



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**PROPUESTA DE METODOLOGÍA DE IMPLEMENTACIÓN DE BIG DATA:
PERSPECTIVA GERENCIAL GUATEMALTECA**

Carlos Antonio Guzmán Coroy

Asesorado por el Ing. Oscar Alejandro Paz Campos

Guatemala, marzo de 2019

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**PROPUESTA DE METODOLOGÍA DE IMPLEMENTACIÓN DE BIG DATA:
PERSPECTIVA GERENCIAL GUATEMALTECA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA

POR

CARLOS ANTONIO GUZMÁN COROY

ASESORADO POR EL ING. OSCAR ALEJANDRO PAZ CAMPOS

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, MARZO DE 2019

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Pedro Antonio Aguilar Polanco
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Luis Diego Aguilar Ralón
VOCAL V	Br. Christian Daniel Estrada Santizo
SECRETARIA	Inga. Lesbia Magalí Herrera López

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Murphy Olympto Paiz Recinos
EXAMINADOR	Ing. Roberto Estuardo Ruiz Cruz
EXAMINADOR	Ing. Marlon Francisco Orellana López
EXAMINADOR	Ing. César Augusto Fernández Cáceres
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

PROPUESTA DE METODOLOGÍA DE IMPLEMENTACIÓN DE BIG DATA: PERSPECTIVA GERENCIAL GUATEMALTECA

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha 31 de octubre de 2017.

Carlos Antonio Guzmán Coroy

Guatemala, 20 de febrero de 2018

Ing. Carlos Azurdia
Coordinador Proyectos de Graduación
Escuela de Ciencias y Sistemas
Facultad de Ingeniería

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante **CARLOS ANTONIO GUZMÁN COROY**, con carné 200010454 titulado: "**PROPUESTA DE METODOLOGÍA DE IMPLEMENTACIÓN DE BIG DATA: PERSPECTIVA GERENCIAL GUATEMALTECA**", y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo propuesto, por lo que firmo la presente para que proceda con los trámites correspondientes.

Sin otro particular, me suscribo de usted.

Atentamente,



ING. OSCAR ALEJANDRO PAZ CAMPOS
COLEGIADO No. 6430

Ing. Oscar Alejandro Paz Campos
Asesor y revisor de Trabajo de Graduación



Universidad San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 28 de Febrero de 2018

Ingeniero
Marlon Antonio Pérez Türk
Director de la Escuela de Ingeniería
En Ciencias y Sistemas

Respetable Ingeniero Pérez:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante **CARLOS ANTONIO GUZMÁN COROY** con carné **200010454** y CUI **2499 23343 0101**, titulado "**PROPUESTA DE METODOLOGÍA DE IMPLEMENTACIÓN DE BIG DATA: PERSPECTIVA GERENCIAL GUATEMALTECA**" y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo aprobado.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,



Ing. Carlos Alfredo Azurdia
Coordinador de Privados
y Revisión de Trabajos de Graduación



UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA



FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA EN
CIENCIAS Y SISTEMAS
TEL: 24767644

*El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del revisor y del Licenciado en Letras, del trabajo de graduación **“PROPUESTA DE METODOLOGÍA DE IMPLEMENTACIÓN DE BIG DATA: PERSPECTIVA GERENCIAL GUATEMALTECA”**, realizado por el estudiante, CARLOS ANTONIO GUZMÁN COROY aprueba el presente trabajo y solicita la autorización del mismo.*

“ID Y ENSEÑAR A TODOS”

Ing. Martín Antonio Pérez Türk
Director

Escuela de Ingeniería en Ciencias y Sistemas



Guatemala, 28 de marzo de 2019

Universidad de San Carlos
De Guatemala



Facultad de Ingeniería
Decanato

Ref. DTG.169.2019

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas del trabajo de graduación titulado: **"PROPUESTA DE METODOLOGÍA DE IMPLEMENTACIÓN DE BIG DATA: PERSPECTIVA GERENCIAL GUATEMALTECA"** presentado por el estudiante: **Carlos Antonio Guzmán Coroy** después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, se autoriza la impresión del mismo.

IMPRÍMASE.

Ing. Pedro Antonio Aguilar Polanco
Decano



Guatemala, Marzo de 2019

/echm

ACTO QUE DEDICO A:

Dios	Por ser y estar siempre presente en mi vida.
Mi madre	Antonia Leticia Coroy Cárcamo, por todo su amor, sacrificios y esfuerzos dedicados a mi formación como persona.
Mi segunda madre	Milagro de Jesús Cárcamo Reyes, por su esmero y cuidados en toda mi vida.
Mi familia	Por su apoyo, presencia y cariño.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por brindarme una formación académica con altas calidades para mi desempeño laboral.
Facultad de Ingeniería	Por ser una casa durante los años de formación académica.
Mis compañeros de estudios en la carrera	Por su apoyo, trabajo conjunto, sacrificios compartidos y experiencias vividas.
Catedráticos de la Escuela de Ingeniería en Ciencias y Sistemas	Por compartir su experiencia y transmitir el conocimiento asimilado durante la carrera.
Ingenieros en Ciencias y Sistemas cercanos	Por su confianza en mis capacidades, su apoyo y motivación para concluir todo el proceso académico.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	VII
LISTA DE SÍMBOLOS	IX
GLOSARIO	XI
RESUMEN.....	XVII
OBJETIVOS.....	XIX
INTRODUCCIÓN.....	XXI
1. BIG DATA EN LAS ORGANIZACIONES EMPRESARIALES	1
1.1. Los ejecutivos de negocios y Big Data	2
1.1.1. Comprensión de Big Data por los ejecutivos	2
1.1.2. Situación de implementación de Big Data.....	3
1.1.2.1. Ámbito mundial.....	3
1.1.2.2. Ámbito Estados Unidos	4
1.1.2.3. Ámbito América Latina.....	6
1.1.2.4. Ámbito Guatemala.....	7
1.2. ¿Qué tomar en cuenta?	8
1.2.1. Digitalización de los datos.....	8
1.2.2. Inteligencia de negocios.....	8
1.2.3. Minería de datos	9
1.2.3.1. Agrupamiento de datos.....	10
1.2.3.2. Redes neuronales.....	10
1.2.3.3. Regresión	11
1.2.3.4. Árboles de decisión	11
1.2.4. Análisis de datos.....	12
1.2.4.1. Análisis descriptivo	12

1.2.4.2.	Análisis de diagnóstico.....	13
1.2.4.3.	Análisis predictivo.....	13
1.2.4.4.	Análisis prescriptivo.....	13
1.2.5.	KPI.....	14
1.3.	El valor de los datos	14
1.4.	Big Data y la estrategia empresarial.....	16
1.5.	Tendencias de estrategias de datos en las organizaciones	17
1.5.1.	Empresa Data Driven Company	18
1.5.2.	Ejecutivo 4.0	20
1.5.3.	Técnicas de análisis de datos no convencionales	25
2.	CONCEPTOS TÉCNICOS GENERALES DE BIG DATA	27
2.1.	Definiendo Big Data.....	27
2.1.1.	Síntesis cronológica.....	27
2.1.2.	Definiciones	29
2.1.3.	Características: las V de Big Data	31
2.1.3.1.	Volumen	31
2.1.3.2.	Velocidad.....	32
2.1.3.3.	Variedad.....	32
2.1.3.4.	Variabilidad.....	33
2.1.3.5.	Veracidad	33
2.1.3.6.	Valor	33
2.1.3.7.	Visualización	34
2.1.3.8.	Vaguedad.....	34
2.1.3.9.	Validez.....	34
2.1.3.10.	Vocabulario	34
2.1.3.11.	Lugar	34
2.1.3.12.	Otras características	35
2.1.4.	Capacidades de Big Data	35

