediante herramientas de visualización, con el fin de que los encargados de la toma de decisiones de la organización puedan hacerlo de manera informada basada en el análisis de datos.

## NoSQL

Las bases de datos tradicionales han sido utilizadas principalmente para almacenar información relevante a los objetivos de la organización. A finales de los años setentas se comenzaron a organizar los datos en formato de tablas, donde cada fila contiene una llave y una serie de atributos. A esto se le conoce como sistema de gestión de bases de datos relacionales (RDBMS).

En 1986 se estandarizó un lenguaje de consultas SQL (lenguaje de consulta estructurado) y se utilizó ampliamente para recuperar información relevante de bases de datos relacionales.

Con los avances tecnológicos cada día más accesibles, los tipos de datos que se generan han ido evolucionando y pasaron de ser solo números y letras a una gran variedad de documentos, imágenes, videos y audios entre otros. Este tipo de archivos tienen la característica de ser muy dinámicos, por lo que la recuperación de datos debe hacerse con una aproximación distinta (V. Rajaraman, 2016).

Además de la gran variedad de tipos documentos que se mencionaron anteriormente, también existen industrias específicas que necesitan hacer uso de bases de datos no relacionales para manejar grandes volúmenes de datos.

Un ejemplo es la industria aeroespacial, donde fabricantes de aeronaves como Airbus, han apostado por usar nuevas tecnologías para el manejo de grandes cantidades de datos. Al existir una mayor demanda de nuevos aviones y un incremento en la producción, también ha habido un aumento de la cantidad de vuelos de pruebas. Como si fuera poco, la tecnología en los aviones cada vez es más compleja, lo que requiere cada vez de más datos.

En cada vuelo de prueba durante la fase de desarrollo de nuevos modelos de aviones, se recopilan cientos de miles de parámetros de los diferentes sensores que miden el desempeño de la aeronave, lo que se traduce en datos cada vez más grandes y variados. Oracle (2015).

Según Araujo, A. (2016), las bases de datos NoSQL aún no cuentan con una definición formal, sin embargo, las define como: *“Las bases de datos NoSQL, también conocidas como “No sólo SQL”, nos referimos a una amplia clase de sistemas de gestión de datos (mecanismos para el almacenamiento y recuperación de datos) que difieren, en aspectos importantes, del modelo clásico de relaciones entre entidades (o tablas) existente en los sistemas de gestión bases de datos relacionales, siendo el más destacado el que no usan SQL como lenguaje principal de consulta.”*.

	SQL	NoSQL
<b>Almacenamiento de datos</b>	Almacenados en un modelo relacional con filas y columnas. El esquema define las tablas, las filas, las columnas, los índices, las relaciones entre las diferentes tablas y otros elementos de la base de datos.	Pueden contener una gran variedad de modelos de datos, incluyendo documentos, gráficos, pares (clave y valor), en memoria y búsqueda.
<b>Esquemas y flexibilidad</b>	Cada entrada cumple con un esquema anteriormente definido, lo que quiere decir que las columnas deben ser decididas y bloqueadas antes de que se ingrese la información. Esto puede ser modificado, pero debe alterarse toda la base de datos y debe desconectarse.	Los esquemas son dinámicos, la información puede ser agregada en vivo y las filas también son flexibles en lo que pueden contener.
<b>Escalabilidad</b>	La escalabilidad es vertical, para almacenar más datos se necesita un servidor más grande lo que puede llegar a ser muy costoso.	La escalabilidad es horizontal, lo que quiere decir que se puede hacer en varios servidores. Muchas tecnologías NoSQL distribuyen los datos en varios servidores de manera automática.
<b>Rendimiento</b>	El rendimiento depende del subsistema de disco. Para mejorar el rendimiento se requiere optimizar consultas, índices y la estructura de las tablas.	Por lo general depende del tamaño del clúster de hardware.

Tabla 1 - Tabla comparativa entre bases de datos SQL y NoSQL. Fuente: McNulty, E (2014) & Amazon AWS

## Hardware y software para el análisis de datos

Analizar big data tiene varios retos que no se limitan al gran tamaño de los datos, sino que también se debe considerar la variedad de los datos y la creciente necesidad de analizar esos datos en tiempo real para que no se vuelvan obsoletos. Se vuelve esencial analizar los datos para reconocer patrones que nos permitan tomar acciones significativas.

El avance exponencial de la generación de datos no está solo, pues la capacidad de procesamiento de las computadoras también avanza rápidamente, duplicando su capacidad de procesamiento cada dos años y su capacidad de almacenamiento cada 15 meses, ambas sin un aumento significativo en el costo (V. Rajaraman, 2016).

Aquí es donde tecnologías como el almacenamiento distribuido y el procesamiento paralelo toman importancia. En la actualidad es posible utilizar grupos de computadoras con miles de procesadores de fácil acceso, tal y como se encuentran en cualquier computadora de escritorio, para realizar análisis de datos, tan grandes y variados como para considerarse big data datos (McAfee y Brynjolfsson, 2012) (V. Rajaraman, 2016).

- **Hadoop**

Hadoop es un framework de código abierto basado en Java que permite el almacenamiento, procesamiento y análisis distribuido de grandes cantidades de datos, donde se hace uso de modelos de programación sobre un grupo de máquinas. Entre las características más relevantes, están el procesamiento en paralelo y la distribución de archivos en nodos.

Al diseñar este framework se tomó en cuenta diversos factores, uno de los más relevantes es que está diseñado para ser usado en *clusters* muy grandes de procesadores de fácil acceso. Al hacer uso de procesadores y discos de almacenaje de

relativamente bajo costo, la probabilidad de que uno o más componentes falle es alta, por lo que se contempló que el framework sea tolerante ante la falla de estos componentes (V. Rajaraman, 2016).

- ***Hadoop Distributed File System (HDFS)***

Está diseñado para almacenar grandes archivos en los cuales la información se escribe una sola vez, pero es leída muchas veces. Los archivos son divididos y almacenados de manera distribuida, tres conjuntos de bloques del mismo archivo son almacenados en tres computadoras distintas. HDFS se utiliza para almacenar trabajos tanto de entrada como salida de MapReduce y datos intermedios en caso de que sea producido alguno en una tarea de mapeo. (V. Rajaraman, 2016).

- ***MapReduce***

MapReduce es un algoritmo que fue diseñado por el equipo de desarrolladores de Google con el objetivo de solucionar los retos presentados por la variedad de datos que debían almacenar y procesar. Hadoop tiene como base el framework de MapReduce.

El término MapReduce es una palabra compuesta y se refiere precisamente a las dos tareas principales que ejecuta (J. Dean, S. Ghemawat, 2014):

- **Map (mapeo):** es una función escrita por los usuarios, toma un par de entrada y produce un conjunto de pares (clave, valor) intermedios. La librería MapReduce agrupa todos los pares intermedios asociados con la misma llave intermedia y pasa los valores a la función Reduce.
- **Reduce:** esta función también es escrita por el usuario, acepta una llave intermedia y un conjunto de valores en formato de pares (clave, valor) para esa llave. Una vez procesados se agrupan y se produce el resultado final.

## Mejores prácticas

El término mejores prácticas puede tener varias connotaciones según el contexto, de manera general se entiende que ciertas técnicas o protocolos son más apropiados y efectivos que otros, siendo la mejora de la calidad el objetivo principal. Sin embargo, surge la pregunta, ¿Qué constituya una “mejor práctica”?

Según Cato (2001), las mejores prácticas están relacionadas con otros dos términos, normas (*standards*) y lineamientos (*guidelines*) y cita las definiciones de la siguiente manera:

**Normas:** (1) una medida, un principio, modelo, etc., establecido para usar como una regla o base de comparación para medir o juzgar capacidad, cantidad, contenido, extensión, valor, cantidad, etc. (Desconocido, 1982).

(2) una codificación de tecnología o procedimiento desarrollado, probado, revisado y publicado por una agencia de gobierno o sociedad profesional; al cual deben apegarse sus miembros y subscriptores; generalmente deben ser seguidos apegándose al detalle (Hatheway, 1992).

**Lineamientos:** formalización presentada por una sociedad técnica o agencia gubernamental como base para la adhesión voluntaria para la aceptación del producto de trabajo; requiere una consideración y una evaluación adicional por parte del profesional para implementar el trabajo en un solo proyecto (Hatheway, 1992).

La principal diferencia entre los lineamientos y las normas, es que los lineamientos sugieren un nivel de calidad, que suele ser un mínimo y la aplicación a estos es voluntaria, recomendada pero no un requerimiento. Por otro lado, las normas tienen una definición más rígida y ofrecen menos espacio de interpretación. La aplicación de éstas es de carácter obligatorio.

Para Kenton (2018), las mejores prácticas son un conjunto de lineamientos o ideas que representan las acciones a seguir más prudentes y eficientes. Usualmente están definidas por una autoridad, ya sea un ente regulador, el gobierno o la regulación interna de una organización.

La Organización de la Naciones Unidas, lo define de la siguiente manera: “una experiencia positiva, probada y replicada en contextos diversos y que, por consiguiente, puede ser recomendada como modelo. Merece ser compartida para que el mayor número de personas pueda adaptarla y adoptarla.”

Las definiciones anteriores coinciden en que las mejores prácticas están compuestas por experiencias que han tenido resultados positivos y han demostrado ser eficaces y útiles en un contexto en concreto. Más puntualmente, algunas de sus características son:

- Presentan un resultado valioso para el usuario
- Son sencillas y fáciles de comprender
- Surge en respuesta a una necesidad de mejorar
- Son sostenibles en el tiempo
- Se presenta como innovación
- Son aplicables en un contexto específico
- Contemplan la retroalimentación posterior a su aplicación

### 2.3 Anotaciones finales

Es frecuente escuchar mencionar el término big data, sin embargo, es un término difícil de cuantificar, aun teniendo a manos las cifras de la cantidad de información que se genera alrededor del mundo cada día. Esa dificultad de cuantificar puede deberse a que los datos no son algo tangible, y se nos hace imposible imaginar lo que un *zettabyte* de información realmente es. Y precisamente debido a esto es que se necesita un esfuerzo en cambiar la manera de pensar en los altos mandos de las organizaciones y comenzar a visualizar el valor que big data puede traer, y ese valor que puede llegar a dar es lo que finalmente lo vuelve tangible.

La tecnología sigue avanzando de manera acelerada, aquellas organizaciones que no se adapten a lo que se ha llamado la cuarta revolución industrial se verán regazadas en un mercado cada vez más competitivo. En el capítulo IV se hará un análisis más profundo de la situación actual de big data y analytics en diversas regiones.

## Capítulo III: Marco metodológico

El diseño de una investigación está compuesto por una serie de características que ayudan al autor a definir cómo proceder. Entre éstas tenemos en primera instancia la definición del enfoque de la investigación, donde se define cómo se aproximará al objeto de estudio desde el punto de vista de paradigma de investigación científica, ya sea de manera cualitativa, cuantitativo o mixto. Al mismo tiempo tomando en cuenta cómo se pretende detallar los resultados, esto puede ser manera experimental, descriptiva o explicativa.

Se detalla también los métodos de recolección de datos y los instrumentos utilizados en este proceso, así como las fuentes de información de donde se obtienen éstos, divididas en fuentes primarias y secundarias.

Con estos puntos definidos, la investigación deberá seguir un proceso ordenado, de manera que permita abordar y solucionar el problema de investigación planteado, así como lograr alcanzar todos los objetivos planteados.

### 3.1 Enfoque de la investigación

El enfoque de esta investigación es de tipo mixto. Esto debido a que se tomó en cuenta tanto el punto de vista cuantitativo como cualitativo. Desde el punto de vista cuantitativo se busca medir la experiencia en otros países en la aplicación de big data y analíticas, enfocado principalmente en la influencia que éstos tienen en la toma de decisiones. En cuanto al punto de vista cualitativo, el hallazgo de nuevos datos ayuda a afinar las preguntas de investigación, permitiendo así volver a etapas previas de la investigación en caso de ser necesario.

La investigación tiene un alcance de tipo descriptivo. Según Sampieri (2010, pag 80) los estudios descriptivos buscan especificar las propiedades, las características y los perfiles de personas, grupos, procesos, objetos o cualquier otro fenómeno que se someta a un análisis.

## 3.2 Tipo de investigación

El diseño para la investigación es de tipo no experimental, se busca demostrar el estado actual y la experiencia de otros países relacionado al uso de big data y analytics. Sampieri (2010, pag 149) define que la investigación no experimental se basa en observar fenómenos tal como se dan en su contexto natural, para posteriormente analizarlos. Las observaciones son de situaciones ya existentes y no provocadas intencionalmente en la investigación.

## 3.3 Sujetos y fuentes de información

### 3.3.1 Sujetos

El sujeto se refiere a los objetos de estudio de la investigación. Esta investigación al ser principalmente una revisión de literatura de la experiencia en otros países relacionada al uso de big data y analytics, no tiene definido un sujeto de estudio.

### 3.3.2 Fuentes de Información

Para la búsqueda de fuentes primarias y secundarias de información se utilizará principalmente la plataforma de búsqueda de la Universidad Nacional con los siguientes criterios de búsqueda:

- Palabras clave:
- “big data”
- “big data analytics”

- “data analytics”
- Filtros:
- Año de publicación: Entre el año 2000 y el 2019
- Tipo de material: Libros, artículos, revistas, tesis y audiovisuales

También se hizo uso de la plataforma de **Google Scholar** para buscar elementos que no estén en la base de datos de **EBSCO** y así obtener un rango más amplio de datos a disposición.

Para la búsqueda de referencias en página de internet, se hará uso del buscador Google. En lo que respecta a datos generales sobre el tema, se utilizarán palabras clave como las siguientes:

- “big data”
- “big data and analytics”
- “big data decision making”

Con el objetivo de hacer referencia a la experiencia en otros países se utilizarán palabras clave que ayuden a concentrar los datos en regiones específicas, cómo, por ejemplo:

- “big data in america”
- “big data en américa latina”
- “big data in europe”
- “big data en España”
- “big data en México”

Se utilizarán nombres de otras regiones para determinar el nivel de aceptación y utilización de big data según las diferentes regiones del mundo. Las palabras claves principalmente se utilizarán en inglés ya que de esta forma se obtiene un mayor número de resultados, sin embargo, también se harán búsquedas en español para obtener datos en este idioma.

### **3.3.2.1 Fuentes primarias de la información**

Como fuentes primarias se comprende que los investigadores proporcionan información de primera mano, donde se presenta la información en su estado original sin ser resumida. En esta investigación se tomarán como fuentes primarias lo siguiente:

- Libros
- Artículos de publicaciones periódicas
- Tesis de grado
- Trabajos presentados en conferencias
- Páginas de internet de autores reconocidos o compañías de trayectoria en el campo
- Documentos oficiales de instituciones u organizaciones
- Elementos audiovisuales

### **3.3.2.2 Fuentes secundarias de información**

Como fuentes secundarias, o derivadas, se entiende que son aquellas cuya información es elaborada producto del análisis, extracción y reorganización de información basada en documentos primarios originales. En esta investigación se tomaron como fuentes secundarias lo siguiente:

- Fuentes de información citadas en el en texto
- Artículos científicos que interpretan otros trabajos
- Cuadros estadísticos
- Publicaciones en páginas de internet relacionadas con el tema

## **3.4 Población**

Tomando en cuenta lo explicado en el apartado 3.3.1 Sujetos, esta investigación al ser una revisión de literatura no cuenta con un sujeto, por lo tanto, tampoco cuenta con una población definida.

### 3.5 Definición de variables

Las variables identificadas en la investigación que ayudaron a conocer la experiencia de otros países y generar una guía de mejores prácticas son las siguientes:

- Situación actual del uso de big data y analytics a nivel internacional
- Tipos de organizaciones que pueden generar valor con el uso de big data y analytics
- Influencia de big data y analytics en la toma de decisiones
- Situación actual del uso de big data en Costa Rica
- Elementos importantes a tomar en cuenta a la hora de implementar big data y analytics en una organización

### 3.7 Descripción de instrumentos utilizados

#### 3.7.1 Entrevistas

Por el momento no se han definido entrevistas relacionadas con la investigación.

#### 3.7.2 Cuestionarios

Se definió un cuestionario que se encuentra detallado en la sección de anexos de este documento.

- Cuestionario sobre la organización y de conocimiento general de big data

### **3.7.3 Análisis de contenido**

Este proyecto se basa en investigación pura, haciendo énfasis en el análisis de documentos mediante una revisión de literatura, que pretende ser objetiva y sistemática en el estudio del contenido.

La revisión de literatura es una herramienta muy útil en el proceso de investigación ya que permite al autor contar con un marco de trabajo para basar su trabajo y además brinda contexto al lograr indicar relaciones con otros estudios previamente realizados. (Williamson y Johanson, 2018)

Además, una revisión de literatura exhaustiva permite identificar sesgos en investigaciones previas, lo que puede ayudar a justificar el enfoque que se le vaya a dar a la investigación.

## Capítulo IV: Diagnóstico de la situación actual

Existe un entendimiento generalizado que el Internet de las cosas (IoT) se va a volver algo omnipresente en la próxima década. Estos dispositivos tendrán la capacidad de generar enormes cantidades de datos, que de ser analizados correctamente pueden crear valor para diferentes organizaciones. El análisis de la información generada por el IoT y otras fuentes de big data requiere una enorme cantidad de recursos. (M. Marjani, 2017) Según el IDC (The International Data Corporation) para este año 2019, se espera que el mercado de big data alcance los 125 mil millones de dólares a nivel mundial.

El Big Data Value Association (BDVA) en su informe publicado en el 2017 sobre el valor estratégico de big data indica que, Estados Unidos de América y Europa, son las regiones donde más se generan iniciativas de big data y analytics. El valor del mercado de datos de Estados Unidos de América ocupó el primer lugar durante los años 2013 y 2014, seguido por Europa en el segundo lugar, región en la que los datos crecieron casi de igual manera.

Manteniendo la misma línea de esta afirmación, durante el evento de negocios digital de CeBIT en Europa en el año 2013, se presentó un informe que concluía que las organizaciones que utilizan big data y analytics tienen una probabilidad cinco veces mayor de tomar decisiones de una manera más expedita que sus competidores. De la misma manera se expuso que el 50% de las organizaciones afirman que el uso de big data les ayuda a cumplir con las demandas de los clientes y les facilita el crecimiento (I-Scoop, 2013).

Este capítulo tiene como objetivo brindar una perspectiva sobre la situación actual del uso de big data y analíticas en varias regiones del mundo, se analizará también el tipo de organizaciones que pueden sacar provecho de esta tendencia y además una mirada al mercado de tecnologías digitales en Costa Rica.