	2.1.4.1.	Almacenamiento	35
	2.1.4.2.	Procesamiento.....	35
	2.1.4.3.	Análisis	36
2.2.		Tendencias tecnológicas que demandan los clientes	36
	2.2.1.	Movilidad.....	36
	2.2.2.	Cloud.....	36
	2.2.3.	Social media	37
	2.2.4.	Análisis en tiempo real.....	37
2.3.		Big Data 2.0	37
	2.3.1.	Tecnologías asociadas	38
	2.3.1.1.	<i>Internet of things (IoT)</i>	39
	2.3.1.2.	<i>Machine learning</i>	39
	2.3.1.3.	<i>Deep learning</i>	39
	2.3.1.4.	<i>Complex event processing</i>	40
	2.3.2.	Tecnologías en fase de crecimiento	40
	2.3.2.1.	Cifrado y enmascaramiento de datos ...	40
	2.3.2.2.	Integración de datos	40
	2.3.2.3.	Virtualización de datos.....	41
	2.3.2.4.	Almacenes de archivos distribuidos.....	41
	2.3.2.5.	Datos en memoria	41
	2.3.2.6.	Monitoreo y administración	42
	2.3.2.7.	Base de datos NoSQL	42
	2.3.2.8.	Análisis predictivo	42
	2.3.2.9.	Búsqueda y descubrimiento de conocimiento.....	43
	2.3.2.10.	Análisis de flujo.....	43
2.4.		<i>Smart data</i>	43

3.	PROPUESTA DE METODOLOGÍA DE IMPLEMENTACIÓN DE BIG DATA EN PLANES ESTRATÉGICOS EMPRESARIALES	45
3.1.	Condiciones iniciales	47
3.1.1.	Pensamiento estratégico	49
3.1.1.1.	Nivel 1: a la deriva	49
3.1.1.2.	Nivel 2: desenfoque.....	50
3.1.1.3.	Nivel 3: aspiración	50
3.1.1.4.	Nivel 4: limitación	50
3.1.1.5.	Nivel 5: propuesta de valor consistente	51
3.1.2.	Enfoque en el cliente	51
3.1.2.1.	Nivel 1: indiferencia	51
3.1.2.2.	Nivel 2: diferenciación	52
3.1.2.3.	Nivel 3: relación establecida.....	52
3.1.2.4.	Nivel 4: acogida.....	52
3.1.2.5.	Nivel 5: dominio.....	53
3.1.3.	Enfoque en los datos	53
3.1.3.1.	Nivel 1: resistencia	53
3.1.3.2.	Nivel 2: consciencia.....	53
3.1.3.3.	Nivel 3: análisis	54
3.1.3.4.	Nivel 4: comprensión estratégica	54
3.1.3.5.	Nivel 5: dirección por datos	54
3.1.4.	Infraestructura de tecnologías de información (TI) ...	54
3.1.4.1.	Nivel 1: inicial	55
3.1.4.2.	Nivel 2: aplicaciones de negocio	55
3.1.4.3.	Nivel 3: gestión de procesos	55
3.1.4.4.	Nivel 4: gestión del negocio.....	55
3.1.4.5.	Nivel 5: innovación	56
3.1.5.	Transformación digital	56
3.1.5.1.	Nivel 1: inmovilismo.....	56

	3.1.5.2.	Nivel 2: activación.....	56
	3.1.5.3.	Nivel 3: proyecto	57
	3.1.5.4.	Nivel 4: estrategia	57
	3.1.5.5.	Nivel 5: innovación.....	57
3.1.6.		Método de análisis	57
	3.1.6.1.	Nivel 1: carencia	58
	3.1.6.2.	Nivel 2: descripción.....	58
	3.1.6.3.	Nivel 3: diagnóstico.....	58
	3.1.6.4.	Nivel 4: predicción	58
	3.1.6.5.	Nivel 5: análisis prescriptivo.....	58
3.1.7.		Fuente de datos	59
	3.1.7.1.	Nivel 1: datos registrados	59
	3.1.7.2.	Nivel 2: datos no registrados	59
	3.1.7.3.	Nivel 3: datos de terceros	59
	3.1.7.4.	Nivel 4: datos abiertos	60
	3.1.7.5.	Nivel 5: datos no tradicionales.....	60
3.2.		Determinación de acciones.....	60
	3.2.1.	Características	61
	3.2.2.	Oportunidades	63
3.3.		Implementación de acciones	65
3.4.		Verificación de logro	65
3.5.		Acciones de mejora en la implementación.....	68
3.6.		Responsable de Big Data	68
3.7.		Ética empresarial	69
4.		PROPUESTA DE ARQUITECTURA MÍNIMA PARA IMPLEMENTAR BIG DATA.....	71
	4.1.	Diseño de arquitectura.....	71
		4.1.1. Nivel Inicial: activación de Big Data	73

4.1.1.1.	Adquisición del dato	74
4.1.1.2.	Generación de información	75
4.1.1.3.	Visualización y análisis.....	76
4.1.2.	Nivel Básico: Big Data no administrado.....	76
4.1.2.1.	Adquisición del dato	78
4.1.2.2.	Generación de información	79
4.1.2.3.	Visualización y análisis.....	80
4.1.3.	Nivel maduro: Big Data administrado	80
4.1.3.1.	Adquisición del dato	82
4.1.3.2.	Generación de información	82
4.1.3.3.	Visualización y análisis.....	82
CONCLUSIONES.....		85
RECOMENDACIONES		87
BIBLIOGRAFÍA.....		89

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Descripción de Big Data por profesionales	2
2.	Proceso implementación big data en planes estratégicos.....	46
3.	Niveles de madurez en ámbitos relacionados con la transformación de los datos en valor	47
4.	Modelo de madurez Big Data.....	49
5.	Características de las acciones para implementar Big Data	63
6.	Estados de datos en proceso para extraer valor estratégico	71
7.	Proceso para extraer valor estratégico de los datos	72
8.	Activación de Big Data	74
9.	Big Data no administrado	77
10.	Big Data administrado	81

TABLAS

I.	Identificación de distintos niveles de madurez en distintas áreas de revisión en modelo de madurez de Big Data.....	48
II.	Ejemplos de indicadores a evaluar oportunidad de mejora Big Data	66

LISTA DE SÍMBOLOS

Símbolo	Significado
EB	<i>Exabyte</i>
PT	<i>Petabyte</i>
TB	<i>Terabyte</i>
ZB	<i>Zettabyte</i>

GLOSARIO

BI	<i>Business Intelligence</i> (BI) incluye las aplicaciones, la infraestructura, las herramientas y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar decisiones.
Cadena de valor	Modo en que se desarrollan las acciones y actividades de una empresa para generar valor al cliente final.
<i>Cloud computing</i>	Computación en la nube relacionado a estilo de computación en el cual las capacidades escalables y elásticas habilitadas por TI se entregan como un servicio usando tecnologías de internet.
<i>Clustering</i>	Capacidad de definir recursos en uno o más sistemas de rango medio interconectados de manera transparente para usuarios y aplicaciones dentro del grupo especificado de sistemas débilmente acoplados en una red de área local o metropolitana.
CRM	<i>Customer relationship management</i> (CRM) es una estrategia comercial que optimiza los ingresos y la rentabilidad a la vez que promueve la satisfacción y lealtad del cliente, haciendo uso de tecnologías que permiten realizar dicha estrategia e identifican y

administran las relaciones con los clientes en persona o virtualmente.

Data mining

Proceso de descubrimiento de correlaciones, patrones y tendencias mediante el filtrado de grandes cantidades de datos almacenados en repositorios, empleando tecnologías de reconocimiento de patrones, así como técnicas estadísticas y matemáticas.

Data science

Campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados, lo cual es una continuación de algunos campos de análisis de datos como la estadística, minería de datos, aprendizaje automático y la analítica predictiva.

DRAM

Dynamic random access memory (DRAM) es un tipo de memoria dinámica de acceso aleatorio que se usa principalmente en los módulos de memoria RAM y en otros dispositivos, como memoria principal del sistema, denominada dinámica ya que para mantener almacenado un dato, se requiere revisar el mismo y recargarlo cada cierto período en un ciclo de refresco.

ERP

Enterprise resource planning (ERP) se define como la capacidad de entregar un conjunto integrado de

aplicaciones comerciales que comparten un proceso común y un modelo de datos que abarca procesos operativos amplios y profundos de extremo a extremo, como los que se encuentran en finanzas, recursos humanos, distribución, fabricación, servicios y la cadena de suministro.

Exabyte

Unidad de medida de almacenamiento de datos cuyo símbolo es el EB. Equivale a 10^{18} bytes.

Flash memory

Tipo de memoria desarrollada a partir de la memoria EEPROM, la cual permite la lectura y escritura de múltiples posiciones de memoria en la misma operación, mediante impulsos eléctricos que permiten velocidades de funcionamiento muy superiores frente a la tecnología EEPROM primigenia.

Fortune 1 000

Es una lista anual mantenida por la popular revista fortune, de las 1 000 empresas estadounidenses más grandes según los ingresos generados por las operaciones principales, las operaciones con descuento y las subsidiarias consolidadas de estas.

GPS

Global positioning system (GPS) es una tecnología de posicionamiento global para determinar la posición de un dispositivo móvil, utilizando de dos a seis de 24 satélites con un alto nivel de precisión, de solo unos pocos metros.

laaS

Infrastructure as a service (laaS) es una oferta estandarizada y altamente automatizada, donde los recursos informáticos, complementados por las capacidades de almacenamiento y de red, son propiedad de un proveedor de servicios y son hospedados por ellos y ofrecidos a los clientes a pedido, pudiendo los clientes autoabastecerse de esta infraestructura, utilizando una interfaz gráfica de usuario basada en la web que sirve como una consola de administración de operaciones de TI para el entorno general.

Insight

Clave o información esencial que permite encontrar u orientas la solución a un problema.

KPI

Key performance indicator (KPI) es una medida de alto nivel, simplificado para la recopilación y revisión en una temporalidad definida, los cuales pueden combinarse con medidas de costo para construir sistemas de métricas operativas clave.

MIPYME

Acrónimo de micro, pequeña y mediana empresa.

PaaS

Platform as a Service (PaaS) es una oferta de plataforma como servicio con una amplia colección de servicios de infraestructura de aplicaciones (*middleware*) (incluida plataforma de aplicaciones, integración, gestión de procesos comerciales) y servicios de bases de datos.

Petabyte	Unidad de almacenamiento de información cuyo símbolo es PB, y equivale a 10^{15} bytes.
PYME	Acrónimo de pequeña y mediana empresa determinadas por volumen de ingresos, valor del patrimonio y número de trabajadores.
ROI	<i>Return on investment</i> (ROI) es un índice financiero que mide y compara el beneficio o la utilidad obtenida en relación a la inversión realizada.
SaaS	<i>Software as a service</i> (SaaS) software que es propiedad, entregado y administrado de forma remota por uno o más proveedores, basados en un conjunto de definiciones de datos y códigos comunes que todos los clientes contratados consumen en un modelo de uno a muchos en cualquier momento, en base al pago por uso o como una suscripción basada en las métricas de uso.
Self service	Término que traducido al español significa autoservicio, relacionado con la práctica de servirse a uno mismo, generalmente en la adquisición de productos o generación de insumos.
SSD	Acrónimo inglés de <i>solid-state drive</i> el cual es un tipo de dispositivo de almacenamiento de datos que utiliza memoria no volátil, como memoria flash para

almacenar datos en lugar de los platos o discos magnéticos de las unidades de discos duros (HDD) convencionales.

Streaming

Técnica que admite la transmisión continua y unidireccional de datos no estructurados como audio o video a través de internet y más recientemente a través de una red móvil.

Terabyte

Unidad de medida informática cuyo símbolo es el TB, y es equivalente a 2^{40} bytes.

TI

Acrónimo de tecnología de la información el cual es la aplicación de ordenadores y equipos de telecomunicación para almacenar, recuperar, transmitir y manipular datos, con frecuencia utilizado en el contexto de los negocios u otras empresas.

Zettabyte

Unidad de almacenamiento de información cuyo símbolo es el ZB, equivale a 10^{21} bytes.

RESUMEN

El crecimiento exponencial de datos que es generado en cada instante, día a día incrementa oportunidades de negocio para las empresas que logran hacer uso de esos datos y obtener información para generar valor estratégico, por lo que es importante que los ejecutivos tengan una adecuada comprensión del uso que se puede dar a los datos para el incremento de nivel competitivo de sus empresas, lo cual se logra mediante la implementación de big data en planes estratégicos.

La situación de implementación de Big Data en diversos ámbitos permite a los ejecutivos situar su empresa sobre su situación en cuanto al aprovechamiento de los datos conociendo los procesos que hoy en día se realizan para el efecto y las tendencias que han surgido.

Big Data es un concepto que ha ido evolucionando de una forma no controlada, pero a un ritmo acelerado, por lo que para su comprensión se ofrece un recorrido a manera de alfabetización digital, desde una síntesis cronológica de su evolución, definiciones y características relacionadas, hasta una descripción de tendencias tecnológicas que han surgido en torno a su evolución.

Como propuesta para que la tecnología de Big Data sea implementada en mayor medida en empresas de Guatemala, se reconoce el alto impacto en la economía que aportan las MIPYMES por lo que para que puedan acceder a mayores ventajas competitivas se ofrece una metodología de implementación de Big Data en planes estratégicos y una propuesta de arquitectura mínima para su implementación, dando prioridad a la generación de valor de los datos,

aprovechando la variedad de datos que existen hoy en día y el bajo costo que ofrecen la tecnologías de información en entornos *cloud*.

OBJETIVOS

General

Generar una propuesta de metodología de implementación de Big Data que oriente de forma técnica a los gerentes de Guatemala a utilizar los beneficios de Big Data en la implementación de sus planes estratégicos.

Específicos

1. Definir los beneficios del uso del Big Data para la toma de decisiones mediante el aprovechamiento del valor de los datos.
2. Proponer una metodología de implementación de Big Data en planes estratégicos empresariales.
3. Definir una arquitectura tecnológica mínima para implementación de Big Data.

INTRODUCCIÓN

Hoy en día se encuentra diversidad de información sobre tecnologías emergentes a partir de Big Data, el cual ofrece una oportunidad de mejora empresarial desde su concepción, pero se reconoce que muy pocos ejecutivos tomadores de decisiones tienen claridad sobre lo que es y los mecanismos que existen para su implementación pareciendo tener una brecha para entre las empresas que hacen uso de los beneficios de Big Data, limitándose a grandes corporaciones, a lo cual es importante brindar una opción de oportunidad de implementación a las pequeñas y medianas empresas.

La tecnología detrás de Big Data tiene un amplio espectro de componentes, oferta y diversidad de configuración que puede llegar a volver complejos los intentos de su implementación, por lo que se considera que existe la posibilidad de implementar Big Data en PYMES, mediante la clarificación de lo que es Big Data, una estrategia empresarial sólida para su implementación mediante una metodología adecuada y el conocimiento de opciones tecnológicas de bajo costo accesibles y de implementación ágil.

Para establecer una definición de Big Data se hace necesario un recorrido histórico de su evolución y una revisión de su concepción por parte de reconocidas consultoras de tecnología como gartner, forrester, IDG y researchgate, entre otras, así como la oferta tecnológica que se tiene a la fecha por parte de los grandes proveedores como microsoft, google y amazon, determinando un concepto propio que pueda ser comprendido en el contexto guatemalteco.

Para que los ejecutivos de PYMES en Guatemala puedan conocer los beneficios de Big Data, se plantea un marco situacional que aborda la concepción y uso en diversos ámbitos, un resumen de lo que hoy en día se está haciendo para aprovechar los datos, y la exposición de tendencias que tanto las empresas como los ejecutivos siguen para mejorar la competitividad de las empresas mediante la transformación en valor estratégico de los datos.

Dada la diversidad de conceptos que existe detrás de Big Data se realiza una exposición de conceptos a manera de alfabetización digital que pueda poner en sintonía a los ejecutivos sobre la tecnología y beneficios en torno a Big Data, contemplando tendencias tecnológicas que van surgiendo.

Como propuesta de metodología de implementación de Big Data se creó un modelo de madurez que mediante un proceso cíclico y de mejora continua, el ejecutivo responsable de las decisiones estratégicas en la empresa pueda ir midiendo y constatando la evolución de la implementación de Big Data en cualquier empresa, partiendo desde niveles nulos de existencia, hasta llegar un nivel de madurez adecuado a los objetivos estratégicos de la empresa.