## 4.1 Situación actual del uso de big data y analytics en diferentes regiones

Big data como un recurso estratégico ha llamado la atención de muchos países desarrollados y también de organizaciones internacionales. Estos últimos inclusive ven en big data una herramienta útil para mejorar la situación de países en vías de desarrollo. La comunidad científica incluso ha comparado su valor estratégico con el del petróleo. (Hajirahimova, Makrufa &,Aybeniz, 2015).

En términos de adopción de big data y analytics, Estados Unidos se encuentra en el primer lugar, en este país un 68% de las compañías del sector privado están implementando de alguna manera big data, seguido por la Unión Europea con un 45%, y en tercer puesto la región de Asia-Pacífico con un 39% (I-Scoop, 2013).

### **Adopción de big data en Estados Unidos**

Los Estados Unidos de América ha visto nacer a varios gigantes de la industria que han adoptado el concepto de big data y analíticas desde su fundación. Un claro ejemplo son las redes sociales, entre las más fuertes Facebook, LinkedIn y Twitter. Todas estas compañías almacenan enormes cantidades de datos de los usuarios, siendo Facebook la que más datos genera, llegando a alcanzar el quinto puesto en la lista de compañías más valiosas en el 2016 según la revista Forbes. Para el 2018 llegó al puesto doce de la misma lista.

A nivel de gobierno, la administración Obama lanzó oficialmente en marzo del 2012 el programa “*Big Data Research and Development Initiative*” con una inversión que superó los 200 millones de dólares (Calil, 2012). Este programa involucró seis agencias gubernamentales:

- Departamento de defensa (DoD)
- Agencia de Proyectos de Investigación Avanzada de Defensa (DARPA)
- Departamento de Energía (DoE)
- Instituto Nacional de la Salud (NIH)

- Fundación Nacional de Ciencia (NSF)
- Servicio Geológico de los Estados Unidos (USGS)

El objetivo del programa es estudiar nuevas metodologías e infraestructuras para la investigación de big data con la finalidad de facilitar las herramientas y técnicas que permitan adquirir conocimiento y nuevas perspectivas en el tema. Específicamente busca desarrollar tecnologías centrales para recolectar, almacenar, manejar, analizar y compartir datos a gran escala, además utilizar estas tecnologías para acelerar el paso del descubrimiento científico y de ingeniería (Calil, 2012).

El universo digital (cantidad de bits creados, replicados y consumidos por año) en Estados Unidos ha presentado un crecimiento constante con una tasa de 25% anual, lo que significa que la cantidad de bytes creados se duplica cada tres años (IDC, 2013).

Este crecimiento constante está alimentado por varios factores como:

- Crecimiento continuo del uso de Internet, redes sociales, uso de teléfonos inteligentes y el IoT.
- Caída en el precio de dispositivos tecnológicos capaces de crear, capturar, manejar, proteger y almacenar información.
- Aumento en los datos generados por máquinas sin interacción humana directa.

A pesar de la gran cantidad de datos generados en Estados Unidos, no toda esa información es capturada y almacenada. La IDC (*International Data Corporation*), estima que en el 2012 únicamente el 18% de la información creada tenía la capacidad de generar valor luego de ser analizada. Estiman también que solo la mitad (9%) fue analizada. Para el año 2020 se estima que el porcentaje de datos que serán analizados alcanzará el 40% (IDC, 2013). Esta diferencia entre los datos generados y los analizados está ligado con la calidad de los datos, que se explicará con más detalle en este capítulo.

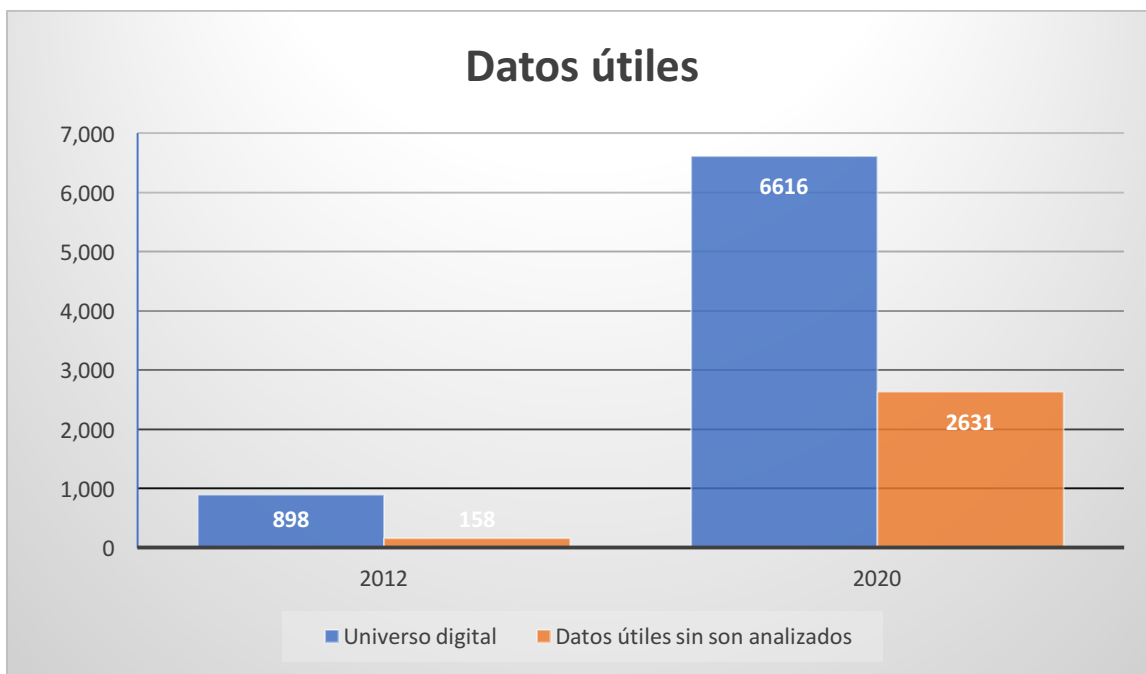


Figura 3 – Cantidad de datos útiles en comparación con los datos generados, expresados en Exabytes (EB).

Fuente IDC, 2013

### Adopción de big data en América Latina

Según IDC, compañía dedicada a la investigación de mercados, América Latina está pasando por un proceso de transformación digital, con alrededor de un 75% de las empresas de la región adoptando dicha transformación.

Se habla también de un crecimiento proyectado hacia el año 2020, con un incremento de un 129% del gasto en big data y analítica. Se estima que esto signifique una inversión cercana a los 14 mil millones de dólares. Sumado a esto, se calcula que, en este mismo periodo, la compra de dispositivos 4G crecerá un 57%, la inversión de sensores IoT un 120% y la infraestructura en la nube un 184% (IDC, 2017).

En el seminario “*Think big: Innovación de datos en América Latina y el Caribe*” organizado por la Comisión Económica para América Latina y el Caribe (CEPAL), con sede en Chile en marzo del 2017, se expuso que uno de los principales desafíos de la región latinoamericana es establecer un diálogo entre las oficinas gubernamentales de

estadística y los grandes actores del big data – en su mayoría del sector privado (CEPAL, 2017). En este capítulo también se hará una comparación acerca del uso de big data en el sector privado y el sector público.

Además, durante el seminario anteriormente mencionado se recalcó la importancia de big data en la región de la siguiente manera:

*“La analítica de grandes datos (big data) puede mejorar la toma de decisiones en áreas críticas del desarrollo, como la salud, el empleo, la productividad, la seguridad y la gestión de desastres naturales, por nombrar algunas, pero para ello se deben generar nuevas alianzas entre todos los actores involucrados e impulsar verdaderas revoluciones educativas y culturales.” (CEPAL, 2017)*

## Argentina

Argentina es uno de los países con mayor industria de software y servicios informáticos de América del sur, con un crecimiento de facturación sostenido en el periodo comprendido entre el 2003 y el 2013, de un 19.8% anual en industrias de este sector. Características que le dan al país una plataforma tecnológica y de conocimiento que facilitarían la adopción de big data.

Sin embargo, una encuesta realizada en este país en el 2014 entre las industrias del sector de servicios informáticos, muestra una demanda prácticamente inexistente y de muy baja complejidad en lo que respecta a big data. Principalmente por falta de conocimiento del tema, limitaciones de infraestructura y problemas institucionales. Además, existe falta de coordinación entre la producción académica y los desarrollos de empresas locales (Malvicino, Yoguel, 2015).

## Adopción de big data en Europa

Mucho antes de que el término “big data” apareciera, Europa ya tenía un pionero en el campo, Alan Turing, el matemático británico que logró descifrar el código enigma de Alemania Nazi, reconocido también como el padre de la computación moderna y los algoritmos. La Unión Europea en la actualidad está impulsándose hacia una economía basada en datos y espera entrenar a la nueva generación de científicos de datos (Blau, 2015).

Si bien es cierto Europa no es la región líder en big data, siendo liderada por Estados Unidos de América, es una de las que más esfuerzo denota en cuanto a organización para recuperar el terreno perdido. Además, como se explicará más adelante, existe un faltante de personal capacitado para hacer frente a las necesidades del mercado, situación que no afecta únicamente a la Unión Europea.

Según la firma Irlandesa William Fry (2017) existe la tendencia de grandes compañías internacionales en localizar sus centros de estrategia y decisión basados en big data, según la región (EEUU, Europa, Asia) a la que estén relacionados. Se habla de que alrededor de un 80% de organizaciones internacionales tienen algún tipo de operación en la Unión Europea.

Algo que está muy arraigado en la cultura europea es el gran valor que se da a los derechos de cada persona, en especial los relacionados a la privacidad. Debido a esto se sienten más seguros sabiendo que sus datos están siendo protegidos por leyes europeas (William Fry 2017).

Algunos de los obstáculos para la adopción de big data en la Unión Europea según Engels (2017) son:

- **Falta de seguridad de datos:** muchas organizaciones perciben que no van a ser capaces de salvaguardar los datos que generen. Tomando en cuenta que en Alemania entre los años 2014 y 2015 el 69% de las compañías fueron víctimas de algún tipo de ataque informático. Entre más datos maneja una compañía, más atractiva es para este tipo de ataques, que le cuesta a Alemania unos diez mil millones de euros al año.
  
- **Alto costo:** el costo no está asociado únicamente a la infraestructura necesaria para el procesamiento de datos. También está relacionado al alto costo de la innovación, que requiere una inversión financiera que tome en cuenta el riesgo de invertir en cambios que podrían resultar no ser los adecuados.
  
- **Falta de conocimiento (know-how):** afecta principalmente a las organizaciones pequeñas. El procesamiento y refinamiento de big data en “datos inteligentes” es una tarea grande que puede llegar a ser abrumador para compañías medianas y pequeñas.

Precisamente para contrarrestar el punto de falta de seguridad de datos y la falta de estandarización, la Unión Europea se encuentra trabajando en regulaciones sobre la privacidad y seguridad de datos (William Fry 2017). Entre estas regulaciones y esfuerzos están:

- Regulación General de Protección de Datos (GDPR). Pretende alcanzar dos objetivos principales
  - Fortalecer los derechos de protección de datos. Inclusive organizaciones internacionales deberán apegarse a esta regulación.
  - Estandarizar la regulación en todo Europa.
  
- **Escudo de Privacidad UE-EEUU (Unión Europea – Estados Unidos).** El principal problema de compartir datos con la Agencia de Seguridad Nacional (NSA) de Estados Unidos era la potencial pérdida de privacidad de los datos. El

escudo de privacidad especifica estrictamente los casos en los que los datos pueden ser transferidos a servidores en Estados Unidos.

## Reino Unido

La cantidad de empleos relacionados a big data en el Reino Unido era de 21,400 puestos en el 2013, para el 2017 ese número había crecido en un 222% alcanzando los 47,600 puestos y para el 2020 se estima que el número alcance los 56,000 puestos, un 117% de incremento con respecto al 2017 (E-skills UK, 2014).

El Reino Unido ha seguido activamente los esfuerzos propuestos por la Unión Europea y su gobierno considera que big data es uno de los principales agentes de cambio en la transformación digital.

La firma de consultoría Wishworks, realizó el estudio “*The State of Big Data in the UK 2017/2018*” con el objetivo de tener una mejor visión de que tan exitosos han sido los esfuerzos de big data en esta región. El estudio se realizó mediante una encuesta aplicada a 100 estrategias de big data, arquitectos y usuarios.

El estudio afirma que un 65% de las organizaciones ya están implementando big data de alguna manera, mientras que el restante 35% se divide en un 6% que no ha comenzado a investigar y un 29% que no han implementado big data, pero se encuentran activamente investigando para comenzar en un corto plazo.

Entre las organizaciones que ya han implementado un 30% asegura que los resultados han sido muy exitosos, mientras el restante 70% considera que ha sido algo exitosa, sin embargo, no cumpliendo sus expectativas. Algo que llama la atención es que ninguna implementación se consideró como no exitosa, lo que puede indicar que no existe una manera adecuada de cuantificar el resultado de un proyecto de big data y así evaluar su nivel de éxito.

## 4.2 Tipo de organizaciones que pueden generar valor mediante el uso de big data y analytics

Los avances tecnológicos están generando oportunidades comerciales por medio de big data que pueden ser aprovechadas por una amplia gama de industrias y sectores que continuamente buscan asegurar una ventaja competitiva. Estos avances pueden ser provechosos inclusive para organizaciones que no están directamente relacionadas con tecnología.

Según la firma William Fry (2017), los datos impulsan a todas las organizaciones, y los cambios tecnológicos recientes han generado oportunidades comerciales derivadas de big data, convirtiéndose esencial para organizaciones en todos los sectores. Ya no es relevante si una organización vende hardware, software, servicios digitales, productos farmacéuticos, servicios financieros o productos en general. Citando a John Frank, Vicepresidente de Asuntos Gubernamentales de la Unión Europea y el director de la oficina de Microsoft en Bruselas, Bélgica, *“todas las compañías son compañías tecnológicas o pronto lo serán”*.

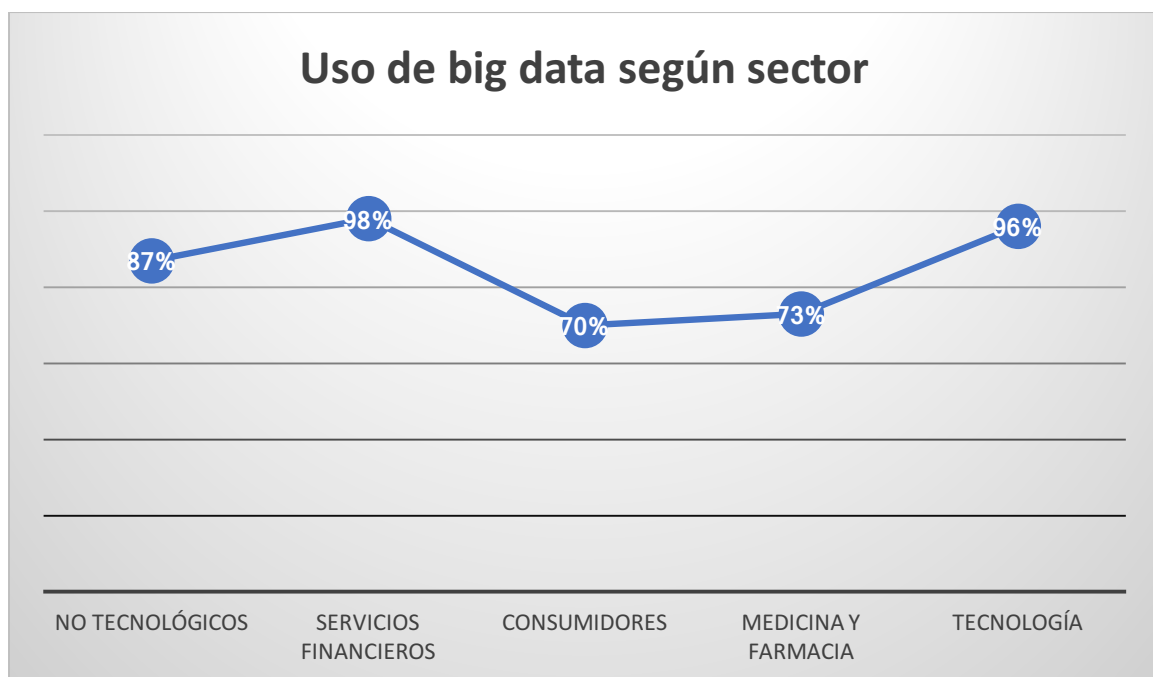


Figura 4 - Uso de big data en Europa según sector. Fuente: William Fry (2017)

En el sector privado, algunas compañías como Google, Facebook y LinkedIn se han hecho grandes utilizando de manera activa big data y analytics. Sin embargo, no todas las organizaciones basan su modelo de negocio en los datos, y pueden encontrarse en un nivel básico de adopción de esta tecnología, aún con dificultades para entender y definir su estrategia en torno a big data y analytics.

En el 2017, Halper en su informe “*Seven Industries Using Analytics for Business Value*”, identifica precisamente siete industrias que en la actualidad usan las analíticas para obtener valor de negocio. Por el tipo de servicios ofrecidos, algunas industrias como los servicios financieros, investigación científica y el mercado de seguros, tienden a adoptar las analíticas temprano. Sin embargo, otras industrias más conservadoras, como el cuidado de la salud, también están adoptando esta tecnología con el fin de mejorar los servicios ofrecidos y los diagnósticos.

Éstas son las siete industrias según Halper y la firma TDWI:

1. Servicios financieros y bancarios

Existen aplicaciones en este sector que son básicas, tales como la detección de fraude, record crediticio y manejo de riesgos. Sin embargo, los servicios financieros se han mantenido alejados de las analíticas de los clientes propiamente, y es en este punto que algunas compañías están obteniendo una ventaja de big data, ofreciendo con servicios como:

- Segmentación de mercado: ofrecen al cliente diferentes opciones según su estilo de vida, su lugar de residencia o trabajo e inclusive haciendo uso de datos geográficos y demográficos.
- Predicción de pérdida de clientes: hace uso de datos históricos de clientes que han dejado de ser clientes de la entidad bancaria para predecir, según el comportamiento, posibles bajas en los servicios bancarios.
- Venta de servicios adicionales

## 2. Salud

La industria de la salud se encuentra actualmente en una transición de pago por servicio ofrecido, a un modelo que genere valor basado en los resultados de los pacientes. Algunos de los casos en los que se utilizan analíticas en el sector médico y salud son:

- Comprender la reincidencia de los pacientes: se mide la reincidencia de padecimientos o admisiones hospitalarias, los datos utilizados pueden ser de diversos centros de salud en lugar de centrarse solo en uno.
- Sanidad poblacional: está más enfocado a servicios de salud del sector público, donde el objetivo es aumentar el bienestar general y reducir los costos.
- Optimización de la operación de los centros de salud y el uso de recursos, tanto de espacios físicos como salas de operación, como de recursos consumibles y medicamentos.