Una vez establecida una estrategia de implementación de Big Data, lo recomendable es tener resultados perceptibles en poco tiempo, por lo que reconociendo que existe diversidad de herramientas tecnológicas y regularmente es requerido personal experto en arquitectura de datos, en arquitectura de infraestructura tecnológica y en otras especialidades, siendo la oferta en Guatemala poca y no accesible para las PYMES que seguramente especulan de una u otra forma sobre el retorno de inversión que tendría implementar una tecnología que mayormente es conocida que es utilizada en grandes corporaciones, se plantean opciones de implementación tecnológica accesibles para las PYMES.

1. BIG DATA EN LAS ORGANIZACIONES EMPRESARIALES

Hoy en día se ha escuchado mucho del término Big Data, y pareciera que es uno de tantos términos de moda que se han adoptado en la jerga de uso común de los responsables de mantener, manipular y administrar datos en las empresas de hoy en día, asumiendo que la mayoría de empresas se encuentran controlando y utilizando de forma adecuada la información que tienen a la mano, y utilizándola como base en la elaboración de planes estratégicos empresariales.

La realidad del uso del Big Data es diferente a lo que creen muchos de los gerentes de empresas de diversos sectores, en cuanto que no están al tanto de la información que no están generando, ni están conscientes de la gran cantidad de datos que la empresa tiene al alcance día con día. Muchos gerentes asumen que, con tener una buena gestión de redes sociales, tener un sistema informático de planificación de recursos empresariales (ERP) y tener un sistema informático de administración de clientes (CRM) que soporten bases de datos con grandes cantidades de datos de los productos, los clientes, los canales de distribución y los canales de venta, ya se encuentran implementando Big Data, lo cual no es suficiente para realizar dicha afirmación.

Las empresas que utilizan sus datos para generar información y esta la utilizan en la elaboración de estrategias empresariales centradas en el cliente y a su vez poseen un ecosistema de información, son las más cercanas a decir que se encuentran implementando Big Data.

1.1. Los ejecutivos de negocios y Big Data

En la actualidad, los ejecutivos de negocios tienen nociones sobre la existencia de Big Data y de alguna forma han tenido un acercamiento con sus beneficios, pero todavía no existe una comprensión total del tema, por lo que es importante realizar una revisión de la comprensión general que se tiene del tema y la situación de su implementación en diversos ámbitos.

1.1.1. Comprensión de Big Data por los ejecutivos

Un estudio realizado en el 2012 por el IBM Institute for Business Value y la Escuela de Negocios Saïd de la Universidad de Oxford muestra que los profesionales afirman que el uso de Big Data brinda una ventaja competitiva y reconocen el potencial de la gama de información en tiempo real que brinda.

Figura 1. Descripción de Big Data por profesionales



Fuente: elaboración propia.

1.1.2. Situación de implementación de Big Data

La implementación de Big Data se ha realizado por diversas empresas en diversos países, existiendo experiencias positivas y negativas en su implementación, observando que es un factor clave considerando dentro de la estrategia empresarial, por lo que es importante revisar experiencias desde diversos ámbitos desde lo general a lo específico.

1.1.2.1. Ámbito mundial

Según revela un estudio realizado en el 2013 por la empresa EMC Corporation¹, el cual tuvo representación de 46 ciudades alrededor del mundo en cuanto a opinión de retos y oportunidades que Big Data y la transformación de TI pueden presentar a las empresas, para ese entonces la gran mayoría de tomadores de decisiones de TI encuestados, consideran que Big Data mejora la toma de decisiones y es un factor clave para mejorar el éxito en el sector empresarial pero contrasta con la realidad de que casi un tercio de los mismos encuestados no tienen planes para implementar la tecnología de Big Data, manifestando que los motivos más comunes identificados como inhibidores de su adopción son la falta de un análisis de rentabilidad clara o de un ROI comprobado, la falta de relevancia para la empresa y la falta de preparación de la cultura empresarial para Big Data².

Las organizaciones que están implementando, planeando o considerando proyectos de Big Data, manifiestan en su mayoría que el objetivo comercial

¹ Dell EMC. *Transforming IT Study Survey Results*. <https://www.emc.com/campaign/global/forum2013/survey.htm>. Consulta: 8 de noviembre de 2017.

² Dell EMC. *EMC Forums 2013 Transforming IT Global Study: Global Results Infographic*. <https://www.emc.com/infographics/forum2013-survey.htm>. Consulta: 8 de noviembre de 2017.

principal al invertir en iniciativas basadas en datos, es mejorar las relaciones con los clientes haciendo que el negocio sea más centrado en los datos mediante el cambio de la forma en que las operaciones están organizadas, según encuesta realizada por IDG en el 2016³. Dicha encuesta también revela una brecha generacional en cuanto que los encuestados más jóvenes están más en acuerdo con el potencial transformador de Big Data, y en cambio los encuestados mayores de 55 años son mucho más inseguros sobre el cambio que puede ofrecer.

1.1.2.2. **Ámbito Estados Unidos**

Tomando como referencia que según el foro económico mundial⁴, los dos países de Norte América (Estados Unidos y Canadá) tienen el mayor índice de competitividad en América y a su vez tienen la mayor capacidad de adopción tecnológica en dicho continente, lo cual implica disponibilidad de las últimas tecnologías, absorción de tecnología a nivel de empresa, medida de inversión extranjera directa en nueva tecnología y su transferencia de tecnología, es importante recordar que no se han de perder de vista los avances de implementación tecnológica en estos países, dado que marcan una brecha para América Latina que normalmente consolida una adaptación de varias de las tecnologías implementadas en estos países algunos años después.

Para el 2017 según la quinta encuesta anual de Big Data de NewVantage Partners⁵ dirigida a los principales responsables de la toma de decisiones empresariales y tecnológicas de empresas incluidas en el listado de *fortune* 1 000, el 48,4 % de los ejecutivos corporativos que fueron encuestados indicaron

³ IDG. *Data & Analytics Survey*, p.3.

⁴ World Economic Forum. *The Global Competitiveness Report 2017–2018*. 2017. ISBN-13: 978-1-944835-11-8.

⁵ NewVantage Partners. *Big Data Executive Survey 2017*. p.6.

que su empresa ha logrado resultados mensurables de sus inversiones en Big Data, un notable 80,7 % de los ejecutivos caracterizaron sus inversiones como exitosas y un 21,0 % de ejecutivos declararon que había sido disruptivo o transformador para su empresa. Asimismo, las empresas se centran en las oportunidades para innovar, al tiempo que reducen los niveles de gastos, teniendo que el 49,2 % informa resultados exitosos de sus esfuerzos de reducción de costos como resultado de las inversiones de Big Data.

Según la encuesta indicada se identifican desafíos que las grandes corporaciones (*fortune* 1 000) aún enfrentan cuando buscan convertirse en organizaciones más orientadas a los datos:

- El 5 % de los ejecutivos informan que su organización ha tomado medidas para crear una cultura basada en datos, pero solo el 37,7 % informa que estos esfuerzos han sido exitosos hasta la fecha.
- La mayoría de los ejecutivos (52,5 %) informan que los impedimentos organizacionales, incluida la falta o alineamiento organizacional, la resistencia empresarial o tecnológica, y la falta de adopción de la gerencia media son factores que limitan el éxito de los esfuerzos de Big Data.
- Mientras que el 60,7 % de los ejecutivos informan que su empresa ha desarrollado una estrategia empresarial de Big Data, el 18 % informa que su empresa carece de una estrategia de datos coherente.

Las grandes firmas se están preparando para una década de cambios disruptivos dado que un 46,6 % de los ejecutivos expresa la opinión de que su empresa puede estar en riesgo de una disrupción importante en la próxima década, vislumbrando una disrupción proveniente de una gama de capacidades

emergentes, que incluyen inteligencia artificial y aprendizaje automático (88,5 %), tecnologías digitales (75,4 %), computación en la nube (65,6 %), tecnología de cadena de valor como servicio (62,3 %), y tecnología financiera como servicio (57,4 %).

1.1.2.3. **Ámbito América Latina**

A nivel de América Latina para el 2017 las empresas han llegado a considerar como mayor prioridad las iniciativas relacionadas con analítica de negocios o Big Data, según TechTarget⁶, lo cual es manifestado por responsables de TI de empresas de 19 países de América Latina, siendo en su mayoría decisores y tomadores de decisiones. A partir del 2014 se observa la maduración que se ha tenido respecto a la implementación de Big Data al ir pasando de un concepto a una consolidación de proyecto e iniciativas en la región de América Latina⁷.