En este sector de servicios de salud, existe resistencia a compartir información, los datos de los pacientes y padecimientos debe ser confidencial, por lo que muchas veces este tipo de cambios está ligado a reformas legales en los distintos países.

## 3. Seguros

Las compañías aseguradoras han hecho uso de analíticas de datos desde sus inicios para estimar el riesgo de varios grupos, llámese finanzas, salud o transporte. El objetivo del análisis de riesgo es estimar que tan frecuentemente ocurren ciertos eventos y cuál sería el impacto si éstos llegaran a ocurrir.

Además de los usos comunes en el campo de los seguros, también se identificaron áreas donde pueden ser implementadas las analíticas:

- Analítica geoespacial para la evaluación de riesgos: con ayuda de datos geográficos se puede determinar con una mayor precisión el riesgo que existe en determinada región, como por ejemplo riesgo de inundaciones o huracanes.

- Uso de telemática para ajuste de precios: el avance en dispositivos capaces de recopilar y transmitir datos, como el IoT, permite medir a analizar datos de utilidad para aumentar o reducir el precio de una póliza. Por ejemplo, en un automóvil, se tomaría en cuenta que tan rápido se conduce, con que fuerza se aplica el frenado e inclusive la frecuencia con la que es utilizado.

#### 4. Ventas al detalle

La llegada de los avances tecnológicos ha tenido un gran impacto en modelo de tiendas al que tradicional. Si bien es cierto el uso de analíticas en el comercio no es algo nuevo, la falta de adopción de un nuevo tipo de analíticas más acorde al mercado actual ha llevado a grandes tiendas a la quiebra, especialmente en Norte América.

Las tiendas más exitosas en la actualidad están enfocadas en conocer los hábitos de sus clientes para ofrecerles los productos adecuados.

- Analíticas omnicanal: el objetivo es brindarle a los clientes lo que quieren, cuando quieren y de la manera que quieren. Los compradores quieren tener la posibilidad de interactuar con el objeto, pero también quieren tener la flexibilidad de poner adquirirlo en línea. Al usar diversos canales, el usuario solo ve una compañía, independientemente si está en la tienda física o si compra por Internet.
- Personalización: se considera crítico en la era digital de las ventas, uno de los mejores ejemplos es el caso de Amazon y las sugerencias de compras basadas en productos vistos anteriormente.
- Analíticas por geolocalización: se puede usar para decidir dónde colocar una tienda o inclusive para conocer el tráfico de personas en una tienda por medio de los teléfonos celulares.

## 5. Manufactura

La industria de manufactura ha sido históricamente lenta en sus cambios, sin embargo, en las últimas décadas esto ha venido cambiando significativamente. Existe una necesidad para innovar, ser eficientes y mantener la competitividad. Por medio de las analíticas se puede:

- Pronosticar y optimizar el inventario: mantener un inventario controlado, elimina el desperdicio y por lo tanto reduce costos. Se utiliza el concepto “*just-in-time*”, justo a tiempo, permitiendo que el inventario necesario esté disponible en la cantidad exacta, en el momento que se necesita y al menor costo posible.
- Sistemas de alertas de control de calidad: mejorar la calidad de la producción ayuda a reducir los costos de producción. Con datos exactos de problemas de calidad se puede trabajar en la optimización del proceso.
- Mantenimiento preventivo: otro ejemplo en el que el IoT es de gran utilidad. La captura de datos de producción en tiempo real puede registrar variaciones de temperatura en la maquinaria que estén fuera del rango normal y detectar vibraciones anormales por medio de sensores. El histórico de información puede ser analizado para encontrar patrones de fallas y determinar cuál equipo o piezas son más susceptibles a fallar.

## 6. Servicios públicos

El sector de servicios públicos (energía, agua y gas en algunos países) genera una enorme cantidad de datos y existe una necesidad de que sean analizados de una mejor manera. Los beneficios no son únicamente para las compañías de servicios públicos, sino también para los usuarios quienes tienen mejor visibilidad de su consumo. Algunos usos de la analítica de datos en esta industria son:

- Detección de fraude: puede ser utilizado principalmente en el servicio eléctrico, mediante la incorporación de medidores inteligentes (de nuevo, la importancia del IoT) que registran información de consumo de manera remota. Estos datos en tiempo

real junto a técnicas de análisis pueden detectar intentos de fraude de manera inmediata.

- Conciencia situacional: los servicios públicos se ven afectados por diversos factores externos, entre ellos el tiempo (lluvia, humedad, viento, polvo y temperatura). La recolección de datos mediante sensores puede ayudar a predecir problemas capaces de generar una interrupción del servicio.

## 7. Tecnología, Software e Internet

Sin lugar a dudas es uno de los sectores que mayor provecho puede sacar de las analíticas, además de ser una de las primeras industrias en adoptar analíticas más complejas. Aquí podemos hablar de inteligencia artificial, aprendizaje automático, procesamiento del lenguaje y otras tecnologías avanzadas.

Este sector presenta también uno de los índices más altos de rotación de personal debido a la competitividad entre empresas por contratar recursos capacitados.

Algunas aplicaciones en esta industria son:

- Análisis de retención de personal: si bien es cierto, esta es una tarea del área de recursos humanos, la implementación de este análisis en un sector tan competitivo como el de tecnología es de suma importancia para mantener la competitividad. Un algoritmo de aprendizaje automático puede llegar a determinar cuándo es más probable que un empleado decida dejar la compañía, usando datos como edad, género, tiempo que le toma en llegar al trabajo, salario y satisfacción con el puesto que desempeña.
- Predicciones de demanda de productos: datos históricos de comportamiento del mercado pueden ayudar a identificar tendencias y desarrollar productos que cumplan las necesidades de un nuevo nicho de mercado.

## Big data en el sector público

Si bien es cierto el desarrollo de aplicaciones de big data es liderado por el sector privado, el sector público también ha comenzado a hacer uso de big data para apoyar la toma de decisiones haciendo uso de datos en tiempo real. En la literatura revisada se expone que se han realizado propuestas para que los gobiernos utilicen big data para comprender las necesidades de sus habitantes. En la conferencia CEPAL 2017 citada anteriormente, se mencionó que la analítica de big data puede mejorar áreas críticas como el desarrollo, la salud, la seguridad y el manejo de desastres naturales.

Es válido hacer una comparación entre el sector privado (negocios) y el sector público (gobierno). Kim, Trimi & Chung (2014) sostienen que, aunque las misiones primarias de ambos sectores no entran en conflicto, si difieren y reflejan distintos objetivos y valores.

	Sector privado	Sector público
<b>Objetivos principales</b>	<p>Proveer productos y servicios para obtener beneficios.</p> <p>Desarrollar y mantener una ventaja competitiva.</p> <p>Satisfacer a los clientes e inversionistas generando valor.</p>	<p>Mantener la tranquilidad en la población.</p> <p>Alcanzar un desarrollo sostenible.</p> <p>Asegurar los derechos básicos de los ciudadanos.</p> <p>Promocionar el bienestar general y el desarrollo económico.</p>
<b>Toma de decisiones</b>	<p>El proceso toma más tiempo y a menudo involucra y la toma de decisiones involucra hacer consultas y consenso entre varias partes involucradas.</p> <p>Una serie de pasos definidos son requeridos para reducir el riesgo.</p>	<p>Necesitan tomar decisiones a corto plazo con un número limitado de involucrados en el proceso en un mercado competitivo.</p>
<b>Estructura organizacional</b>	Jerárquica	Gobernanza
<b>Recursos financieros</b>	Ganancias (Revenue)	Impuestos

*Tabla 2 - Diferencia entre sector público y privado. Fuente Kim et al. 2014*

Las organizaciones gubernamentales tienen que enfrentar sus propios retos, además de tener que lidiar con los problemas generales de la integración de big data proveniente de diferentes fuentes, formatos y costos. Uno de los principales desafíos es la recolección de datos, esto debido a que los mismos no solo provienen de diferentes canales (redes sociales, IoT e Internet en general), sino que también de diferentes fuentes (otros países, instituciones y agencias). De estos retos, el más complicado es compartir datos con otros países, especialmente por los diferentes idiomas y posibles problemas de privacidad (William Fry, 2017). Sin embargo, hay casos, como la prevención y lucha contra el crimen en lo que es de vital importancia compartir la información (Kim et al. 2014), existiendo una regulación que proteja la privacidad de los datos.

### **Tecnologías emergentes**

Existen además, cambios tecnológicos que permiten la aplicación de big data en campos que antes no se pensaba posible. Uno de estos campos es la agricultura, que si bien es cierto no es una novedad que se haya visto beneficiado con los avances científicos y tecnológicos, hoy en día tiene a su disposición sensores de todo tipo, que pueden medir desde cambios climáticos a gran escala, hasta minúsculas variaciones en la acidez y la temperatura del suelo. Los datos recolectados por cientos de miles de sensores están abriendo campo a lo que se conoce como agricultura inteligente.

El IoT es una de las tecnologías revolucionarias que obtiene todos los beneficios ofrecidos por el Internet. Este tipo de dispositivos tiene un sinnúmero de aplicaciones, como el mencionado anteriormente, también se puede citar otros ejemplos como sistemas para ciudades inteligentes, oficinas inteligentes, ventas inteligentes (*smart retail*), suministro de agua inteligente, transporte inteligente, cuidado de la salud inteligente y energía inteligente. (Ruivo P, Oliviera T, 2017).

### 4.3 Influencia de big data y analytics en la toma de decisiones

Big data no abarca únicamente datos, sino también los medios, tanto tecnológicos como metodológicos mediante los cuales los datos son almacenados, procesados y analizados (Van Puyvelde, 2017). Los datos por si solos no son más que una serie de símbolos y se convierten en información únicamente cuando son procesados para que cobren sentido.

Tanto las compañías como otras organizaciones pueden beneficiarse del análisis de big data, utilizándolo para obtener un mejor entendimiento de los datos y así tomar decisiones eficientes y bien informadas. Por ejemplo, Chen, Chiang & Storey (2012) sugieren que, las analíticas de negocios y otras tecnologías similares pueden ayudar a las organizaciones a tener un mejor entendimiento de la ventaja que los datos masivos y analíticas dirigidas a un sector en específico puede significar para su negocio y mercado en general.

LaValle, Lesser, Shockley, Hopkins & Kruschwitz (2010) descubrieron en un estudio que las organizaciones con mayor desempeño (*top-performing*) son aquellas que toman decisiones basadas en análisis rigurosos a una velocidad dos veces mayor que las organizaciones con el menor desempeño (*lower performing*). Además, las organizaciones que lideran el mercado usan la perspectiva analítica para guiar tanto estrategias futuras como la operación del día a día.

Este análisis de datos debe estar a cargo de “*data miners*” y “*data scientists*” quienes tienen la formación necesaria para extraer información valiosa haciendo uso de técnicas de minería de datos que ayuden a hacer predicciones, identificar tendencias recientes, encontrar información oculta y finalmente asistir en la toma de decisiones (M. Marjani, 2017).

Precisamente bajo esa misma línea de tipo de perfil profesional que se necesita para obtener valor de big data, el Consejo de Habilidades Sectoriales del Reino Unido para la Industria de Tecnología de la Información (E-skills UK) realizó un estudio en conjunto con

SAS, una de las compañías más importantes en software de analíticas, inteligencia de negocio y manejo de datos, denominado “Evaluación de la demanda de mano de obra y habilidades 2013–2020”, con el objetivo de determinar las necesidades de profesionales preparados con miras al futuro.

En el estudio, SAS define big data como el nuevo petróleo y que será el combustible que impulse la economía en las siguientes décadas. Al igual que el petróleo, los datos deben ser “refinados” para obtener ese potencial y ahí es donde se necesita el personal con las habilidades necesarias y las herramientas adecuadas para convertir el conocimiento en acciones decisivas. La demanda de este personal ya está sobrepasando la oferta existente en el mercado, ya que, según datos del estudio, la demanda en la Unión Europea crece en promedio un 212% (E-skills UK, 2014).

Los beneficios que se puedan obtener de los datos dependen en gran medida a la calidad de los mismos. El concepto de calidad puede variar según cada organización, pero en términos generales datos de calidad son aquellos aptos para el uso inmediato. La calidad de los datos surge como uno de los temas más importantes de la gestión de datos (Ruivo P, Oliviera T, 2017).

Los problemas en la calidad de los datos pueden generar un impacto considerable en el proceso de toma de decisiones de las organizaciones, esto debido a que la calidad del análisis está correlacionada completamente con la calidad de los datos que se utilizan para dicho análisis (M, Campanelli. 2007). Esto además puede traducirse en afectaciones en el rendimiento financiero de la organización, según Madan (2010), se estima que la mala calidad de los datos les cuesta a negocios estadounidenses alrededor de un 30% de sus ganancias, unos 700 mil millones de dólares anualmente. Gartner (2010) por su parte estima que algunas organizaciones, de manera individual, pierden hasta 100 millones de dólares anualmente debido a este problema.

Sharma, Mithas & Kankanhalli (2014) consideran que la teoría que sostiene que la analítica de datos lleva a generar valor, necesita ser analizada más a profundidad.

Especialmente porque los roles asociados al proceso de toma de decisiones deben ser comprendidos de una mejor manera antes de lograr asimilar cómo las organizaciones pueden generar valor de la analítica de datos.

Este razonamiento está basado en que habitualmente se ha enfocado en buscar maneras en las que la gerencia pueda tomar mejores decisiones usando big data, cuando el enfoque debería ser contar con mejores datos y herramientas para la toma de decisiones.

La literatura también refuerza que las organizaciones pueden generar valor de los datos mientras continúan funcionando de la misma manera que antes (Sharma et al. 2014). Por el contrario, la literatura nos demuestra que se necesita una transformación en el proceso de toma de decisiones para poder obtener mejor rendimiento con base en las inversiones realizadas en analíticas de datos.

Sharma et al. (2014) plantea tres preguntas que pueden dar un mejor panorama de la situación antes mencionada:

- ¿Cómo influye el uso de analítica de datos en el proceso de toma de decisiones en la organización?
- ¿Cómo influye el proceso de toma de decisiones en la organización sobre el uso de analítica de datos?
- ¿Cuáles son los efectos conjuntos del uso de analítica de datos y el proceso de toma de decisiones de la organización en el rendimiento organizacional?

#### 4.4 Situación actual en Costa Rica

El desarrollo de la industria de software en Costa Rica fue el resultado del espíritu empresarial de una generación de jóvenes de la década de los ochentas, época de crisis económica en este país (Vargas, 2012). La Cámara de Tecnologías de Información y Comunicación (CAMTIC), concuerda en que Costa Rica se ha venido desarrollando en el sector de las tecnologías digitales desde hace poco más de 3 décadas.

En el 2003 CAMTIC propuso la Estrategia Nacional de las Tecnologías de la Información y de la Comunicación: “Costa Rica: Verde e Inteligente” ® como marco conceptual y estratégico para promover una visión integral sobre las acciones necesarias para impulsar este sector.

Esta estrategia ha evolucionado desde su creación, ajustándose a los nuevos retos y al cambiante panorama nacional, actualmente se trabaja en la segunda versión, donde se fija la visión de la estrategia como:

*“Costa Rica será un proveedor líder de productos y servicios de Tecnologías Digitales innovadoras y de alto valor agregado a nivel nacional e internacional; un centro de convergencia para los negocios y el desarrollo humano en un ambiente ecológicamente amigable, democrático y seguro “.*

Como parte de la estrategia, se ha creado una subdivisión por categorías según el tipo de actividades que realizan las empresas del sector de tecnologías digitales en Costa Rica.

1. Desarrolladores de Software
2. Multimedia Digital
3. E-commerce
4. E-learning
5. Tecnología de la Información
6. Telecomunicaciones y Redes
7. Comercialización de Tecnologías

8. Servicios Habilitados por las tecnologías digitales
9. Manufactura de Componentes Digitales

En el resumen ejecutivo de Costa Rica Verde e Inteligente, se menciona que, según cifras del 2010, existían en Costa Rica unas 805 empresas de tecnologías digitales. Creciendo entre el 2003 y el 2008 en un 101%, siendo el subsector de servicios de software el que mayor crecimiento presentó. Se realizó la consulta en CAMTIC y en PROMOCER, sin embargo, aún no cuenta con datos publicados más actualizados.

Más de la mitad de las empresas de tecnologías digitales en Costa Rica, se identifican con el subsector de tecnologías de la información (55%) o con el desarrollo de software (52%). Cabe hacer la anotación que solo una tercera parte de las empresas se identifican como parte de un solo subsector, mientras que las restantes dos terceras partes realizan actividades que pertenecen a dos o más subsectores.

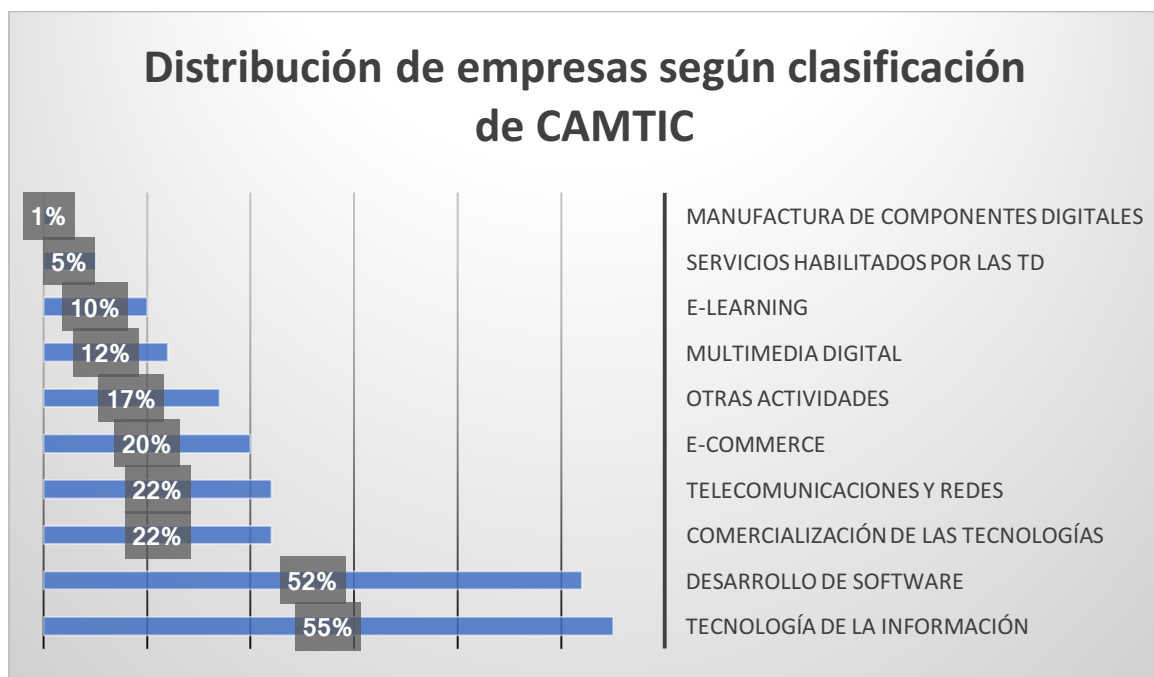


Figura 5 – Distribución de empresas según clasificación de CAMTIC. Fuente: CAMTIC, 2014

CAMTIC identificó en el informe “Mapeo sectorial de tecnologías digitales 2014” que un 11% del total de empresas de tecnologías digitales en Costa Rica manejan actividades

relacionadas a big data, y dentro del subsector de tecnologías de la información representan un 19,8% de las empresas. De esas empresas que trabajan con big data, un 87,5% ofrecen los servicios a clientes nacionales, un 66,7% a transnacionales en Costa Rica y un 58,3% únicamente a clientes en el exterior.

Por su parte, el Banco Central de Costa Rica (BCCR) realizó un estudio denominado “Exportaciones de servicios por medio de redes TIC” con el objetivo de cuantificar y caracterizar las exportaciones de servicios suministradas por medio de redes de tecnología, información y comunicación (TIC). Convirtiendo a Costa Rica en el primer país del mundo en realizar esta cuantificación. (BCCR, 2018)

El estudio tomó en cuenta información brindada por 984 empresas que exportan servicios tecnológicos, divididas en las nueve categorías mencionadas anteriormente según la clasificación de CAMTIC. Durante el 2016 Costa Rica realizó exportaciones de servicios por medio de tecnologías de información y comunicación por un valor de 3.310 millones de dólares, lo que equivale a un 5.8% del producto interno bruto del país, similar al porcentaje que represente la actividad agrícola.

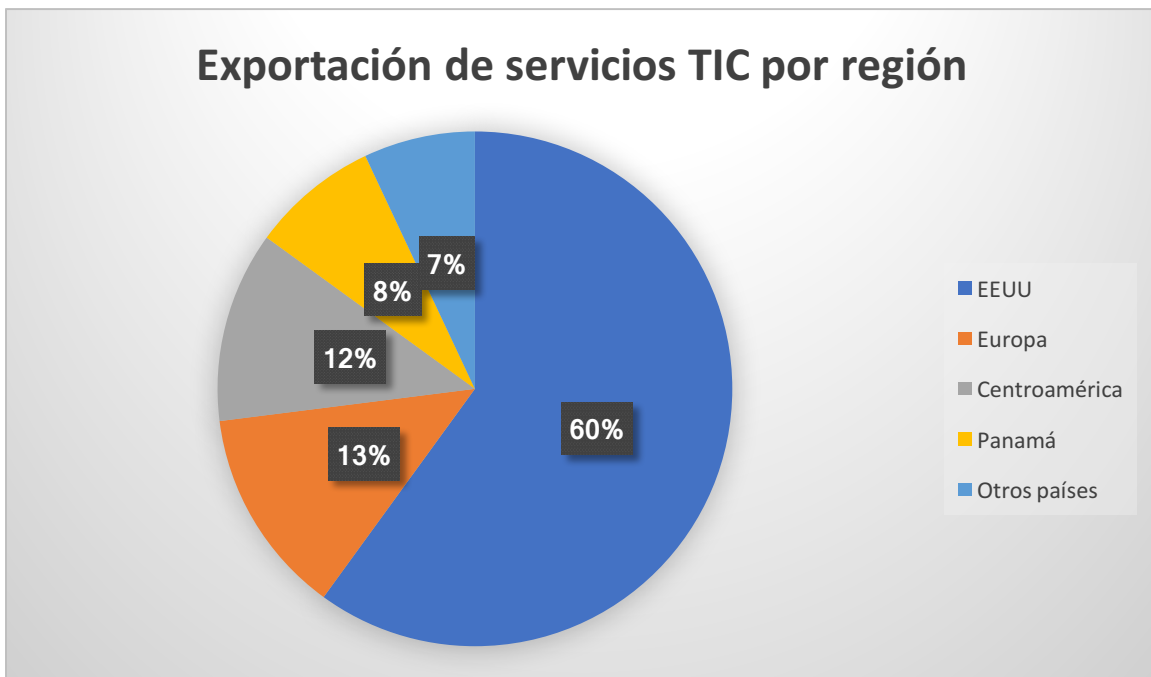


Figura 6 - Exportación de servicios TIC por país o región de destino. Fuente BCCR 2018

## 4.5 Anotaciones finales

El análisis de big data es una tendencia que va en aumento según lo demuestran las proyecciones de inversión en esta área. Existe también una creciente necesidad de profesionales preparados para hacerle frente al análisis de datos cada vez más complejos, motivados principalmente por los avances en la tecnología que permiten recopilar datos que antes no era posible obtener.

Como se ha mencionado previamente, big data por sí solo no tiene mayor valor, toma valor cuando es analizado de una manera que esté alineada a la visión de la organización. Además, se debe tomar en cuenta la calidad de los datos, en especial cuando el análisis de éstos influye en la toma de decisiones.

Aunque Estados Unidos lidere la adopción de big data, los esfuerzos que se están gestando en la Unión Europea y el Reino Unido tienen como objetivo cerrar esa brecha que existe entre ambas regiones y mejorar su nivel de competitividad con la formación de profesionales mejor preparados y leyes más robustas que respalden el uso de los datos. América Latina por su parte está en un proceso de adaptación, y no debe perderse de perspectiva que existe un gran potencial en el análisis de datos para mejorar las condiciones de los países en vías de desarrollo.

Es importante reconocer que la aplicación de big data y analytics no es algo exclusivo del sector de tecnología de información. Por el contrario, industrias como la banca y los seguros han hecho uso de análisis de datos desde mucho antes de que se utilizara el término big data.

De la misma manera, se hace énfasis en que big data no favorece únicamente al sector privado. Precisamente, ese potencial que tiene big data de ayudar a los países en vías de desarrollo puede ser explotado por el sector público, mejorando la calidad de vida de los habitantes de una región utilizando los datos para entender mejor sus necesidades.

## Capítulo V: Solución del problema

En el capítulo II, correspondiente al marco teórico, se expusieron varias definiciones del concepto de mejores prácticas, así como algunas características relevantes para este trabajo de investigación. En este contexto, se entiende como “Guía de mejores prácticas”, una serie de acciones para la implementación de big data y analytics enfocado en la toma de decisiones organizacional. Se hace la salvedad que la guía está compuesta por recomendaciones, no requerimientos y cada organización puede aplicar lo que mejor se adapte a su estructura y necesidades.

Esta guía de mejores prácticas está sustentada en la información recopilada por medio de investigación tanto en el marco teórico como en el análisis de la situación actual. Se tomó en cuenta la experiencia en otros países, así como el mercado de las tecnologías de información en Costa Rica.

### 5.1 Desarrollo de la solución

La transformación digital representa una serie de retos para la mayoría de las organizaciones. Esta solución está enfocada principalmente en identificar, capturar, manejar y analizar big data. Una gran parte las organizaciones están dispuestas a utilizar la información resultante del análisis datos, para descubrir las ideas que impulsarán las decisiones estratégicas de negocios o mejorarán los servicios ofrecidos según el sector en el que se desarrolle.

La guía de mejores prácticas está dividida tres secciones principales:

1. Manejo del negocio
2. Manejo de los datos
3. Análisis de datos

El objetivo de esta división es delimitar el enfoque de cada área llevando un orden específico en su ejecución, comenzando por tener un claro entendimiento del manejo del negocio, posteriormente definir el manejo de los datos con los objetivos de negocio en mente y finalmente establecer las pautas para el manejo de datos, de igual forma tomando en cuenta los dos puntos anteriores.

Dentro de estas secciones se definen las mejores prácticas identificadas durante la investigación, a su vez estructuradas de la siguiente manera:

- ¿Cuál es la mejor práctica?
- ¿Por qué debe ser seguida?
- Otros aspectos relevantes a la fase específica

Posteriormente en el apartado 5.2 se presenta el procedimiento de implementación, se detalla cómo puede ser implementada la mejor práctica, siguiendo la misma estructura anteriormente mencionada. A continuación se explica en detalle las secciones de la guía propuesta.

## 5.1.1 Manejo del negocio

### 5.1.1.1 – Comenzar con conocimiento

Es necesario un consenso entre los diferentes departamentos cuando se manejan términos que son nuevos y que pueden generar confusión. Debe existir un conocimiento general que ayudará a definir expectativas claras sobre la incorporación de una nueva tecnología.

#### ¿Por qué?

La existencia de un lenguaje común con respecto a big data y analytics fomentará la generación de ideas a través de diversos equipos internos de la organización. Uno de los factores que frenan la implementación de big data es la brecha cognitiva que existe en los diferentes departamentos de las organizaciones.

Big data tiene características que varían según la literatura que se consulte, en especial cuando se habla más allá de sus cuatro principales, volumen, variedad, velocidad y veracidad. Las analíticas por su parte también tienen conceptos que deben quedar claros, como por ejemplo los tipos de analíticas que se pueden aplicar según las necesidades de la organización.

### 5.1.1.2 – Alineamiento de Big Data con los requerimientos de negocio

La organización debe tener una idea clara del propósito de los datos, asegurándose que dicho propósito está alineado con los objetivos de la organización.

#### ¿Por qué?

Conocer claramente cuáles son los objetivos del negocio no puede perderse de vista independientemente del tipo de proyecto. Como se ha mencionado anteriormente, big

data debe ser una iniciativa que responda a las necesidades de la organización, no una tarea propiamente del departamento de tecnologías de información, por lo tanto, es importante comprender que el proyecto tendrá un objetivo orientado al negocio y no a la tecnología.

## **5.1.2 Manejo de los datos**

### ***5.1.2.1 – Evaluar los requerimientos de datos***

Con los objetivos de negocio claros, se definen los requerimientos de los datos mediante un análisis de lo que se tiene hasta el momento y lo que se desea alcanzar mediante el uso de big data.

#### **¿Por qué?**

Tener una gran cantidad de datos no es equivalente a buenos datos, y como se ha expuesto anteriormente, big data se trata de datos masivos sumamente variados siendo creados constantemente. Puede que la organización cuente con los datos correctos mezclados entre muchos datos que no pueden generar valor para la organización. El no determinar cuáles son los datos realmente útiles significa realizar trabajo que pudo evitarse habiendo definido los objetivos con claridad. Entre más datos aleatorios se recopilen, más a menudo se desorganizarán.

Parte de la definición de big data, en la que coinciden diversos autores, es que tiene cuatro dimensiones principales, volumen, velocidad, variedad y veracidad. Además de estas cuatro dimensiones algunos autores agregan otras como: valor, validez y volatilidad. Es recomendable evaluar big data en todas sus dimensiones.

Lo que se busca en este punto es determinar qué tipo de datos necesita la organización para alimentar la herramienta que le genere una ventaja competitiva.

### **5.1.2.2 – Definir una estrategia de Big Data**

Tener una estrategia de big data definida es esencial para que el proyecto de implementación sea exitoso. Las organizaciones deben buscar una ventaja competitiva constantemente y por ende deben también enfocarse en obtener, almacenar y analizar los datos que brinden esa ventaja.

#### **¿Por qué?**

Uno de los principales problemas a los que se enfrentan las organizaciones es que no hacen un uso adecuado de big data, se enfocan en recolectar la mayor cantidad de datos sin tener una estrategia de cuáles son los datos que en realidad deberían recolectarse.

Como si fuera poco, se debe hacer énfasis en la velocidad de los cambios con respecto a big data, es probable que cuando se defina una estrategia, ésta ya esté desactualizada y amerite una revisión. De aquí la importancia de hacer una implementación iterativa.

Una vez se hayan determinado qué tipo de datos necesita la organización, se procura definir cómo obtendrán éstos precisamente.

### **5.1.2.3 – Establecer parámetros calidad (datos correctos)**

Se continúa trabajando con relación a los objetivos de la organización y las metas definidas para el uso de los datos. Tener políticas claras sobre el tipo de datos que se desea almacenar y procesar es la base para una mejor toma de decisiones.

#### **¿Por qué?**

Las organizaciones necesitan datos de buena calidad para que los encargados de la toma de decisiones puedan tener la confianza de basar sus decisiones en datos. Más que la cantidad de los datos, importa que sean los datos adecuados y de la mejor calidad

posible. Con los datos correctos se puede obtener información valiosa sobre los clientes, predecir tendencias futuras y crear nuevas oportunidades de negocio.

Se debe tener presente que una de las características de big data es la veracidad, donde por definición se dice que los datos son relevantes cuando son confiables, que éstos no deben tener sesgos ni ruido.

Anteriormente ya se tuvo que haber definido el tipo de datos que necesita la organización y cómo obtenerlos, en este momento corresponde definirse cómo diferenciar los datos realmente útiles.

#### ***5.1.2.4 – Revisión de legislación vigente***

La protección de datos es un tema que ha tomado fuerza con el avance de big data, la cantidad de información personal que se recolecta, almacena y analiza tiene un potencial muy grande y abre las puertas a un sinnúmero de aplicaciones positivas, sin embargo, se debe tener extremo cuidado pues las implicaciones negativas por el mal uso de los datos también pueden tener un gran impacto en la organización.

#### ***¿Por qué?***

Big data y analytics es una herramienta bastante nueva y sin lugar a dudas muy poderosa. Se desea eliminar cualquier duda sobre posibles usos que se le pueda dar a los datos de una manera ilícita o sin respaldo legal. Como se ha mencionado anteriormente, los datos son considerados el petróleo del futuro, por lo que la cautela en el aspecto legal es más que necesaria.

Internet es un servicio global, sin embargo, la privacidad no tiene el mismo alcance. Las leyes de privacidad, las medidas y los organismos reguladores varían de un país a otro. Por ejemplo, en la Unión Europea existe la Regulación General de Protección de Datos (GDPR) que rige toda la región, pero cada país introduce las directivas de manera concreta y con algunas variantes.

### **5.1.2.5 – Creación de plataforma que trabaje para la organización**

La infraestructura que acompaña a los datos es igual de importante que los datos mismos. Crear una plataforma que ayude a cumplir las metas organizacionales debe ser una tarea continua.

#### **¿Por qué?**

Los cambios tecnológicos ocurren todos los días y las organizaciones deben estar preparadas para hacerle frente a estos cambios y seguir siendo competitivos. Una base sólida que sea flexible con los cambios en el entorno asegurará una mejor adaptabilidad en el futuro.

### **5.1.2.6 – Utilizar una implementación iterativa**

Los proyectos de big data no deben ser demasiado ambiciosos en un inicio, en especial cuando no se tiene experiencia en el tema. Se recomienda comenzar con un proyecto piloto o prueba de concepto (*proof of concept*) que sea pequeño y fácil de manejar. Posteriormente se puede aplicar lo aprendido en proyectos de mayor escala y complejidad.

#### **¿Por qué?**

Al igual que con la incorporación de nuevas tecnologías, existe una curva de aprendizaje cuando se habla de análisis de big data, y se debe evitar abarcar más de lo que la organización puede manejar.

Es muy común que las necesidades de la organización cambien conforme van entendiendo los datos, en especial cuando se comienza a comprender el valor potencial de los mismos. Además, las organizaciones deben estar atentas en la creación de nuevas fuentes de datos, por lo que debe existir flexibilidad en el planteamiento inicial.

Hacer una implementación iterativa y con metodologías ágiles permite brindar soluciones rápidas a necesidades inmediatas, que no podría hacerse cuando se hace un desarrollo sin seguir un proceso particular y sin tener certeza clara de los requerimientos.

#### ***5.1.2.7 – Entender el impacto de Big Data***

La organización debe tener un claro entendimiento de los requerimientos de big data, en primera instancia para ser recolectados y almacenados y posteriormente para ser analizados y obtener un beneficio de los mismos. Se busca explotar la capacidad que tiene big data de generar valor y ventaja competitiva.

#### ***¿Por qué?***

La implementación de big data puede impactar la arquitectura de tecnologías de información de la organización de múltiples maneras. Por ejemplo, muchas organizaciones tienen hardware estandarizado, sistemas de manejo de bases de datos y plataformas de análisis de datos que no serán suficientemente capaces de manejar el volumen, velocidad o variedad de información, así como tampoco la capacidad de procesamiento de información que demanda big data.

#### ***5.1.2.8 – Planear para el futuro***

Las organizaciones que ya estén implementando o tengan planeado implementar big data analytics, deben tener en cuenta los constantes cambios tecnológicos y de mercado para tomar acciones proactivas en lugar de reactivas.

## ¿Por qué?

Las fuentes y metodologías de big data cambian constantemente, a esto se le suman otros factores como los avances en los campos de aprendizaje automático (*machine learning*) e inteligencia artificial.

El análisis de big data seguirá tomando fuerza en el futuro como lo demuestran las proyecciones detalladas tanto en el marco conceptual como en el diagnóstico de la situación actual. Sin lugar a dudas será pieza clave para la toma de decisiones en muchas organizaciones.

En un artículo reciente sobre la industria médica (Abraham, T, 2018), se hace referencia a un estudio de la firma americana de consultoría McKinsey donde se expone que los servicios de análisis de datos e información crecerá entre un 16 y 18 por ciento en los próximos cinco años, mientras que los servicios administrativos se espera que disminuyan debido a la automatización.

Los encargados del manejo de las tecnologías de información en la organización que ponga en práctica proyectos de big data deberá aceptar que el cambio será la constante y que la organización debe estar lista para acomodarse a esos cambios para seguir siendo competitivos.

### 5.1.3 Análisis de datos

#### 5.1.3.1 – Estrategia de análisis de datos

Tener definida una clara estrategia para el análisis de datos enfocada en resolver problemas de negocio debidamente definidos, que de ser resueltos tendrán un impacto sobre el rendimiento de la organización.

#### ¿Por qué?

Al igual que es importante tener una estrategia de big data definida, el análisis de datos también debe estar claramente delimitado. El análisis de datos es la extracción del conocimiento de enormes cantidades de información que por sí solos no tienen valor alguno.

En la estrategia de análisis de datos se debe tomar en cuenta la formulación de hipótesis que darán lugar a nuevos descubrimientos. Estas hipótesis normalmente están basadas en experiencia previa y en el descubrimiento por medio de investigación.