Según el foro económico mundial⁸, en la región de América Latina y el Caribe existen grandes brechas de competitividad que cubrir, incluyendo temas de innovación que a su vez incluyen capacidad de innovación y adquisición de productos de tecnología avanzada, así como a nivel de tecnología se incluyen temas de disponibilidad de las últimas tecnologías, transferencia de tecnología y absorción de la misma por parte de las empresas; aunque también se tiene progresos en infraestructura y capital humano, que juntos con mejoras en la preparación tecnológica e innovación, ayudan a que en América Latina y el

⁶ PÉREZ ARBESÚ, Lizzette. *Prioridades de TI para las empresas de América Latina en 2017*. <http://searchdatacenter.techtarget.com/es/cronica/Prioridades-de-TI-para-las-empresas-de-America-Latina-en-2017>. Consulta: 8 de noviembre de 2017.

⁷ OSORES, Melisa. *Crece Big Data en AL en 2015, impulsada por software y servicios*. <http://searchdatacenter.techtarget.com/es/cronica/Crece-big-data-en-AL-en-2015-impulsada-por-software-y-servicios>. Consulta: 8 de noviembre de 2017.

⁸ World Economic Forum. *The Global Competitiveness Report 2017–2018*. 2017. ISBN-13: 978-1-944835-11-8.

Caribe se descubran nuevas fuentes de crecimiento y garantizar un crecimiento inclusivo y amplio.

1.1.2.4. Ámbito Guatemala

A nivel de Guatemala se cuenta con registros de que instituciones como FUNDESA, CACIF y Cámara de Industria han reconocido desde el 2014, los beneficios de Big Data y como ejemplo de ello, lo han propuesto como solución a la problemática del contrabando y defraudación aduanera⁹.

En el sector gobierno, los ministerios de educación y de finanzas de alguna forma han evolucionado en el uso de herramientas de inteligencia de negocios para el tratamiento de grandes datos, realizando análisis de datos con altos volúmenes y asimismo, mediante la existencia de la carta iberoamericana de calidad en la gestión pública que promueve el establecimiento de un enfoque común acerca de las nociones de calidad y de excelencia en la gestión pública, se están realizando esfuerzos para implementación de datos abiertos y servicios electrónicos disponibles a la ciudadanía.

Para tener una mejor idea de lo que Big Data puede aportar a las empresas es importante ahondar un poco sobre su propósito, algunos conceptos asociados al tratamiento tradicional de los datos en empresas que no han implementado Big Data y los beneficios que puede brindar a las empresas el darle mayor relevancia al valor de los datos.

⁹ ORELLANA CÁMBARA, Maritryny. *Cumplimiento de las disposiciones y recomendaciones internacionales de la organización mundial de aduanas, caso: contrabando en la frontera la mesilla del departamento de Huehuetenango, Guatemala.* p.64.

1.2. ¿Qué tomar en cuenta?

En la actualidad las empresas utilizan diferentes métodos, herramientas y tecnología para el tratamiento de sus datos y extraer conocimiento de los mismos para facilitar la toma de decisiones en diversos niveles operativos, tácticos y estratégicos, de los cuales es importante que los ejecutivos tengan conocimiento de la forma en que las empresas de actualidad están dando uso a sus datos, mencionando a continuación las más relevantes.

1.2.1. Digitalización de los datos

Hoy en día se puede digitalizar casi todo y las empresas ante la necesidad de información han adoptado procesos de digitalización de sus documentos para acceder a datos que se encuentran en ellos, lo cual constituye un aprovechamiento de los datos que ya se poseen, pero no habían sido accesibles para ser analizados por encontrarse en papel. Las empresas también han hecho conciencia la existencia de datos que se encuentran externos a su organización, pero son generados y capturados en todo momento tal como información biométrica de pulseras de movimiento, aplicaciones de seguimiento de actividades deportivas, posición GPS de usuarios de dispositivos móviles, datos demográficos capturados por redes sociales, *streaming*, videoconferencias, sensores de electrodomésticos, entre otros.

1.2.2. Inteligencia de negocios

Desde hace mucho tiempo se tiene la conciencia en las empresas de la ventaja que ofrece el contar con la habilidad de transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios, y esto es logrado mediante la

inteligencia de negocios que permite responder a diversas preguntas con el uso de la consolidación de los datos existentes.

La implementación de inteligencia de negocios conlleva la implementación de procesos de transformación y traslado de almacenamiento de los datos, así como la correspondiente visualización de valores resultantes de información precisa, lo cual ofrece aumento en la eficiencia en la obtención de respuestas, disminuyendo los tiempos para su obtención y de esta forma realizar un mejor control del cumplimiento de metas empresariales conociendo el estado interno de la organización.

1.2.3. Minería de datos

Al contar con bases de datos que contengan detalles de las operaciones de la empresa, se hace uso de métodos matemáticos y estadísticos para reconocer hechos significativos, patrones, tendencias, relaciones y anomalías que no son perceptibles a simple vista, y que ayudan a mejorar la toma de decisiones, lo cual es conocido como minería de datos, dada la analogía de contar con una montaña de datos en donde se escarba mediante la aplicación de diversas técnicas para extraer conocimiento, tomando en cuenta que idealmente esta aplicación debiera ser automática o semiautomática.

Para aplicar la minería de datos en la toma de decisiones empresariales existe una variedad de técnicas entre las cuales es necesario conocer las más relevantes para reforzar decisiones que aumenten la competitividad y generación de nuevas oportunidades de negocio.

1.2.3.1. Agrupamiento de datos

Es importante tener presente que al hablar de datos se debe tener claridad de la información que se quiere generar, y ante la gran cantidad de datos que pueden existir se puede tomar en cuenta el agrupamiento estadístico el cual consiste en que cuando existen gran cantidad de datos (discretos y continuos) que se encuentran muy dispersos, las distribuciones de frecuencias sin agrupar no son la mejor opción para realizar una organización de datos, por lo cual se hace necesario realizar una distribución en intervalos o clases, que hagan posible un resumen de los datos de la variable en estudio, para de esta manera concentrar los datos y así acumular el número de observaciones o frecuencias contenidas para cada clase facilitando su presentación, además de permitir un análisis de aspectos resaltantes que serían muy difícil de observar con datos individuales. Lo anterior se puede explicar de mejor forma haciendo una semejanza a divide y vencerás, en tanto que en lugar de analizar un volumen de datos por completo, se pueden realizar agrupaciones de los mismos.

Al tener agrupaciones de datos resulta interesante observar y explorar sobre el comportamiento y la información que se puede obtener sobre los diversos grupos, esto es mediante el uso del análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados datos.

1.2.3.2. Redes neuronales

El conocimiento actual de neurociencia ha permitido la elaboración de modelos computacionales que permiten reconocer patrones y aprender de los mismos en situaciones donde existen unidades de procesamientos que tienen

gran cantidad de conexiones entre si conformando una estructura, no basándose en un modelo preestablecido y trabajando con complejidad de variables, lo cual ha dado paso a lo que se conoce como redes neuronales.

En el campo empresarial existen entornos dinámicos que dependen de diversas variables que determinan un comportamiento atípico difícil de pronosticar a futuro, como por ejemplo en los mercados financieros, la propagación de una promoción o comportamientos de los procesos de producción, donde las redes neuronales brindan una herramienta para encontrar predicciones donde el análisis numérico de datos históricos se queda limitado en estos escenarios.

1.2.3.3. Regresión

En toda decisión estratégica es importante conocer la relación que existe al modificar valores clave del modelo de negocio que se tiene, como por ejemplo el precio de un producto y el costo de fabricación y como estos aumentarán o disminuirán las ventas, por lo que se hace necesario investigar la relación entre diferentes variables, lo cual se logra mediante el análisis de regresión. Se utiliza cuando se sospecha que una variable puede estar afectando (variable independiente) al comportamiento de otra (variable dependiente) u otras, y en base al estudio de la relación entre estas variables se busca predecir comportamientos futuros.