## 5.2 Procedimiento de implementación

En esta sección se explica el procedimiento de implementación de la guía tomando en cuenta la misma estructura sobre la que la guía está creada.

### 5.2.1 Manejo del negocio

#### 5.2.1.2 – *Comenzar con conocimiento*

Capacitar a los involucrados en el proyecto en los temas que son nuevos para la organización, haciendo énfasis en los conceptos de big data y analytics, sus características, beneficios, repercusiones y costos asociados.

Posterior a la capacitación se planean talleres donde los involucrados en el proyecto comparten su perspectiva de big data y analytics para asegurarse que entre todos los presentes exista un consenso de los conceptos básicos de big data. Se documenta el resultado de los talleres para tener una futura referencia en caso de que se incorporen nuevas personas al proyecto.

#### 5.2.1.2 – *Alineamiento de Big Data con los requerimientos de negocio*

Se inicia por recopilar, analizar y entender las necesidades del negocio. Este proceso debe ser completado antes de comenzar el proceso de planificación para aprovechar el análisis de Big Data.

La gerencia de la organización debe tomar la iniciativa y el departamento encargado de tecnologías de información debe trabajar según lo que se ha definido como necesidades de negocio. Si la gerencia no establece las metas claras, no se podrá recopilar y generar datos de manera correcta. Se debe evitar la recolección sin objetivos claros, pues de lo contrario se recopilarán todos los datos que se encuentren para posteriormente trabajar

en definir lo que se necesita y lo que se puede desechar, generando confusión, trabajo innecesario y costos elevados.

Las metas y objetivos definidos en esta fase, serán la base para las decisiones que se tomarán en el futuro del proyecto. Con una base clara los cambios tecnológicos no serán un problema, siempre y cuando se trabaje de manera iterativa, pues el objetivo del negocio no debería verse afectado por los avances en tecnología, a fin de cuentas, los objetivos son de negocio y no tecnológicos.

En este punto se recomienda definir los KPIs (medidores de desempeño) de negocio que se utilizarán para evaluar el comportamiento del proyecto de datos. Algunos ejemplos que pueden ayudar a definir los KPIs son:

- Lecciones aprendidas de la implementación de Big Data
- Ahorro obtenido luego de la implementación
- Aumento en ventas o retención de clientes
- Mejoras de procesos

### ***Otros puntos importantes***

Se recomienda además tomar en cuenta y documentar los siguientes aspectos:

- Conocer el contexto organizacional
- Estrategias y plan de negocios
- Principios de negocio, objetivos y motivadores
- Existencia de *frameworks* implementados
- Gobernanza y marcos legales
- Estrategia de tecnologías de información

## **5.2.2 Manejo de los datos**

### ***5.2.2.1 – Evaluar los requerimientos de datos***

Se elabora una lista de las dimensiones de big data y lo que la organización considera importante de cada una de ellas. El objetivo es tener una visión clara de las características de big data y el impacto que pueden tener.

- Volumen: cantidad de datos esperada de cada fuente. Se debe tener en cuenta que la cantidad de datos generada aumenta considerablemente cada año.
- Variedad: los datos pueden ser obtenidos de diferentes formas y técnicas, entre más variadas las fuentes mejor. La mayoría de los datos suele venir de redes sociales.
- Velocidad: el tiempo que transcurre desde que se recolectan los datos hasta que se realiza un análisis de ellos.
- Veracidad: verificar que los datos que se obtienen realmente sirven para tomar decisiones basados en ellos.

### ***5.2.2.2 – Definir una estrategia de Big Data***

Lo principal en la definición de la estrategia es cómo obtener el conocimiento de big data, que por sí solo no tiene mayor valor. Es común que las organizaciones enfrenten obstáculos técnicos, tanto maximizando los sistemas existentes como por otro lado valorando la compra de nuevos sistemas.

En primera instancia se debe hacer un análisis que considere con que datos cuenta la organización actualmente. Hacer un recuento de lo que no se tiene es igual de importante que determinar con que se cuenta en un inicio, como por ejemplo una organización ya que cuente con una fuente de datos propia.

Una vez que se hayan recolectado los datos necesarios para un proyecto, se procede a identificar lo que puede estar faltando. Es posible que inicialmente no se pueda contemplar con exactitud todos los datos que se necesiten, por lo que es importante tener en cuenta la flexibilidad y no perder la noción de que se trata de un proyecto iterativo.

En la gran mayoría de casos será necesario probar los datos y revisar los resultados. Existe la posibilidad de que en las primeras iteraciones no se obtengan los resultados esperados, por lo que se recomienda repetir el proceso con diferentes variables en una pequeña escala hasta encontrar los resultados deseados para posteriormente entrar de lleno con el set de datos completo. Se amplían detalles sobre la implementación iterativa en el punto 5.2.2.6 de este documento.

Cuando se tenga definido que datos hacen falta, se debe determinar de dónde y cómo obtener dichos datos para alcanzar los objetivos previamente definidos. Algunas fuentes de datos a considerar con las siguientes:

- **Documentos internos:** cualquier documento propio de la organización, sean archivos de procesamiento de texto, hojas de cálculo, PDF o similares.
- **Datos archivados:** versiones digitalizadas de documentos físicos.
- **Archivos multimedia:** fotografías, audios y videos.
- **Bases de datos:** pueden ser bases de datos convencionales SQL, NoSQL o sistemas Hadoop.
- **Aplicaciones de negocio:** software ERP, puntos de venta, y de manejo de contenido entre otros.
- **Datos de redes sociales:** datos no estructurados de diversas redes sociales.
- **Datos generados sin interacción humana:** proveniente de sensores integrados en dispositivos inteligentes (IoT) u otro tipo de máquinas. Con información variada como geolocalización, temperatura, ruido o hasta datos biométricos.

Posteriormente deben definirse parámetros de calidad, mantener un orden en los datos que se almacenan y se procesan simplificaran la tarea del análisis posterior.

#### ***5.2.2.4 – Establecer parámetros calidad (datos correctos)***

El equipo de datos y la unidad de negocios deben trabajar en conjunto para alcanzar las metas de negocio fijadas en la etapa anterior. Los encargados del análisis de datos (científicos de datos o afines) hacen su trabajo mediante el modelado de datos, y se espera que ellos comprendan lo que los encargados de la unidad de negocio desean obtener. Por otro lado, se espera que los líderes del negocio tengan al menos una noción básica de lo que se puede lograr o no con el uso de los datos.

Se deben definir políticas del tipo de datos que se desea almacenar. Existe una tendencia de guardar todos los datos obtenidos de manera cruda (sin ningún tipo de procesamiento), a pesar de que la gran parte de esos datos nunca se vayan a utilizar, esto debido a que existe un temor generalizado de que en un futuro esos datos pueden llegar a ser necesarios.

Desarrollar criterios y métodos para la eliminación de datos que estratégicamente no son relevantes a corto y largo plazo es una forma de controlar la cantidad de datos que se requiera almacenar, reduciendo además el costo asociado por guardar y procesar estos datos.

Se recomienda también determinar cuáles son los parámetros que tendrán un mayor impacto en la organización, se podrá hacer un mejor uso de los datos que se tiene a través de un enfoque más centrado en lugar de intentar clasificarlos todos.

#### ***5.2.2.4 – Revisión de legislación vigente***

Se debe realizar una revisión de la legislación local y de otros países donde se tenga operación o donde se vayan a almacenar los datos en caso de que aplique. En el caso de Costa Rica la legislación vigente se encuentra en el sistema SINALEVI (Sistema Nacional de Legislación Vigente). Lo más relevante es contar con la siguiente información:

- Leyes de retención de datos
- Determinando cuánto tiempo los datos pueden ser retenidos por una organización para fines comerciales o legales
- Leyes de privacidad
- Determina cuales datos pueden ser almacenados
- Leyes o regulaciones de uso de datos
- Determina el ámbito en el que pueden ser usados los datos

#### **5.2.2.5 – Creación de plataforma que trabaje para la organización**

El entorno de big data está cambiando sumamente rápido en la actualidad para considerar la construcción de una base definitiva o de larga duración. Por el contrario, se debe planificar adelantándose a los cambios disruptivos que puedan ocurrir en el entorno, tales como:

- Nuevos tipos de datos
- Retos de competitividad
- Cambios en enfoques de programación
- Hardware
- Tecnologías de redes y comunicaciones
- Servicios ofrecidos por proveedores de big data

Esta fase está relacionada con la siguiente, donde se sugiere realizar una implementación iterativa, principalmente motivado por los cambios tecnológicos. Una solución a este reto, es mantener un balance entre diferentes implementaciones, haciendo uso de una mezcla de, por ejemplo:

- Almacenamiento distribuido con un *framework* como *Hadoop*
- Sistemas distribuidos tradicionales (*grid computing*)
- Equipos locales
- Computación en la nube
- Servidores centrales

El punto clave en la creación de la plataforma es tomar en cuenta la flexibilidad. Se asume que en un periodo corto (de 2 a 3 años) se deberá reprogramar y reubicar todas las aplicaciones de big data, por lo que se recomienda tomar en cuenta esta posibilidad. Una alternativa es hacer uso de desarrollo basado en metadatos sin tomar en cuenta el código de programación, únicamente los atributos de la aplicación, lo que a la larga incrementará la productividad y ayudará a aislar la implementación de los cambios tecnológicos subyacentes.

Lo anterior mencionado tiene relación con la planificación a futuro. Según se demostró en el marco conceptual, la cantidad de datos generados se duplica cada dos años y la capacidad de las computadoras, tanto en procesamiento como almacenamiento lo hace a un ritmo similar.

### Requerimientos tecnológicos a considerar

Dependiendo de las necesidades de la organización, se puede elegir entre bases de datos SQL y noSQL, y entre estas opciones una gran cantidad de variantes. Se deben considerar aspectos como:

- Los datos a analizar son estructurados o no estructurados: De esto dependerá si se necesita una base de datos tradicional o una más flexible
- Requerimiento de análisis en tiempo real: Existen opciones como Apache Spark (*framework* de computación en clúster de código abierto) para procesamiento en tiempo real
- Se harán evaluaciones posteriormente: Plataformas como Hadoop permiten el procesamiento por lotes

### Big data en la nube

La computación en la nube es otro de los aspectos de la tecnología de información que aún está en auge. Se debe considerar que los servicios en la nube suelen ser cobrados

por consumo (transferencia de datos y procesamiento) y cuando se trata de big data se puede esperar procesar mucha información. Por el lado positivo, los servicios en la nube pueden establecerse y crecer rápidamente.

Ciertos proveedores de servicios como Amazon EMR (Elastic MapReduce), una plataforma que utiliza Hadoop para el procesamiento de grandes cantidades de datos, ofrecen herramientas que ayudan al prototipado del ambiente de desarrollo, lo cual puede ser muy útil para realizar pruebas iniciales. Una vez se alcance un modelo operativo sólido, se puede replicar en equipo propio para evitar los cargos de uso de la nube.

#### ***5.2.2.6 – Utilizar una implementación iterativa***

Definir un plan piloto con pocos datos y de fácil manejo. Los resultados de este proyecto sirven para formular nuevas pruebas y a la larga definir los futuros proyectos de big data que le generarán valor a la organización.

Seleccionar un área de la organización en la que se quiere mejorar el proceso del negocio y que al mismo tiempo no represente un gran impacto en la misma en el caso de que el proyecto no sea exitoso.

Hacer uso de metodologías ágiles en el proceso, donde se permita una implementación iterativa. El objetivo es tener la habilidad de brindar soluciones rápidas en pequeñas iteraciones basadas en necesidades específicas, en lugar de hacer todo el mismo tiempo. Los requerimientos de la organización irán cambiando conforme se logre tener un mejor entendimiento de big data.

También se puede dividir el proyecto en partes más pequeñas, identificando oportunidades específicas con el potencial de generar gran valor, sin perder de perspectiva cuál es el objetivo principal de la organización.

Por último, y este es un proceso iterativo en sí mismo, la organización debe estar atenta a los cambios en el entorno y no únicamente en cuanto a tecnología. La velocidad a la que se mueven los negocios gracias a un mercado globalizado es muchísimo mayor a lo que era hace unos años. Los datos cambian constantemente creando nuevas oportunidades.

#### ***5.2.2.7 – Entender el impacto de Big Data***

Capacitar a los encargados del área de tecnologías de información e infraestructura para que tengan una mentalidad abierta a formas innovadoras de procesamiento y enfoques híbridos para acomodar la variedad de datos, que se encontrarán durante la implementación del proyecto, tanto estructurados como no estructurados, internos y externos.

Planificar acorde al volumen, velocidad y variedad de fuentes de datos que deban ser manejados en la organización. Es muy probable que los datos no puedan ser almacenados y procesados en la misma localización, en especial si se utilizan marcos de trabajo distribuidos. En estos casos, un enfoque de almacenamiento de datos lógico puede ser más apropiado porque puede proporcionar análisis de datos casi en tiempo real.

Tomando en cuenta la implementación iterativa del punto anterior, se debe planear una arquitectura que permita la expansión para acomodar una carga de trabajo difícil de predecir en etapas iniciales. El almacenamiento dinámico y capacidad de procesamiento que ofrece la nube puede ser una manera una solución atractiva.

#### ***5.2.2.8 – Planear para el futuro***

Crear un plan de formación continua, en especial para los involucrados en la toma de decisiones de la organización y los encargados de tecnologías de información. Para mantener la competitividad es esencial poder tomar acciones sobre los datos, independientemente de los cambios tecnológicos que sucedan en el entorno.

Se recomienda prestar atención en aprendizaje automático e inteligencia artificial. La tendencia en la actualidad es crear algoritmos que puedan recolectar datos de manera más eficiente, analizarlos y obtener valor de ellos sin la necesidad de interacción humana.

## 5.2.3 Análisis de datos

### 5.2.3.1 – Estrategia de análisis de datos

Se debe definir uno o varios enfoques, según las necesidades de la organización, pues con un mismo conjunto de datos se pueden obtener resultados distintos, el enfoque debe ir alineado nuevamente con los objetivos de negocio de la organización.

Los enfoques de análisis de datos son:

- Descriptivo: utiliza métricas para analizar datos crudos y obtener información sobre lo que pasó.
- Descubrimiento: utiliza datos históricos para diagnosticar por qué está pasando algo en particular, ayuda a la identificación de tendencias y patrones.
- Predictivo: utiliza la extrapolación de información para pronosticar comportamientos futuros. Está relacionado con inteligencia artificial y aprendizaje automático.
- Prescriptivo: utiliza los resultados del análisis descriptivo y de descubrimiento para determinar acciones a tomar. Muy utilizado en la industria del transporte.

Estos enfoques están ordenados por complejidad por lo que debe tomarse en cuenta la madurez de la organización en el análisis de big data al hacer la evaluación.

### 5.3 Pruebas y resultados

En este apartado se presentan las pruebas y el análisis de los resultados obtenidos luego de aplicar el piloto de la solución planteada. A grandes rasgos el proyecto busca brindar a las organizaciones una herramienta para facilitar la aplicación de big data y analytics como apoyo en la toma de decisiones.

La aplicación del piloto se basa en probar la implementación de la guía de las mejores prácticas propuesto en el apartado anterior. Para efectos de este proyecto y por limitaciones de tiempo, no se puede aplicar la guía de mejores prácticas en su totalidad, por lo que se han seleccionado una serie de etapas que fueron identificadas como esenciales antes de emprender un proyecto de big data y análisis de datos.

Específicamente se tomaron en cuenta en cuenta los siguientes puntos:

- Comenzar con conocimiento
- Alineamiento de big data con los requerimientos de negocio
- Evaluar los requerimientos de big data
- Definir una estrategia
- Utilizar una implementación iterativa
- Entender el impacto de big data

Cada uno de los puntos mencionados anteriormente tiene su propio procedimiento de implementación, sin embargo, las necesidades de cada organización son distintas, por lo que se detalla lo que fue aplicado en cada uno de estos puntos.

El objetivo principal de este piloto es someter a prueba la propuesta de solución para verificar su funcionalidad y encontrar posibles puntos de mejora. En la etapa de pruebas estuvieron involucrados varios directores de la organización de la cual se detallará más a fondo en el siguiente apartado.

Durante el pilotaje se realizaron varios talleres de capacitación y sesiones de trabajo con la finalidad de explicar cada etapa, los conceptos relacionados, y definir puntos relevantes para el proyecto de implementación.

### **5.3.1 Sobre la organización**

Las pruebas del piloto fueron aplicadas en Commercial Solutions, una microempresa costarricense dedicada al comercio internacional y la búsqueda de soluciones amigables con el medio ambiente.

Para conocer detalles más específicos sobre la organización y su relación actual con big data, se aplicó el cuestionario sobre la organización, que se encuentra adjunto en este documento como anexo 1.

#### **Reseña de la organización**

Commercial Solutions fue fundada en el año 2017 por tres socios, quienes en la actualidad se encargan a nivel gerencial de la misma. La compañía tiene como objetivo brindar soluciones comerciales en diversas áreas, enfocándose principalmente en la innovación tecnológica.

Desde el 2018 comenzaron con la línea CS iluminación, con la misión de brindar soluciones en iluminación eficiente y de menor consumo. Esta línea está orientada a proyectos públicos, compras privadas, distribución de luminarias y todo lo referente a soluciones de iluminación a la medida y amigables con el ambiente.

Se eligió esta organización por ser un proyecto emprendedor que busca abrirse un espacio en el mercado de la iluminación y la innovación tecnológica, dentro de su plan está hacer uso de análisis de datos para conocer mejor el mercado y enfocar sus esfuerzos de una manera más eficiente. Big data puede brindar muchas oportunidades

si se usa adecuadamente, por lo que es importante tener una base de conocimiento que permita obtener una ventaja competitiva.

### 5.3.2 Taller de capacitación

Como parte del apartado **Comenzar con conocimiento** se llevó a cabo un taller de capacitación con las personas a nivel de gerencia interesadas en aplicar big data y analytics en la organización.

La capacitación tuvo como finalidad introducir a los participantes en temas como big data, analytics, sus características, beneficios, repercusiones y costos asociados.

Antes de dar inicio a la sesión, se les pidió a los asistentes llenar un cuestionario para conocer su percepción actual de big data y analytics, así como otros datos relevantes de la organización. Dicho cuestionario se encuentra adjunto a este documento como anexo 1. Los datos recopilados resultaron importantes no solo para el taller de capacitación, sino también para otras etapas del piloto.

Lista de participantes en la sesión

Nombre	Puesto
Julian Jarquín	Director de operaciones
Federico Umaña	Director de mercadeo
Luis Diego Soto	Director de finanzas
Eddy Leitón	Encargado de tecnologías de información

*Tabla 3 - Participantes en el taller de capacitación*

A manera de ayuda audiovisual se utilizó una presentación de diapositivas digitales titulada “Conceptos y definiciones: Big Data & Analytics”, la misma se encuentra adjunta como anexo 2 en este documento.

Durante la presentación se respondieron dudas sobre los conceptos expuestos, además se presentaron casos de éxito de organizaciones que han hecho uso de big data para obtener una ventaja competitiva y mejorar su rendimiento. Se pudo observar que había temas que aún luego de responder las preguntas no quedaban debidamente aclarados, por lo que se tomó nota para profundizar nuevamente durante la siguiente etapa del taller.

Una vez finalizada la presentación se abrió un espacio para que los participantes de la sesión compartieran su perspectiva sobre big data y analytics. Se le dio tiempo a cada uno de los participantes para que expusieran desde su punto de vista de que se trataban los conceptos expuestos en la presentación. Cada intervención fue anotada para posteriormente buscar un consenso entre los presentes y así tener una misma definición para todos.

Cuando se logró alcanzar el consenso sobre las definiciones de los diferentes conceptos, se documentó cada una de ellas para futura referencia del proyecto.

Concepto	Definición
<b>Big data</b>	Son datos que no pueden ser almacenados ni procesados en una cantidad de tiempo definida utilizando métodos tradicionales.  Por si solos los datos no aportan valor, éstos deben ser procesados y analizados para lograr obtener información útil y obtener una ventaja competitiva.
<b>Características de big data</b>	Caracterizados principalmente por su volumen, variedad, velocidad y veracidad.
<b>Analytics</b>	Es el proceso de examinar conjunto de datos para extraer conocimiento de los mismos y que permita tomar acciones basadas en los hallazgos.

<b>Enfoques de analytics</b>	Existen diferentes enfoques que ayuden a cumplir con objetivos distintos, entre ellos están las analíticas descriptivas, de descubrimiento, predictivas y prescriptivas.
<b>Ciclo de vida de big data analytics</b>	<ol style="list-style-type: none"> <li>1. Comprender el negocio</li> <li>2. Comprender los datos</li> <li>3. Minería y modelado de datos</li> <li>4. Interpretación del modelo</li> <li>5. Puesta en marcha del modelo</li> <li>6. Monitoreo</li> </ol> <p>El ciclo puede ser iterativo en cualquier etapa</p>
<b>Beneficios de big data analytics</b>	<p>Aumento de productividad</p> <p>Posibilidad de crear ventaja competitiva</p> <p>Mejor conciencia situacional (con respecto al mercado)</p>
<b>Retos de big data analytics</b>	<p>Cambios tecnológicos constantes</p> <p>Necesidad de manejo de varias estructuras de datos</p>

*Tabla 4 - Documentación de definiciones para el proyecto*

### 5.3.3 Alinear requerimientos

Como parte del cuestionario adjunto como anexo 1, se le pidió a los involucrados en el proyecto información sobre los objetivos del negocio, además que definieran lo que esperaban ellos obtener con la implementación del análisis de datos.

En la etapa anterior se logró llegar a un consenso sobre las definiciones de los conceptos relevantes para la aplicación de big data y analytics. Esto nos sirvió de base para definir los requerimientos de la organización de una manera similar.

Se les pidió a los directores de las áreas de operaciones, mercadeo y finanzas asistir a otra sesión de trabajo y así analizar, discutir y entender las necesidades del negocio que se plantearon como parte de las respuestas del cuestionario anteriormente mencionado. Además, se les brindó la oportunidad de proponer necesidades no visualizadas antes del taller de capacitación.

#### Lista de participantes en la sesión

Nombre	Puesto
Julian Jarquín	Director de operaciones
Federico Umaña	Director de mercadeo
Luis Diego Soto	Director de finanzas

*Tabla 5 Participantes en la sesión de alineación de requerimientos*

No se requirió la presencia del encargado del área de tecnologías de información puesto que el proyecto debe responder a las necesidades de negocio principalmente. Se les recuerda a los presentes que los objetivos fijados en esta etapa no deben verse afectados por los avances tecnológicos, por lo que no tienen que ser fijados a tecnologías en específico.

Durante la sesión, se revisaron los temas relacionados a los objetivos de negocio, se discutieron los puntos aportados en el cuestionario por los directores encargado de cada área. En primera instancia se aseguraron de que todos estaban siguiendo una misma línea de trabajo y posteriormente se procedió a analizar cada uno de estos puntos y determinar cuales tienen más relevancia para el futuro de la organización.

Se identifican y se documentan las principales necesidades del negocio de la siguiente manera:

- Tener un mejor entendimiento del mercado de iluminación en Costa Rica
- Mejorar el posicionamiento de la compañía en el mercado
- Aumentar la captura de información de fuentes de datos
- Manejo óptimo de inventario

Adicionalmente se documentan los siguientes aspectos:

- Estrategias y plan de negocios
  - La estrategia de negocios está orientada en proyectos tanto a nivel públicos como privados, donde exista la necesidad de iluminación.
- Principios de negocio, objetivos y motivadores
  - Brindar una solución óptima de iluminación al cliente
  - Adaptación a las necesidades del mercado
  - Distinción por la calidad del servicio
  - Tener presencia en el diseño de proyectos de iluminación
- Existencia de frameworks implementados
  - No hay frameworks implementados en la organización
- Gobernanza y marcos legales
  - Se rigen bajo el marco legal costarricense y leyes de comercio internacional

#### 5.3.4 Evaluación de requerimientos de big data

Una vez definidos los requerimientos del negocio, se procedió con otra sesión de trabajo para evaluar y definir los requerimientos de big data que van a ayudar a cumplir los objetivos de negocio.

Lista de participantes en la sesión

Nombre	Puesto
Julian Jarquín	Director de operaciones
Federico Umaña	Director de mercadeo
Eddy Leitón	Encargado de tecnologías de información

*Tabla 6 - Participantes en la sesión de requerimientos de big data*

Durante la sesión se hizo una revisión de las principales características de big data: volumen, variedad, velocidad y veracidad. Se explicó nuevamente en qué consiste cada

una de estas características, así como el impacto que cada una de éstas puede tener sobre la organización.

Las características de big data fueron ordenadas por la importancia que tienen para alcanzar los objetivos de negocio, quedando de la siguiente manera:

1. Veracidad
2. Variedad
3. Volumen
4. Velocidad

Se determina que para la organización lo principal es tener la confianza de contar con datos verdaderos y confiables con lo que puedan apoyar la toma de decisiones. El director de operaciones mencionó que ellos no hacen nada con un enorme volumen de datos si no tienen la certeza de su veracidad.

Por otra parte, se puntualizó que se considera importante contar con datos de fuentes variadas. Según las respuestas del cuestionario del anexo 1, la organización pretende utilizar datos de la siguiente manera:

Implementado actualmente	De 1 a 5 años	Más de 5 años
Fuentes de datos públicos (por ejemplo, SICOP)	Correos electrónicos	Datos de punto de venta
	Redes sociales	
	Sensores e IoT	
	Datos geo-espaciales	
	Datos de búsqueda	

Tabla 7 - Implementación de datos en el tiempo

La velocidad quedó ocupando el último lugar de la lista de importancia, todos los involucrados coincidieron que si bien es cierto se necesitan datos actuales, no requieren datos en tiempo real, pues su orientación va más hacia medir la situación actual del

mercado y en un futuro determinar posibles comportamientos de consumo. Esto último está relacionado con los requerimientos de la organización, entre ellos un adecuado manejo de inventario.

### 5.3.5 Definición de estrategia de big data

Se comienza la sesión de trabajo retomando los temas tratados anteriormente, haciendo énfasis tanto en los requerimientos de negocio como los de big data. Además, se les recuerda a los asistentes que la estrategia de big data que se defina dependerán los resultados que se obtengan más adelante.

El director de mercado cuestiona si la estrategia que se defina será definitiva, a lo que se le responde que la estrategia puede y debe ser revisada periódicamente según los resultados obtenidos. Se aprovecha para abordar el tema del concepto de implementación iterativa y se les comenta que explicará más a fondo en una futura sesión de trabajo.

Lista de participantes en la sesión

Nombre	Puesto
Julian Jarquín	Director de operaciones
Federico Umaña	Director de mercadeo
Eddy Leitón	Encargado de tecnologías de información

*Tabla 8 - Participantes en la sesión de definición de estrategia de big data*

Como segundo punto en la agenda de la sesión de trabajo, se hace un recuento de los datos con los que cuenta la organización. Esta información se obtuvo como parte del cuestionario adjunto como anexo 1, y se les consultó a los asistentes si tenían algo más que agregar a la lista, sin embargo, no hubo nada más que agregar.

Se documentó que cuentan con las siguientes fuentes de información propia:

- Documentos internos:
- Hojas de cálculo:
- Manejo de inventario
- Contactos de proveedores
- Documentos PDF
- Catálogo de productos
- Manual de productos
- Facturas
- Otros
- Datos archivados:
- Respaldos digitales de facturas
- Respaldos digitales de contratos
- Archivos multimedia:
- Fotografías
- Fotografías de los productos
- Fotografías de trabajos realizados
- Videos
- Videos demostrativos
- Bases de datos:
- MySQL
- Base de datos de clientes e interesados
- Aplicaciones de negocio:
- Software ERP – FinanzaPro

Cuando se tuvo claridad de los datos con los que cuenta la organización, se procedió a hacer un listado de la información que se considera necesaria para alcanzar los objetivos de la organización.

Se documentó que los directores de la organización consideran necesario obtener los siguientes datos:

- Datos de búsqueda de luminarias en Costa Rica (en buscadores como google)
- Datos de empresas con alto consumo eléctrico asociado con iluminación
- Datos de fabricación de luminarias a nivel mundial
- Datos de reducción de consumo con implementación de luminarias de última tecnología
- Datos de proyectos públicos y privados donde se necesiten luminarias

El director de mercadeo mencionó que hay fuentes de datos ya existentes que son relevantes para la organización que aún no han podido aprovechar por diferentes motivos, entre ellos la falta de presupuesto, dificultad de acceso a la información en algunas entidades como la ESPH (Empresa de Servicios Públicos de Heredia) y la falta de recursos.

También se documentaron estas fuentes de información para considerar en el futuro:

- Connexis Leads: empresa costarricense dedicada al servicio de análisis de datos de procesos de compras públicas del gobierno de Costa Rica.
- SICOP (Sistema Integrado de Compras Públicas)
- CCC (Cámara Costarricense de la Construcción): cuentan con datos de construcciones tanto a nivel público como privado. Tienen un listado de las empresas que solicitan permisos de construcción, contacto de los encargados y nombre de los proyectos.
- PROCOMER (Promotora de Comercio Exterior de Costa Rica) y el Ministerio de Hacienda cuentan con datos de importación de productos, lo que puede ayudar a medir las tendencias del mercado en Costa Rica.

A pesar de que en el plan piloto no se incluyó la tarea de establecer los parámetros de calidad de los datos, se discutió sobre este tema debido a que la veracidad de los datos fue la característica calificada como de mayor importancia durante la evaluación de requerimiento de big data.

Con respecto a los parámetros de calidad de datos se establecieron una serie de lineamientos que servirán de base para cuando se realice la implementación completa. Estos serían:

- No se almacenarán datos “crudos”, esto con la finalidad de reducir el espacio necesario de almacenamiento y el costo asociado a esto
- Se buscará obtener datos de fuentes confiables (motores de búsqueda, instituciones gubernamentales o datos propios)

### 5.3.7 Taller de implementación iterativa

Al igual que el primer taller, este tuvo el objetivo de aclarar conceptos que podían ser nuevos para algunos de los involucrados en el proyecto, así como definir sobre qué área u objetivo de la organización se puede definir el primer proyecto de big data a modo de prueba.

Lista de participantes en la sesión

Nombre	Puesto
Julian Jarquín	Director de operaciones
Federico Umaña	Director de mercadeo
Luis Diego Soto	Director de finanzas
Eddy Leitón	Encargado de tecnologías de información

*Tabla 9 - Participantes en el taller de implementación iterativa*

Como ayuda audiovisual se utilizó una presentación de diapositivas digitales titulada “Conceptos de implementación iterativa”, la misma se encuentra adjunta como anexo 3 en este documento.

Se hizo énfasis en que para la implementación de un proyecto de análisis de big data, es importante tomar en cuenta el uso metodologías ágiles, en las que el proceso sea iterativo e incremental, pues los requisitos irán evolucionando junto con las necesidades de negocio para mantener la competitividad.

Posteriormente se trabajó sobre las necesidades de la organización con el objetivo de delimitar un proyecto pequeño de big data y analytics que sirva de piloto de prueba, tanto de los puntos definidos en las etapas anteriores, como para el proceso iterativo, donde según los resultados obtenidos se podrán ir haciendo los ajustes necesarios conforme se comprenda mejor la relación entre los datos y el análisis hasta encontrar lo que funcione mejor para la organización.

Entre los directores de la organización acordaron que el primer esfuerzo de big data y analytics debería estar orientado a conocer mejor el mercado de las luminarias en Costa Rica, pues sin esta base no podrían determinar por ejemplo las necesidades del mercado ni trabajar en mejorar el manejo de inventario.

### 5.3.8 Impacto de big data

Esta sesión de trabajo estuvo orientada hacia el encargado de tecnologías de información, sin embargo, se consideró importante la asistencia de los directores para que estuvieran al tanto del impacto que puede tener big data en la organización, no solo a nivel de infraestructura de tecnología.

Lista de participantes en la sesión

Nombre	Puesto
Julian Jarquín	Director de operaciones
Federico Umaña	Director de mercadeo
Luis Diego Soto	Director de finanzas
Eddy Leitón	Encargado de tecnologías de información

*Tabla 10 - Participantes en la sesión de trabajo de impacto de big data*

Se explicó el impacto que big data puede tener sobre la organización, principalmente a nivel de infraestructura. Se habló sobre los servicios en la nube que consume

actualmente la organización, su infraestructura local actual y el enfoque que deben tomar para alcanzar los objetivos de la organización mediante el uso de datos y analíticas.

El principal servicio externo que consume Commercial Solutions es FinanzaPro Cloud, que brinda una solución de facturación y manejo de inventario en la nube, además de reportes variados. Todos los demás datos se manejan de manera local.

#### 5.4 Anotaciones finales

La aplicación del piloto fue muy valiosa para determinar campos de mejora en la organización donde fue aplicado, pues sus directores consideran que necesitan informarse más sobre el análisis de datos para considerarlos dentro de la cultura de toma de decisiones. De igual manera sirvió también para aplicar cambios en la solución propuesta.

En cuanto a la guía de mejores prácticas, se evidenció la importancia de que exista un conocimiento previo sobre lo que es big data y analytics antes de iniciar cualquiera de las demás etapas.

La propuesta de solución sugiere que la guía de mejores prácticas puede ser aplicada a discreción de la organización y según sus necesidades, y que por lo tanto no es necesario que se cumplan todas las etapas. Sin embargo, luego de la experiencia del pilotaje, la recomendación es que la primera etapa relacionada a la introducción de conceptos básicos, sea completada siempre como primer paso. Esto debido a que se deben crear las bases de conocimiento definiendo un lenguaje común para todos los involucrados en el proyecto.

Como resultado de la aplicación del plan piloto se obtuvo la documentación necesaria para la implementación de las fases restantes de la guía de mejores prácticas, así como un mejor entendimiento sobre el análisis de big data por parte de los directores Commercial Solutions, quienes consideran necesarios hacer varios ajustes en la

organización, así como la asignación de una mayor partida del presupuesto operativo para el área de tecnologías de información, tal y como se ve reflejado en sus respuestas en el cuestionario de valoración de aplicación del piloto, adjunto en este documento como anexo 4.

## Capítulo VI: Análisis financiero

El análisis financiero se justifica principalmente por el aumento en la productividad de la organización, en el caso de Commercial Solutions en la cual se aplicó el piloto, esto se traduce en un aumento en las ventas, bajo el supuesto de la aplicación de la guía de mejores prácticas para un proyecto de big data y analytics.

Según datos recopilados en el capítulo II marco teórico, un estudio del Parlamento de la Unión Europea demostró que las organizaciones que aplican big data analytics incrementan su productividad entre 5% y un 10% (UK Intellectual Property Office Informatics Team, 2014). Por su parte McAfee y Brynjolfsson (2018) aseguran que el uso de big data analytics para la toma de decisiones resulta en un aumento de productividad de un 5% en promedio, y un aumento de rentabilidad de un 6% en comparación a sus competidores.

### 6.1 Inversión inicial

En la tabla 11 se detallan los componentes incluidos en la inversión inicial que se requiere para la puesta en marcha de la solución planteada. Las actividades que conforman la inversión inicial son aquellas detalladas la solución, que corresponde al capítulo V de este documento.

El costo por hora refleja el promedio del cobro de un consultor en el área de las tecnologías de información.

Costo por hora: ₡ 25.000,00

Actividad	Horas estimadas	Costo Total
Implementación mejores prácticas relacionado al manejo del negocio	40	₡ 1.000.000,00
Implementación mejores prácticas relacionado al manejo de los datos	150	₡ 3.750.000,00
Implementación mejores prácticas relacionado al análisis de datos	40	₡ 1.000.000,00
<b>Costo total de la implementación</b>	<b>230</b>	<b>₡ 5.750.000,00</b>

Tabla 11 – Detalle de la inversión inicial. Fuente: Elaboración propia

El costo total de la inversión inicial se estima en ₡ 5,750,000.00.

## 6.2 Recuperación de la inversión inicial

La empresa espera recuperar la inversión inicial por medio de un aumento en las ventas y una reducción de los gastos operativos relacionados a esta actividad, con acciones como la mejora de manejo de inventario y búsqueda de proveedores que se adapten de una mejor manera a las necesidades de los clientes.

Se toma como supuesto, basado en la revisión de literatura, que la empresa mejorará su rentabilidad en al menos un 5% luego de la implementación de big data y analytics para la toma de decisiones.

Por lo tanto, se proyecta que los ingresos aumentarán en un 5% y al mismo tiempo los gastos operativos también disminuirán en el mismo porcentaje.

### 6.3 Proyección del flujo de efectivo

La proyección del flujo de efectivo se realiza a un plazo de 5 años, tiempo en el que la organización estableció como límite para la implementación completa de big data y analytics como apoyo para la toma de decisiones.

Para la proyección no se toma en cuenta la depreciación, pues la solución planteada está orientada hacia el conocimiento, y la aplicación de mejores prácticas en la organización.