1.2.3.4. Árboles de decisión

Como es sabido la labor de los ejecutivos en los diversos niveles involucra mucho la toma de decisiones, y en muchos de los casos las mismas tienen una relevancia de alto impacto en la empresa, y a la vez, es requerido que dichas

decisiones sean tomadas con rapidez, por lo que es importante tener claridad en las estrategias empresariales y utilizar los datos existentes para tener los mejores aciertos en la toma de decisiones, y es en esta situación que el uso de la técnica de árboles de decisión permitirá mejorar el proceso de toma de decisiones y plantear estrategias más adecuadas.

Los árboles de decisión permiten analizar los riesgos y beneficios en la toma de decisiones considerando todas las opciones que se planteen y anticipando los posibles resultados de las decisiones a tomar, y logrando mayor precisión al tener la mayor cantidad de datos posible.

1.2.4. Análisis de datos

Todo el proceso que se realiza para extraer información de valor a partir de los datos se conoce como análisis de datos, a lo cual es importante que los ejecutivos tengan presente que este análisis tiene dimensiones de análisis acorde a la acción que se quiera realizar con la información adquirida, por lo que es posible realizar diversos tipos de análisis de datos.

1.2.4.1. Análisis descriptivo

Consiste en análisis de datos que tiene como resultado el brindar respuestas descriptivas del pasado y del presente, dando información sobre qué es lo que ha pasado y está pasando de forma que brinda un mayor conocimiento y comprensión del comportamiento de la empresa.

1.2.4.2. Análisis de diagnóstico

La necesidad de brindar respuestas a las razones y causas por las cuales está ocurriendo el presente, da lugar al análisis de diagnóstico, dando información sobre los por qué se da el presente y determinando las causas, mediante el uso de información existente sobre relaciones que existen entre los datos y la identificación de problemas a través de una interpretación de lo ocurrido.

1.2.4.3. Análisis predictivo

Mediante el análisis de los datos también es posible tener un acercamiento con el futuro y qué sucederá, a lo cual el análisis predictivo hace uso de los datos existentes para determinar el resultado probable de un evento futuro, o bien, una probabilidad de ocurrencia de una situación que ya se produjo, logrando identificación de patrones de comportamiento en la implementación de operaciones de la empresa.

1.2.4.4. Análisis prescriptivo

En la búsqueda de saber cómo hacer que algo ocurra, es posible utilizar los datos y realizar un tipo de análisis que permita accionar con mayor confianza de éxito en el resultado esperado, lo cual es realizado por medio del análisis prescriptivo que realiza la identificación de oportunidades para incrementar la probabilidad de ocurrencia del escenario que se tiene previsto brinde el mayor beneficio, logrando una mayor optimización al determinar qué acciones son las que se tienen que tomar para que este suceda.

1.2.5. KPI

Finalmente se desea tener algo que diga un resultado, a lo cual toda la información que se ha agrupado y explorado debe reflejar el comportamiento de indicadores de desempeño (KPI) que se tiene en la empresa.

Los anteriores conceptos parecen sencillos, pero en la práctica las empresas están aún lejos de hacer realidad dicha implementación, lo que se observa con pocos ejemplos claros de implementación de Big Data en Latinoamérica.

1.3. El valor de los datos

Estando en una época digital que ha propiciado grandes avances en la comunicación, la forma de realizar los procesos, propiciado el manejo de mayor información y ha revolucionado los modelos empresariales, se exige a las empresas la toma de decisiones con celeridad, incrementando de esta forma la importancia de los datos con que cuenta, de los cuales pueda extraer la mejor información para apoyo a la toma de decisiones estratégicas.

La época actual ha permitido que una gran cantidad de datos se encuentre al alcance de muchos y para los que los saben aprovechar, pueden obtener información relevante que puede ofrecer a las empresas ventajas competitivas, de forma que los datos se han convertido en un bien intangible en las empresas que conforman un activo de gran valor empresarial.

El poder que los datos proporcionan ha sido reconocido por ejecutivos que les han dado una importancia primordial para la toma de decisiones estratégicas, siendo los datos un recurso valioso, a nivel de cualquiera de los activos que

brindan capacidad operativa a la empresa, y más aún, los datos brindan un valor intangible a tiempo futuro dado que permiten llegar a cierto nivel de predicción de las operaciones de la empresa y su implementación de planes estratégicos.

Hasta la fecha existe mucha información importante que se encuentra en medios que no han sido procesados o fuentes accesibles que no han sido consultadas, razones por las cuales no se ha realizado su correspondiente análisis, por lo que su valor inmerso no ha sido convertido en un beneficio para la empresa, teniendo los ejecutivos un recurso muerto en sus organizaciones. El desafío para los ejecutivos se encuentra en encontrar la información que puedan proporcionar los datos, sacar el mejor provecho de la misma obteniendo conocimiento que provea una ventaja competitiva.

El ciclo de vida conocido de los datos para extraer su valor se da mediante la identificación de las fuentes de acceso a los datos y su recopilación, realizar el almacenamiento adecuado, analizar los mismos para concluir con una transformación de los mismos en conocimiento y aprendizaje de gran valor para la empresa. La tecnología brinda la capacidad de automatizar este ciclo de vida de los datos.

Para el 2015, según un estudio sobre análisis de datos realizado por la consultora KPMG Internacional a ejecutivos de diversos niveles y de diversos sectores en 15 países¹⁰, se tiene que sólo el 19 % de las empresas consigue extraer verdadero valor de los datos, el 86 % admite que no tiene las habilidades suficientes para tratar sus datos y el 58 % de organizaciones tienen dificultades evaluando la calidad y confiabilidad de los datos. Lo anterior evidencia la brecha que existe en la obtención del valor que tienen los datos en las empresas y el

¹⁰ KPMG International Cooperative. *Going beyond the data turning data from insights into value*. p.6.

potencial que existe para transformarlos en conocimiento que brinde beneficios de competitividad.

1.4. Big Data y la estrategia empresarial

El alto grado de competitividad al que se enfrentan hoy en día las empresas, exige de ellas respuestas acertadas en los procesos de implementación estratégica para lograr éxito y rentabilidad. Las decisiones eficientes determinadas en la estrategia empresarial toman en cuenta el conocimiento y aprendizaje existente y el correspondiente análisis de los posibles escenarios a los cuales se enfrenta el rumbo corporativo, requiriendo la disponibilidad adecuada de la información obtenida de los datos alrededor de la empresa.

Las herramientas tecnológicas brindan una oportunidad muy importante para lograr automatizar el ciclo de vida de los datos para su conversión en conocimiento y aumentar el valor que aportan los mismos para la empresa, logrando la toma de decisiones en los diversos niveles ejecutivos (operativos, tácticos y estratégicos) de forma oportuna, rápida, efectiva y eficiente.

Como parte de las estrategias empresariales que tienen contemplada la transformación digital y la transformación de los datos en valor se tiene que Big Data ofrece un aporte considerable para lograr llevar a cabo la realización de dichas estrategias.

Big data permite automatizar la transformación de datos para la obtención de su valor mediante diversas herramientas que facilitan extrapolar, enfocar, visualizar, reflexionar, perfeccionar, modelar y predecir resultados con

tecnologías de aprendizaje automático que entienden los datos para ayudarnos a la toma de decisiones estratégicas incluidas en planes estratégicos.

Big Data genera diversos beneficios como parte de la estrategia empresarial que se puede planificar, tal como:

- Prontitud en la toma de decisiones mediante un análisis de la información recopilada en tiempo real.
- Implementación de inteligencia en el desarrollo de las operaciones implementadas por planes estratégicos inteligentes, dada la automatización de procesos con conocimiento de tendencias y pronósticos.
- Eficiencia en cuanto a implementación de estrategias enfocadas a optimización de costos, dado el conocimiento adquiridos sobre la evolución del producto y su ciclo de vida.
- Incremento de comunicación y efectividad con estrategias enfocadas a clientes, así como mejora en la comunicación con los aliados estratégicos de la empresa.

1.5. Tendencias de estrategias de datos en las organizaciones

Dada la importancia de los beneficios que brinda Big Data y la velocidad de su maduración, los ejecutivos deben conocer las estrategias recientes que las empresas están implementando y el perfil de ejecutivos que se está formando.