	Periodo 0	Periodo 1	Periodo 2	Periodo 3	Periodo 4	Periodo 5
Inversión	₪ (5,750,000.00)					
Ingresos	₪-	₪13,831,498.80	₪14,523,073.74	₪15,249,227.43	₪16,011,688.80	₪16,812,273.24
Gastos Operativos	₪-	₪12,450,352.00	₪11,827,834.40	₪11,236,442.68	₪10,674,620.55	₪10,140,889.52
Flujo neto de efectivo	₪ (5,750,000.00)	₪1,381,146.80	₪2,695,239.34	₪4,012,784.75	₪5,337,068.25	₪6,671,383.72

Tabla 12 – Proyección del flujo de efectivo

### 6.4 Análisis final de la inversión

Para determinar la rentabilidad del proyecto, se tomaron en cuenta los parámetros del Valor Actual Neto (VAN) y la Tasa Interna de Retorno (TIR). El VAN indica el valor del proyecto en la actualidad en valores monetarios, mientras que el TIR es una medida porcentual.

La tasa de descuento utilizada para las proyecciones es de un 15%, esto se refiere a la tasa de rendimiento o rentabilidad mínima que se espera obtener como resultado.

Se considera que cuando el VAN es mayor o igual a 0 el proyecto es rentable. Cuando el VAN es igual a 0 se está cumpliendo con la tasa de descuento. Un valor mayor quiere decir que se cumplió con la tasa y además se generó un beneficio adicional.

En cuanto al TIR, se considera que la inversión es aceptable cuando el valor resultante es mayor que la tasa de descuento. Cuando el valor resulta igual a la tasa de descuento es indiferente si se realiza el proyecto o no, y cuando es menor, el proyecto no será rentable.

<b>Tasa de descuento (referencia)</b>	15%
<b>VAN</b>	₪ 6,495,799.36
<b>TIR</b>	45.41%
<b>PER</b>	₪ 8,089,170.89
<b>Periodo de recuperación</b>	2 años y 5 meses

*Tabla 13 – Detalle del análisis financiero*

Luego de realizar los cálculos, se estima que el VAN es de ₪6,495,799.36 y el TIR es de 45.41%, lo que se traduce en un 30.41% sobre la tasa de descuento. Ambos parámetros avalan la factibilidad financiera del proyecto. Además, se calculó que el periodo de recuperación de la inversión será de 2 años y 5 meses.

## Capítulo VII: Conclusiones y recomendaciones

En este capítulo se presentan las conclusiones y recomendaciones como resultado de la realización de este proyecto.

### 7.1 Conclusiones

1. Mediante la revisión de literatura se logró confirmar que big data es una tecnología disruptiva que está cambiando múltiples industrias. Entre esos cambios está darles valor a grandes grupos de datos sin aparente relación entre ellos por medio del uso de técnicas avanzadas de análisis de datos. El resultado de ese análisis de datos puede usarse como apoyo en la toma de decisiones que eventualmente generen una ventaja competitiva.
2. A nivel mundial existen grandes esfuerzos por incorporar big data tanto a nivel privado como en el gobierno. En la actualidad, la implementación de big data está liderada por Norteamérica, seguido de Europa y Medio Oriente. La Unión Europea está trabajando en programas que incentiven el uso de big data en las organizaciones, haciendo énfasis especialmente en la formación de profesionales y en la creación de leyes que protejan la privacidad de los datos.
3. Se logró identificar que big data puede ser usado en prácticamente cualquier industria, independientemente si esté o no relacionada con la tecnología. Existen sectores que han hecho uso de análisis de datos desde su concepción, como el sector de servicios financieros y el sector de seguros. Estos han tenido que adaptarse a los cambios tecnológicos y al gran volumen y variedad de datos que tienen a disposición en la actualidad.  
Otros sectores por el contrario que han sido históricamente más lentos en adaptar cambios, como la manufactura y los servicios públicos, también se han visto beneficiados por el análisis de datos y se han visto forzados a actualizarse para mantenerse competitivos.

4. Existe una relación entre el rendimiento de las organizaciones con mejor desempeño y la toma de decisiones basadas en el análisis riguroso de datos. Además, dichas organizaciones suelen hacer uso de una perspectiva analítica a la hora de plantear estrategias futuras en la operación diaria.

Cabe destacar que las organizaciones que tengan entre sus planes hacer uso de big data para apoyar la toma de decisiones, deben contar con los perfiles profesionales adecuados, con la capacidad de discernir cuáles son los datos correctos para la organización y así lograr obtener el mejor rendimiento posible.

La calidad de los datos es más importante que la cantidad de datos, la estrategia debe estar enfocada en conseguir datos de mejor calidad y herramientas que permitan brindar un soporte para la toma de decisiones.

5. El desarrollo de la industria de software en Costa Rica ha presentado un gran crecimiento en las últimas tres décadas y se perfila hacia ser un proveedor líder en productos y servicios de tecnologías de información. Casi tres cuartas partes de las empresas del sector exportan sus servicios a Norteamérica o a Europa, regiones que a la vez lideran la implementación de big data a nivel mundial. A pesar de esto adopción de big data en nuestro país ha sido lenta y no se cuentan con estudios recientes que muestren una proyección de crecimiento.

6. Durante la elaboración de la guía de mejores prácticas se recopilieron las necesidades identificadas durante la revisión de literatura. Las organizaciones que deseen implementar big data y analytics para la toma de decisiones, deben tener claro que los objetivos de negocio deben liderar el resto de esfuerzos que se tengan en mente.

La recolección, almacenamiento y análisis de datos no puede ser ejecutados a la ligera, una debida planificación y revisión constante de los requerimientos será necesaria para alcanzar el éxito.

7. La puesta en práctica de la solución propuesta mediante un piloto para verificar su efectividad es de gran importancia. Durante la investigación se recopilan los

datos necesarios para proponer una solución, pero no es sino hasta la aplicación donde se evidencian posibles áreas de mejora y se puede obtener una retroalimentación real.

## 7.2 Recomendaciones

1. El investigador les recomienda a los directores de la empresa invertir en la formación sobre big data, tanto a nivel gerencial como operativo. A nivel gerencial se debe tener una mejor claridad de lo que se desea obtener de los datos, y al mismo tiempo debe existir un entendimiento general de los conceptos y de lo que implica la implementación.

A nivel operativo haría falta alguien con conocimientos en análisis de datos, no necesariamente alguien a nivel de un científico de datos, pero si con la suficiente experiencia como para detectar patrones en los datos.

2. Uno de los puntos en los que coincidieron los directores de la empresa en cuando a los principales impedimentos es que no se cuentan con los recursos económicos necesarios. Se le recomienda al director financiero revisar el presupuesto anual de la empresa para definir una partida específicamente para tener acceso a los datos que necesitan y que a la larga aumentarán la rentabilidad.

El presupuesto además de contemplar los accesos a las distintas bases de datos de pago, también deberá considerar la inversión necesaria para la formación continua del equipo de trabajo.

3. Se le recomienda al director de mercadeo definir una segmentación de mercado hacia el cuál se quiere dirigir con el objetivo de mejorar la lealtad y retención del cliente.

4. En cuanto a la implementación de la solución, se recomienda hacer la implementación completa a modo de prueba tomando en cuenta solo uno de los objetivos de la organización, documentando los resultados y tomar en cuenta el desarrollo iterativo para ir buscando mejores resultados.

## Capítulo VIII: Análisis retrospectivo

Big data y analytics ha sido un tema que me ha llamado la atención desde que lo escuché nombrar por primera vez cuando realizaba mi práctica profesional durante mis estudios de bachillerato de ingeniería en sistemas. Tuve la oportunidad de trabajar durante 10 meses en una empresa dedicada al desarrollo y verificación de software de aviónica, y durante uno de los cursos de entrenamiento se le dio énfasis a la importancia del análisis de datos en la industria aeronáutica.

Sin embargo, mi carrera profesional posteriormente se fue orientando más a la industria de la publicidad y el desarrollo web, donde las analíticas de datos son usadas ampliamente para seguir el comportamiento de los usuarios y según la información recopilada planear cambios o compañías futuros. Mi experiencia en este sentido está más orientado a la implementación de los sistemas de seguimiento que el análisis propio de los datos.

Cuando se nos presentaron las opciones para los temas de investigación, el que más me llamó la atención fue big data y analytics, pues me pareció una excelente oportunidad para estudiarlo más a profundidad. Según lo planteado se debía realizar un análisis de la experiencia de otros países en lo que respecta a datos no estructurados, objetivo que se logró mediante una revisión de literatura. Del mismo modo se investigó también sobre la situación actual del mercado de las tecnologías de información en Costa Rica.

Sobre esto último me pareció curioso no encontrar datos más actualizados sobre una industria que ha crecido tanto en los últimos años y que les da empleo a tantas personas. Retomando mi experiencia profesional, únicamente he tenido la oportunidad de trabajar en empresas que brindan servicios a compañías extranjeras, exportando servicios principalmente a Estados Unidos, Canadá y Europa. Con empresas de diferentes sectores, entre ellos finanzas, seguros, farmacia y automotriz.

Para esta primera etapa de la investigación fue necesario realizar una revisión de literatura. En mi opinión la charla impartida por el profesor Fulvio Lizano fue clave para poder llevar a refrescar los conocimientos adquiridos en el que fuera uno de los primeros cursos de la maestría. Muy agradecido por la dedicación de la Escuela de Informática en brindarnos las herramientas necesarias para realizar un mejor trabajo.

Justamente en la revisión de literatura, uno de los temas de investigación estaba relacionado con la adopción de big data y analytics a nivel mundial. No fue sorpresa encontrar a Norteamérica ocupando el primer lugar, pues es la región donde están basados grandes referentes de la industria. Europa, actualmente en segundo lugar, realiza un esfuerzo grande para mantenerse competitivo. Considero muy acertado que, entre los planes de esta región, se tome la formación de profesionales como uno de los puntos más importantes, debido a la escases de científicos de datos y afines que se espera en los próximos años.

Encontrar información sobre big data y analytics no fue una tarea difícil, sin embargo, si se debió filtrar la información para obtener datos relevantes. Hay aspectos de big data que están muy bien explorados, como sus características y los posibles beneficios de su implementación. No obstante, hay otros aspectos, como la influencia que tiene big data en la toma de decisiones es un campo muy poco explorado, afirmación que es respaldada por los artículos científicos consultados durante la investigación.

El tema de las mejores prácticas también ha sido algo que ha despertado mi interés, la formación de ingeniero en sistemas y mis estudios en aviación, me han enseñado que la mejor manera de hacer algo es seguir una serie de pasos previamente definidos y que han demostrado dar buenos resultados. Aunque reconozco que en algunos casos es necesario desviarse de esos pasos para obtener resultados distintos.

La realización de este proyecto de investigación me hizo valorar el aprendizaje adquirido en los diferentes cursos de la Maestría en Tecnologías de Información (MATI) y al mismo

tiempo recordar las largas horas de estudios, los consejos de los profesores y la gratificación de alcanzar un objetivo con esfuerzo.

Quedo muy agradecido con todo el personal de la MATI, quienes siempre se mostraron en la mejor disposición de ayudar. El apoyo brindado por la tutora fue esencial para guiarme durante todo el proceso de algo que era completamente desconocido para mí.

## Capítulo VIII: Bibliografía

Abraham, T. (2018). McKinsey forecasts high growth in data and analytics, more vertical integration. Recuperado 3 marzo, 2019, de <https://www.healthcarediver.com/news/mckinsey-forecasts-high-growth-in-data-and-analytics-more-vertical-integra/522522/>

Airbus (2013). Airbus and IBM to help aircraft operators optimize fleet management and operations. Recuperado noviembre, 17, 2018 de <https://www.airbus.com/newsroom/press-releases/en/2013/04/airbus-and-ibm-to-help-aircraft-operators-optimize-fleet-management-and-operations.html>

Amazon (s.f.). Bases de datos no relacionales | Bases de datos de gráficos | AWS. Recuperado noviembre, 17, 2018 de <https://aws.amazon.com/es/nosql/>

Araujo, A. (2016). ¿Qué es una Base de Datos NoSQL? Recuperado noviembre, 17, 2018 de <https://blogs.oracle.com/spain/qu-es-una-base-de-datos-nosql>

Bagci, F. (s.f.). Opportunities and Challenges of Big Data Analytics.

Barranco, R. (2012). ¿Qué es Big Data? Recuperado noviembre, 17, 2018 de <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html>

BCCR (2018). Resultados del Estudio: Exportaciones de servicios por medio de redes TIC.

BDVA (2017) Big Data Value Strategic Research and Innovation Agenda. The New Economic Asset for Europe.

Blau, J. (2012) Big Demand For Big Data Scientists in Europe. Revista Research-Technology Management. Pp 3-6.

Callil, T. (2012). Big Data is a Big Deal. Recuperado 10 febrero, 2019, de <https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal>

Campanelli, M. (2007). Dirty data is a business problem, not an IT problem, says Gartner. Recuperado febrero, 3, 2019, de <https://www.dmnews.com/data/news/13068182/dirty-data-is-a-business-problem-not-an-it-problem-says-gartner>

CAMTIC (2014). Costa Rica: Verde e Inteligente 2.0. Estrategia de desarrollo del sector digital

CAMTIC (2015). Mapeo sectorial de tecnologías digitales 2014. Unidad de investigación empresarial

Capgemini. (s.f) The Deciding Factor: Big Data & Decision Making. Economist Intelligence Unit – The Economist

Cato, S (2001). BEST PRACTICES – what does that imply? San Diego Museum of Natural History

Chen H, Chiang RH & Storey VV (2012) Business intelligence and analytics: from big data to big impact. MIS Quarterly 36(4), 1165–1188.

Columbus, L. (2017). 53% Of Companies Are Adopting Big Data Analytics. Recuperado noviembre, 19, 2018 de <https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/#7b946a3939a1>

Consultancy UK. (2018). 70% of Big Data projects in UK fail to realise full potential. Recuperado 13 febrero, 2019, de <https://www.consultancy.uk/news/16839/70-of-big-data-projects-in-uk-fail-to-realise-full-potential>

Dean, J., Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113.

Dresner Advisory Services (2017) Big Data Analytics Market Study

E-skills UK (2014) Big data analytics: assessment of demand for labour and skills 2013-2020, Big data analytics, e-skills UK, London

Engels, B. *Intereconomics* (2017) 52: 213. <https://doi.org/10.1007/s10272-017-0677-4>

European Parliament (2016). Big data and data analytics. The potential for innovation and growth

Fernando, A. (s.f.). Types of data analytics. Recuperado noviembre, 16, 2018 de [https://medium.com/@amritha\\_fernando/types-of-data-analytics-c03a6f22ede2](https://medium.com/@amritha_fernando/types-of-data-analytics-c03a6f22ede2)

Fosso Wamba, Akter, Edwards, Chopin, Gnanzou. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165(C), 234-246.

Gartner. (2011) Findings from primary research study: Data quality issues create significant cost, yet often go unmeasured.

Gartner. (2016). What Is Big Data? - Gartner IT Glossary - Big Data. Recuperado noviembre, 14, 2018 de <https://www.gartner.com/it-glossary/big-data/>

Hajirahimova, Makrufa & Aliyeva, Aybeniz. (2015). Big Data strategies of the world countries.

Halevi, G. (2012). The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature. Recuperado noviembre, 17, 2018 de <https://www.researchtrends.com/issue-30-september-2012/the-evolution-of-big-data-as-a-research-and-scientific-topic-overview-of-the-literature/>

Halper, F (2017) Seven Industries Using Analytics for Business Value. TDWI Research

IBM. (s.f.). The Four V's of Big Data. Recuperado noviembre, 18, 2018 de <https://www.ibmbigdatahub.com/infographic/four-vs-big-data>

IBM. (s.f.) What is data science? Recuperado noviembre, 18, 2018 de <https://www.ibm.com/analytics/in/en/technology/data-science/>

IBM. (2015). What is MapReduce? Recuperado noviembre, 18, 2018 de <https://www.ibm.com/analytics/hadoop/mapreduce>

IDC (2013). THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East — United States.

IDC (2017). En América Latina, habrá un incremento del 129% de gasto en Big Data y analítica para mejorar la experiencia del cliente hacia 2020: IDC Releases. Recuperado febrero, 2, 2019 de [http:// mx.idclatin.com/releases/news.aspx?id=2227](http://mx.idclatin.com/releases/news.aspx?id=2227)

I-Scoop. (2013). Big data analytics: from big data to smart decisions. Recuperado febrero, 3, 2019, de <https://www.i-scoop.eu/big-data-action-value-context/big-data-analytics-from-big-data-to-smart-data-and-decisions/>

Kim, Gang-Hoon & Trimi, Silvana & Chung, Ji-Hyong. (2014). Big Data Applications in the Government Sector: A Comparative Analysis among Leading Countries. Communications of the ACM. 57. 78-85. 10.1145/2500873.

Kenton, W. (2018). Best Practices. Recuperado 22 febrero, 2019, de [https://www.investopedia.com/terms/b/best\\_practices.asp](https://www.investopedia.com/terms/b/best_practices.asp)

Kubick, W.R. (2012). Big Data, Information and Meaning. In: Clinical Trial Insights, pp. 26–28

LaValle, S., Lesser, E., Shockley, R., Hopkins, M., & Kruschwitz, N. (2010). Big Data, Analytics and the Path From Insights to Value | MIT Sloan Management Review. Recuperado 8 febrero, 2019, de <https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/>

Madan, S. (2010). Best practices for evaluating data quality tools. Ovum.

Malvicino, F., Yoguel, G. (2015) Big Data: avances recientes a nivel internacional y perspectivas para el desarrollo local. Centro Interdisciplinario De Estudios En Ciencia, Tecnología e Innovación. Gobierno de Argentina.

M. Marjani et al., "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges," IEEE Access, vol. 5, pp. 5247-5261, 2017.  
doi: 10.1109/ACCESS.2017.2689040

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D., (2012). Big data. The management revolution. Harvard Business Review, 90(10), 61–67.

McKinsey Global Institute (2011) Big data: The next frontier for innovation, competition, and productivity

McNulty, E. (2018). SQL vs. NoSQL- What You Need to Know. Recuperado noviembre, 20, 2018 de <https://dataconomy.com/2014/07/sql-vs-nosql-need-know/>

Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L., & Canhoto, A. (2018). Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, 93, 67-78.

Normandeau, K. (2013). Beyond volume, variety and velocity is the issue of big data veracity. Recuperado noviembre, 16, 2018 de <http://inside-bigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>

Oracle. (2015). Airbus Uses Big Data Appliance to Improve Flight Testing. Recuperado noviembre, 17, 2018 de <https://video.oracle.com/detail/videos/featured-videos/video/4662924811001>

Organización de las Naciones Unidas. (s.f.). Mejores prácticas. Recuperado 23 febrero, 2019, de <http://www.fao.org/capacity-development/resources/good-practices/es/>

PitneyBowes (2011). Data Quality? That's IT's problema not mine. What Business Leaders Should Know About Data Quality

Rajaraman, V. (2016) Big Data Analytics. *Resonance*, 695-716

Rajeev Sharma, Sunil Mithas & Atreyi Kankanhalli (2014) Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations, *European Journal of Information Systems*, 23:4, 433-441, DOI: 10.1057/ejis.2014.17

Ricardo Energy & Environment (2017) Study into Deriving Transport Benefits from Big Data and the Internet of Things in Smart Cities.

Ruivo P, Oliviera T. (2017). Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value? *NOVA*

Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal

Sampieri H, Roberto (2010). Metodología de la Investigación. Editorial McGRAW – HILL INTERAMERICANA México

TDWI Research (2011) Big Data Analytics

The economist (2014) Who's big on Big Data

The economist (2012) Big data Lessons from the leaders

UK Intellectual Property Office Informatics Team (2014). Eight Great Technologies. Big Data. A patent review.

Vargas, L. (2012) La industria del software en Costa Rica un exitoso resultado desde la Universidad Pública.

Van Puyvelde, D. (2017). Beyond the buzzword: big data and national security decision-making

William Fry (2017) Europe for Big Data. An analysis of the key factors driving big data strategies.

Williamson, K., Johanson, S., G (2018). Research Methods. Information, Systems. And Contexts. Second Edition.

Wishworks (2018). The State of Big Data in the UK 2017/2018

## Glosario

**Analytics:** Data analytics (análisis de datos) es un enfoque que implica el análisis de datos (big data, en particular) para sacar conclusiones. Al usar data analytics, las empresas pueden estar mejor equipadas para tomar decisiones estratégicas y aumentar su volumen de negocios.

**Apache Spark:** sistema de computación que se basa en Hadoop Map Reduce y que, principalmente, permite dividir o paralelizar el trabajo, ya que normalmente se instala en un clúster de máquinas.

**Clúster:** conjunto de computadoras que se relacionan entre sí a través de una red de alta velocidad, actuando como una sola computadora.

**Datos crudos:** son los datos que se recogen desde una fuente. Datos crudos no han sido sujetos al procesamiento u otra manipulación; también se conocen como datos primarios.

**Datos no estructurados:** son datos binarios que no tienen estructura interna identificable. Es un conglomerado masivo y desorganizado de varios objetos que no tienen valor hasta que se identifican y almacenan de manera organizada.

**Dispositivos inteligentes:** son dispositivos electrónicos conectados a otros dispositivos o redes a través de diferentes protocolos.

**Framework:** es una estructura conceptual y tecnológica de soporte definida, normalmente con artefactos o módulos de software concretos, en base a la cual otro proyecto de software puede ser organizado y desarrollado. Típicamente, puede incluir soporte de programas, librerías y un lenguaje interpretado entre otros programas para ayudar a desarrollar y unir los diferentes componentes de un proyecto.

**Internet de las cosas (IoT):** es un sistema de dispositivos de computación interrelacionados, máquinas mecánicas y digitales, objetos, animales o personas que tienen identificadores únicos y la capacidad de transferir datos a través de una red, sin requerir de interacciones humano a humano o humano a computadora.

**Metadatos:** son datos que se usan para describir otros datos. Los metadatos consisten en información que caracteriza datos, describen el contenido, calidad, condiciones, historia, disponibilidad y otras características de los datos.

**M2M (machine to machine):** es la comunicación entre máquinas (llámese un dispositivo electrónico, robot, automóvil o cualquier otro objeto que no sea una persona).

## Anexos

### Anexo 1 – Cuestionario sobre la organización

#### Sección 1 – Información sobre la organización

##### 1.1 Datos generales

Sector	
<input type="checkbox"/> Agricultura	<input type="checkbox"/> Transporte
<input type="checkbox"/> Manufactura	<input type="checkbox"/> Servicios de salud
<input type="checkbox"/> Entretenimiento	<input type="checkbox"/> Servicios financieros
<input type="checkbox"/> Telecomunicaciones	<input checked="" type="checkbox"/> Ventas
<input type="checkbox"/> Energía	<input type="checkbox"/> Educación
<input type="checkbox"/> Sector público	<input type="checkbox"/> Investigación
<input type="checkbox"/> Otro:	

Cantidad de empleados	
<input type="checkbox"/> 1-5	
<input checked="" type="checkbox"/> 6-29	
<input type="checkbox"/> 30-99	
<input type="checkbox"/> 100-999	
<input type="checkbox"/> Más de 1,000	

##### 1.2 ¿Cuáles son los principales objetivos del negocio?

El principal objetivo de negocio es posicionar a la empresa entre los principales cinco proveedores de iluminación en Costa Rica.

##### 1.3 ¿Cuenta con una estrategia y plan de negocios? En caso afirmativo por favor elaborar

Si     No

Datos adicionales:

La estrategia de negocios está enfocada en proyectos, tanto del sector público como privado.

Se trabaja con un inventario mínimo en el país y las compras se hacen bajo pedido especial.

Activamente se buscan reuniones con desarrolladores de proyectos privados.

En el sector público se participa en licitaciones relacionadas con iluminación, tanto de construcción como de vía pública.

#### 1.4 ¿Cuáles son los principios de negocio, objetivos y motivadores?

Darle una solución en iluminación al cliente

Adaptarse a las necesidades del cliente y no que el cliente deba adaptarse solo a lo que el mercado ofrece. Esto se resume en brindar un servicio personalizado.

Poder llegar al encargado de diseño de proyectos con la finalidad de tener conocimiento precio ve las especificaciones técnicas en cuanto a iluminación.

## Sección 2 – Generalidades de big data

### 2.1 ¿Está familiarizado con big data?

Si     Medianamente     No

### 2.2 ¿Qué es para usted Big data?

Cualquier base de datos que tenga gran volumen y cumpla con cierta estructura.

### 2.3 ¿Qué es para usted el análisis de datos?

Comprender de qué forma se utilizan los datos para extraer información

### Sección 3 – Big data en la organización

#### 3.1 ¿Cuenta la organización con una estrategia de big data o analytics?

Sí  No

#### 3.2 ¿De cuales fuentes de datos su organización recolecta, o espera recolectar información?

*Seleccionar sola una opción por fila.*

	Actualmente	De 1 a 5 años	Más de 5 años	No está en planes	Desconocido
Registros	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Transacciones	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Eventos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Correos electrónicos	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Redes sociales	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sensores	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
IoT	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Datos públicos	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uso de teléfonos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Datos RFID	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Datos de POS	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Datos geo espaciales	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Textos en general	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Audio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Fotografías	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Videos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Otros: Datos de motores de búsqueda	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3.3 ¿Qué tipo o tipos de datos considera su organización que son relevantes, pero aún no han podido aprovechar?

Datos de licitaciones, en el sector público existen bases de datos como Connexis Leads, que tiene un costo mensual de uso, y también datos del SICOP, que son gratuitos, pero necesitan interpretación.

La Cámara Costarricense de la Construcción tiene datos del sector privado. Llevan un listado de las empresas que solicita permisos de construcción, encargados y nombre del proyecto.

A nivel de importación, datos de cuáles son las líneas de producto que más se están importando (tendencia del mercado). PROCOMER y el Ministerio de Hacienda manejan esta información (partidas arancelarias por mes).

3.3 ¿Cuáles son los principales impedimentos para tener acceso a los datos que son relevantes para su organización?

Presupuesto para comprar datos de diversas plataformas

Falta de acceso a información (ESPH no provee datos).

Algunas fuentes no son accesibles (por confidencialidad)

Tiempo

Recursos

## Sección 4 – Cultura organizacional

4.1 ¿Cuenta la organización el conocimiento necesario para manejar big data?

Si     No

4.2 ¿A nivel de gerencia se tiene conocimiento sobre big data?

Si     No

4.3 ¿La toma de decisiones basada en datos es parte de la cultura organizacional?

Si     No

4.4 Comentarios adicionales

Si exista toma de decisiones basada en datos, pero se hace de manera empírica, no se hace un análisis como tal.

## Sección 5 – Infraestructura

4.1 La recolección de datos en la organización es:

- Propia
- Sub-contratada
- Ambas
- Aún no se recolectan datos

4.2 El análisis de datos en la organización es:

- Propio
- Sub-contratado
- Ambos
- Aún no se analizan datos

#### 4.3 ¿La organización comparte datos con terceros (clientes, proveedores, gobierno)?

Si     No     No aplica

Se comparten datos con proveedores, más puntualmente sobre licitaciones.

### Sección 5 – Innovación con Big Data

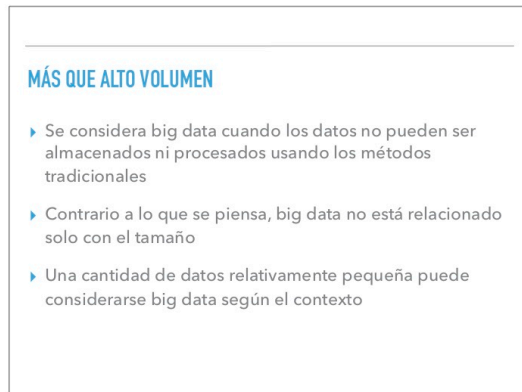
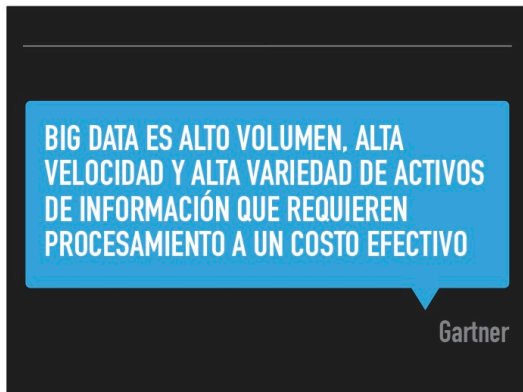
#### 5.1 ¿De qué manera big data genera o se espera que genere valor en la organización?

*Seleccionar sola una opción por fila.*

	Actualmente	Dentro de 1 a 5 años	Dentro de más de 5 años	No se crea valor	Desconocido
Mejorar la segmentación de los clientes	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mejorar la lealtad y retención de clientes	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mejorar la eficiencia	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mejorar el diseño de producto	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Manejo de riesgo	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Manejo de finanzas	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mejorar el modelo de negocios	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Otros:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

#### 5.2 ¿Qué espera obtener de la implementación de big data?

Conocer mejor el mercado, enfocar los esfuerzos de una forma más eficiente y mejor manejo de inventario



---

### PRINCIPALES CARACTERÍSTICAS

- ▶ Volumen (dimensión de los datos)
- ▶ Variedad (diferentes formas de datos)
- ▶ Velocidad (análisis de transmisión de datos)
- ▶ Veracidad (incertidumbre de los datos)

---

### PRINCIPALES RETOS

- ▶ Más datos = más espacio de almacenamiento
- ▶ Los datos se generan más rápido
- ▶ Necesidad de manejar varias estructuras de datos
- ▶ Requerimientos de negocios ágiles

---

### CREACIÓN DE VALOR

- ▶ Aumento de productividad
- ▶ Ventaja competitiva

¿QUÉ ES

---

# ANALYTICS?

ES EL PROCESO DE EXAMINAR SETS DE DATOS CON EL OBJETIVO DE EXTRAER CONOCIMIENTO DEL CUAL SE PUEDA TOMAR ACCIONES

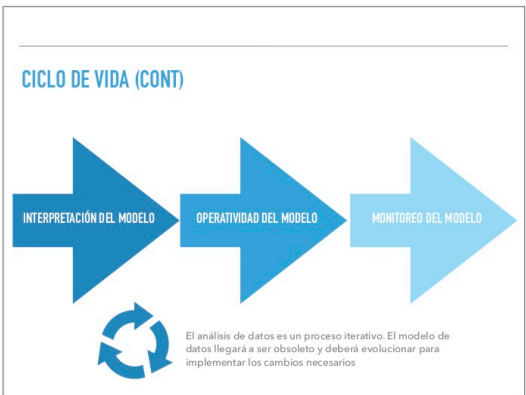
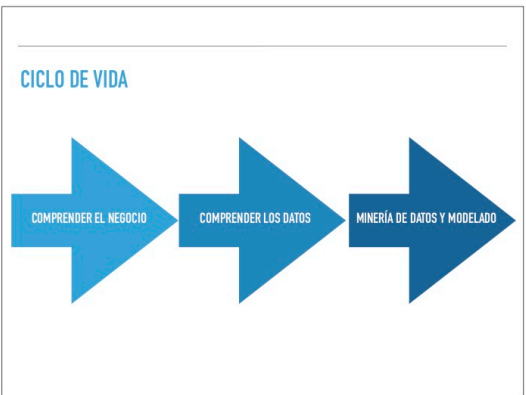


**ANÁLISIS DE BIG DATA**

- ▶ Uso de técnicas avanzadas de análisis para trabajar sobre grandes sets de datos
- ▶ Los sets de datos por si solos no representan conocimiento

**DIFERENTES TIPOS**

- ▶ Descriptivo
- ▶ Descubrimiento o diagnóstico
- ▶ Predictivo
- ▶ Prescriptivo



## Anexo 3 – Presentación: Conceptos de implementación iterativa



---

### ¿QUÉ ES?

- ▶ Enfoque de desarrollo que permite la convergencia de una solución con las metas deseadas a través de sucesivos perfeccionamientos a los requerimientos y los artefactos de desarrollo
- ▶ Normalmente es un concepto de desarrollo de software
- ▶ Las iteraciones son ciclos de perfeccionamiento

---

### ¿QUÉ NO ES?

- ▶ El término *iteración* a menudo se usa de manera errónea como sinónimo de *incremento*
- ▶ Si bien los enfoques iterativos se pueden usar con modelos de entregas incrementales, ambos términos no son sinónimos

---

### PERFECCIONAMIENTO

- ▶ Una iteración no necesariamente incluye todas las etapas de un proyecto
- ▶ Es posible tener iteraciones en etapas específicas
  - ▶ Como por ejemplo, iteraciones de análisis e iteraciones de captura de datos

## Anexo 4 – Cuestionario de valoración de aplicación del piloto

Respuesta del director de operaciones y director de mercadeo de Commercial Solutions.

### 1 ¿Cuál es la percepción de la organización de Big Data y Analytics?

Big data es un término que nos ha llamado mucho la atención, mas nunca nos habíamos tomado el tiempo para explorarlo a profundidad. Teníamos la percepción que por los altos costos era algo que solo estaba al alcance de grandes compañías.

Ahora tenemos un mejor panorama de big data y lo que se necesita para analizar los datos y obtener un valor real de éstos. Los objetivos que tiene nuestro negocio pueden ser potenciados y alcanzados si logramos definir una estrategia de datos eficiente.

### 2 ¿Qué cambios considera necesarios en la organización para poner en marcha el proyecto de implementación?

La compañía aún cuenta con recursos limitados para invertir, el capital se compone de aporte de los socios y también de las utilidades generadas por ventas. El presupuesto ha estado orientado a la investigación de los proveedores, asistiendo a ferias mundiales de iluminación en diferentes países y formando una relación con fabricantes y representantes.

Estamos llegando a un punto donde necesitamos conocer mejor el mercado local y consideramos que es prudente asignar un mayor presupuesto al área de tecnologías de información para así poder invertir en infraestructura, ya sea local o servicios de datos en la nube, según lo que se adapte más a nuestras necesidades.

### 3 ¿Cuál es la percepción de la guía de mejores prácticas en general?

En la industria de la iluminación existen diferentes estándares y mejores prácticas, por lo que tenemos experiencia siguiendo procedimientos o requerimientos, en especial cuando se trata de licitaciones públicas. Hemos tenido incluso que aplicar pruebas

destructivas a los productos que ofrecemos para demostrar el cumplimiento con ciertas normas internacionales.

La guía nos parece de gran utilidad, sin embargo, consideramos que se necesita un conocimiento más amplio en el tema de big data antes de tan siquiera pensar en adentrarse en un proyecto de este tipo.

#### 4 ¿Cómo beneficiaría a la organización la aplicación de la guía de mejores prácticas?

Sin lugar a dudas necesitamos una manera de guiar el proceso de implementación de big data en nuestra empresa. En nuestro caso cada etapa representaba un reto nuevo, y una vez finalizada teníamos una mejor percepción de cómo orientar big data hacia los objetivos que tenemos como organización.

#### 5 ¿Cuál es la opinión sobre la aplicación del piloto?

Nos hubiera gustado tener más tiempo para ver algunos temas más a fondo y también poder hacer algunas pruebas con datos reales. Sin embargo, somos conscientes de las falencias de nuestra organización en cuanto a conocimiento y estructura para hacerle frente a un proyecto de big data en este momento.

## Índice de tablas

Tabla 1 - Tabla comparativa entre bases de datos SQL y NoSQL. Fuente: McNulty, E (2014) & Amazon AWS .....	20
Tabla 2 - Diferencia entre sector público y privado. Fuente Kim et al. 2014 .....	46
Tabla 3 - Participantes en el taller de capacitación .....	78
Tabla 4 - Documentación de definiciones para el proyecto .....	80
Tabla 5 Participantes en la sesión de alineación de requerimientos.....	81
Tabla 6 - Participantes en la session de requerimientos de big data .....	82
Tabla 7 - Implementación de datos en el tiempo .....	83
Tabla 8 - Participantes en la session de definición de estrategia de big data.....	84
Tabla 9 - Participantes en el taller de implementación iterativa .....	87
Tabla 10 - Participantes en la session de trabajo de impacto de big data .....	88
Tabla 11 – Detalle de la inversion inicial. Fuente: Elaboración propia .....	92
Tabla 12 – Proyección del flujo de efectivo .....	93
Tabla 13 – Detalle del análisis financiero .....	94

## Índice de figuras

Figura 1 - Diferentes términos usados para describir lo que sería el análisis de big data. Fuente: TDWI, 2011 .....	6
Figura 2 - Porcentaje de adopción en organizaciones según región geográfica. Fuente: Dresner Advisory Services, 2017 .....	11
Figura 3 – Cantidad de datos útiles en comparación con los datos generados, expresados en Exabytes (EB). .....	35
Figura 4 - Uso de big data en Europa según sector. Fuente: William Fry (2017) .....	40
Figura 5 – Distribución de empresas según clasificación de CAMTIC. Fuente: CAMTIC, 2014 .....	52
Figura 6 - Exportación de servicios TIC por país o región de destino. Fuente BCCR 2018 .....	53