1.5.1. Empresa Data Driven Company

Los ejecutivos de hoy en día siguen requiriendo tomar decisiones objetivas basadas en información de la empresa lo más real y reciente posible, lo cual exige que la obtención de la información utilice métodos, procesos y sistemas acorde a la existencia de datos que regularmente pueden estar incompletos, desordenados, difíciles de administrar por su amplia cantidad pero que de igual forma es necesario sean analizados para obtener información que sea insumo para la toma de decisiones estratégicas. Lo anterior tiene como resultado que las empresas de hoy en día utilicen nuevos paradigmas científicos para el tratamiento de los datos, lo cual es ofrecido por el campo de *data science*, y en consecuencia surgen las empresas Data Driven Company que básicamente son empresas basadas en la información, centralizadas en los datos para la realización de los procesos y toma de decisiones en todos los niveles, apoyándose con el uso de *data science* como estrategia central de su negocio.

En una empresa Data Driven Company se encuentran características como:

- Organizacionalmente utilizan jerarquías horizontales en la que empleados y directiva están en contacto continuo.
- Toman decisiones muy rápidamente y en muchos casos de forma automática, en los diversos niveles operativos de la organización basados en la correspondiente analítica de datos.
- Trabajan con base en objetivos con adecuado seguimiento y control de KPI's.

- Evalúan a sus empleados con base en resultados.
- Plantean las decisiones como si fueran problemas científicos que deben ser resueltos con la ayuda de los datos.
- Fomentan la curiosidad de sus empleados.
- Incentivan la búsqueda de nuevas formas de explotar la información disponible en empresa.
- Tienen una política de datos muy consistente que incluye una gobernanza de datos orientada a *self service* de datos centralizados y organizados adecuadamente.
- Entienden a la perfección qué pretende el cliente de ellas y ajustan sus experiencias a estas expectativas.
- Buscan utilizar la última tecnología de análisis y medición de datos.

Según Christopher Penn, líder reconocido en marketing digital y conferencista en el campo de data driven, existen 5 etapas en las que se puede encontrar una empresa respecto a los datos¹¹:

- Resistencia: donde se identifica a las empresas que son resistentes al cambio, creyendo que no existe necesidad de evolucionar puesto que siempre han hecho las cosas de la misma manera.

¹¹ PENN, Christopher. *The Evolution of the Data-Driven Company*. <http://www.christopherspenn.com/2016/11/the-evolution-of-the-data-driven-company/>. Consulta: 8 de noviembre de 2017.

- Consciencia y conocimiento de datos: donde se identifica a las empresas que son conscientes del valor de los datos, pero están centradas únicamente en su recolección, sin extraer su valor.
- Análisis y guía por los datos: donde se identifica a las empresas que analizan lo que les dicen los datos, extraen conclusiones y obtienen valor de los datos. Esta etapa es común que sea donde pasen muchas empresas por muchos años.
- Comprensión y conocimiento estratégico de los datos: donde se identifica a las empresas que prestan atención a los datos de forma estratégica, descubriendo el porqué de estos datos mediante investigación interna, extrayendo *insights*.
- Data driven: donde se identifica a las empresas que van más allá del análisis de datos y extracción de *insights*, puesto que son capaces de contestar a la pregunta: ¿qué viene después?, mediante el uso de datos en todos los niveles de la organización, impulsando la empresa por los datos de una forma estratégica.

1.5.2. Ejecutivo 4.0

Los avances de la tecnología y su ritmo acelerado de actualización e implementación en las empresas, permiten que hoy en día una gran cantidad de las grandes empresas hayan logrado el aprovechamiento de las capacidades digitales que ofrece la tecnología y a su vez hayan logrado su implementación en todos y cada uno de los niveles: proceso, producto y modelo de negocio. Lo anterior exige que los ejecutivos modifiquen y actualicen su forma de gerencia y

toma de decisiones, convirtiéndose en líderes de una administración digital en empresas que están siendo parte de la denominada cuarta revolución industrial, que no es más que efecto de la digitalización de las mismas.

Los ejecutivos de hoy en día deben tener las capacidades de contar con suficiente conocimiento, poseer habilidades y la adecuada actitud que le permita relacionar el mundo físico y el digital, así como manejar en su cotidianidad la comunicación y el tratamiento de datos e información obtenidos en un entorno seguro, cambiante y complejo, logrando realizar velozmente el análisis de la información para la toma de decisiones estratégicas. Estos ejecutivos que se desenvuelven en organizaciones que han pasado a la cuarta revolución industrial provocada por la digitalización, son conocidos como los ejecutivos 4.0.

Los ejecutivos que desean entrar en la categoría de ejecutivos 4.0 tienen el reto de reinventarse rápida y profundamente para tener el adecuado perfil que les permita el aprovechamiento de las capacidades que brinda la tecnología. Si los ejecutivos de la vieja escuela se aferran a una actitud de la empresa de TI tradicional, se arriesgan a convertirse en ciudadanos de segunda clase en la suite ejecutiva¹², en el sentido que únicamente estarán dirigiendo operaciones, pero se irán alejando de la verdadera labor de la toma de adecuadas decisiones estratégicas para la empresa.

Los ejecutivos deben ser conscientes que aunque la tecnología ofrezca muchas ventajas para la digitalización de las empresas, dichas empresas no están solo conformadas por máquinas y componentes físicos que pueden prestarse a la mejor innovación tecnológica, sino que también existen personas que tienen la dirección del caminar de la organización, quienes deben acompañar

¹² TYNAN, Dan. *The skills and traits of a next-generation CIO*. <https://www.cio.com/article/3207542/leadership-management/the-skills-and-traits-of-a-next-generation-cio.html>. Consulta: 8 de noviembre de 2017.

adecuadamente las transformaciones tecnológicas, pues de no ser así disminuyen dicha innovación e impacto de implementación digital. Los ejecutivos 4.0 comprenden a cabalidad los conceptos tecnológicos y más importante aún, comprenden las capacidades y bondades que la tecnología brinda a la empresa.

La consultora RocaSalvatella propone 8 competencias digitales para el éxito profesional¹³ que todo ejecutivo debe desarrollar al estar al frente del proceso de transformación digital en las empresas, logrando conseguir un mayor impacto dentro de un enfoque hacia resultados. Dichas competencias profesionales para la transformación digital son:

- **Conocimiento digital:** que se comprende como una de las competencias más intangibles, siendo una capacidad para desenvolverse profesional y personalmente en la economía digital, pero que finalmente aporta a la empresa un aumento de productividad y aceleramiento de innovaciones, así como seguridad en los cambios.
- **Gestión de la información:** que se comprende como la capacidad para buscar, obtener, evaluar, organizar y compartir información en contextos digitales, particularmente con análisis de grandes datos, aportando a la empresa competitividad e inteligencia de negocio con información convertida en conocimiento relevante.
- **Comunicación digital:** que se refiere a la capacidad para comunicarse, relacionarse y colaborar de forma eficiente con herramientas y en entornos digitales, propiciando una empresa conectada, aportando a la empresa un adecuado conocimiento interno y un mejor posicionamiento de marca.

¹³ RocaSalvatella. *Cultura Digital y Transformación de las Organizaciones*. p.8.

- Trabajo en red: como capacidad para trabajar, colaborar y cooperar en entornos digitales, sacando el mayor provecho de espacios virtuales para el efecto, y otros mecanismos en red, aportando a la empresa un mayor sentido de pertenencia.
- Aprendizaje continuo: dado como capacidad para gestionar el aprendizaje de manera autónoma, conocer y utilizar recursos digitales, mantener y participar de comunidades de aprendizaje, mediante procesos formales y no formales, aportando a la empresa mayor competitividad.
- Visión estratégica: entendida como capacidad para comprender el fenómeno digital e incorporarlo en la orientación estratégica de los proyectos de su organización, particularmente en el entendido que el líder es el principal impulsor de la transformación digital, aportando a la empresa una visión clara, motivación y sostenibilidad del proceso de transformación digital.
- Liderazgo en red: que consiste en la capacidad para dirigir y coordinar equipos de trabajo distribuidos en red y en entornos digitales, cobrando relevancia la característica de equipos distribuidos que exigen los entornos globalizados, requiriendo el adecuado manejo de herramientas digitales, aportando a la empresa un mejor uso de los recursos y una integración de equipos de trabajo distribuido.
- Orientación al cliente: dada como la capacidad para entender, comprender, saber interactuar y satisfacer las necesidades de los nuevos clientes en contextos digitales, aportando a la empresa calidad y excelencia en las experiencias de clientes.

Adicional a las propias habilidades, los ejecutivos 4.0 deben velar porque dentro de las empresas se encuentren habilidades en el personal para un mayor éxito en el proceso de transformación digital, entre las cuales se pueden mencionar las habilidades para un mundo digital propuestas por la organización para la cooperación y el desarrollo económicos¹⁴:

- Análisis de datos: especialmente de grandes volúmenes con características diversas y complejas, que permita tomar las mejores decisiones.
- Desarrollo de software: para la implementación de automatización de procesos mediante software.
- Habilidades de seguridad cibernética: dada la necesidad de cumplimiento de regulaciones legales, así como salvaguarda de la información y conocimiento generado en la empresa.
- Habilidades efectivas de comunicación y análisis: fundamentales para proporcionar y compartir información en toda la organización a medida que se produzcan cambios a causa de la transformación digital.
- Creatividad, adaptabilidad y capacidad de aprendizaje: dada la modificación de funciones laborales que propicia la transformación digital.
- Alfabetización informática: para utilizar adecuada y oportunamente la tecnología en las labores.

¹⁴ OECD. *Skills for a digital world*. p.4.

1.5.3. Técnicas de análisis de datos no convencionales

El incremento de la variabilidad de fuentes de datos que se encuentran hoy en día, así como los avances tecnológicos dan lugar a aplicar técnicas de análisis de datos bajo la nueva generación digital de la información tal como:

- Análisis de sentimiento que permite determinar la actitud de un individuo o grupo hacia un tema particular, especialmente utilizada para el análisis de los datos provenientes de redes sociales.
- Análisis semántico de textos que mediante un proceso trata de extraer valor, a través del análisis semántico de grandes volúmenes de textos.
- Análisis de imagen, voz y vídeo utilizado por grandes empresas para optimizar procesos y realizar pronósticos.
- Implementación de modelos inteligentes con inteligencia artificial en procesos que requieren análisis de grandes cantidades de datos y obtener mayor precisión en los resultados y automatizando el aprendizaje de la empresa.

2. CONCEPTOS TÉCNICOS GENERALES DE BIG DATA

El concepto de Big Data es conocido mayoritariamente por la referencia de almacenamiento de altos volúmenes de datos. Esta conceptualización de Big Data en parte puede ser por su traducción literal y en parte porque la tecnología implicada en su implementación permite almacenar grandes cantidades de datos, pero, el manejo de altos volúmenes de datos es solo una característica, y en este sentido es importante destacar que la importancia no radica en contar con la capacidad de almacenar grandes volúmenes de datos, sino en realizar el análisis de los datos existentes, generar ideas, tomar decisiones y materializar estrategias empresariales en base a la información que se tiene.

2.1. Definiendo Big Data

Siendo Big Data bastante popular hoy en día, no hay una definición específica que fije con claridad, exactitud y precisión su significado, por lo que para comprender su significado se presenta a continuación los hitos más importantes que han reconocido su existencia, definiciones más reconocidas y una descripción de sus características.

2.1.1. Síntesis cronológica

Al rastrear en el tiempo el surgimiento del término Big Data se tiene una primera referencia utilizada en un artículo escrito en 1989 por Erik Larson¹⁵ sobre como las empresas utilizan la gran cantidad de información recopilada de las

¹⁵ LARSON, Erik. *How Did They Get Your Name? Direct-mail Firms Have Vast Intelligence Network Tracking Consumers.* http://articles.orlandosentinel.com/1989-07-26/lifestyle/8907254531_1_subscribe-to-magazines-subscriber-list-junk-mail. Consulta: 22 de noviembre de 2017.

personas para fines de comercialización, pero algunos autores consideran que no siendo una definición técnica solo es una referencia¹⁶.

El término Big Data se usó por primera vez de forma técnica en julio de 1997 en un artículo¹⁷ realizado por los investigadores de la NASA Michael Cox y David Ellsworth cuando empezaron a encontrarse con la dificultad de tratar y visualizar grandes cantidades de información. Ambos afirmaron, que el ritmo de crecimiento de los datos empezaba a ser un problema para los sistemas informáticos actuales. Esto se denominó el problema del Big Data.

En el año 2001 Doug Laney de la empresa consultora y de investigación de TI Gartner Inc., define las 3 V del Big Data¹⁸ (volumen, velocidad y variedad), las cuales son presentadas como 3 dimensiones que son desafíos en la administración de datos ante el aumento del comercio electrónico en ese momento.

La revista de tecnología Wired hizo popular el Big Data mediante la publicación de un artículo en el 2008¹⁹, donde presenta la existencia de cantidades masivas de datos y el concepto de Big Data, así como reconoce la oportunidad diferenciadora que brinda al encontrar respuestas ante grandes cantidades de datos.

¹⁶ BARKER, Dan. *The Real Original Source of the Phrase "Big Data"*. <http://barker.co.uk/bigdata>. Consulta: 22 de noviembre de 2017.

¹⁷ COX, Michael y ELLSWORH, David. *Application-Controlled Demand Paging for Out-of-Core Visualization*. p.1.

¹⁸ LANEY, Doug. *Application Delivery Strategies*. p.1.

¹⁹ ANDERSON, Chris. *The end of theory: the data deluge makes the scientific method obsolete*. *Wired*. <https://www.wired.com/2008/06/pb-theory/>. Consulta: 11 de noviembre de 2017.

En ese mismo año un grupo de científicos de Computing Community Consortium realizaron una publicación de un artículo²⁰ que hace un aporte en la credibilidad de Big Data, dado que le presentan como una innovación de la década y reconociendo su potencial en transformar las actividades y operaciones de diversos sectores de la industria.

En el 2012 la *information, communications, and society journal*, publica un artículo²¹ que presenta el Big Data como un fenómeno cultural, tecnológico y académico que se basa en la interacción de la tecnología, el análisis y la mitología y bajo esta situación promueve discutir sobre los problemas que conlleva.

2.1.2. Definiciones

Para establecer una definición conjunta de Big Data se presentan 3 definiciones realizadas por empresas reconocidas en el ámbito de tecnologías de la información.

El diccionario de tecnología de la empresa Gartner Inc.²² define: “Big data son activos de información de gran volumen, de alta velocidad, de veracidad, de valor y de gran variedad, que demandan formas rentables e innovadoras de procesamiento de la información para mejorar la comprensión, la toma de decisiones y automatización de procesos”.

²⁰ BRYANT, KATZ y LAZOWSKA. *Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society*. p.7.

²¹ DANAHO BOYD, Kate Crawford. *Critical Questions for Big Data*. p.662.

²² Gartner. *Big Data*. Gartner. <https://www.gartner.com/it-glossary/big-data/>. Consulta: 11 de noviembre de 2017.

La firma de investigación Forrester a través de su vicepresidente Gualtieri²³ define: “Big data es la capacidad de una empresa para almacenar, procesar y acceder a todos los datos que necesita para operar de manera efectiva, tomar decisiones, reducir riesgos y atender a los clientes”.

La firma IDC a través de su vicepresidente Cartner²⁴ indica que “las tecnologías de Big Data describen una nueva generación de tecnologías y arquitecturas diseñadas para extraer valor económico de grandes volúmenes de datos heterogéneos, habilitando una captura, identificación o análisis a alta velocidad”

Adicionalmente en el ámbito académico se comprende que la acelerada y caótica evolución de la literatura de Big Data ha impedido el desarrollo de una aceptación universal y formal para una definición para Big Data existiendo diversas propuestas, en tanto que ninguna ha impedido que los trabajos posteriores modificaran o ignorasen definiciones anteriores y sugirieran nuevas²⁵. La Universidad de Amsterdam en conjunto con la Universidad Politécnica de Hong Kong ofrece una redefinición de lo expresado por Gartner²⁶ indicando que “Las tecnologías de Big Data (datos intensivos) tienen como objetivo procesar datos (estructurados y no estructurados) de alto volumen, alta velocidad y alta variedad, para extraer el valor de los datos previstos y garantizar la alta veracidad de los datos originales y la información obtenida que demanda una relación costo-beneficio, formas innovadoras de procesamiento de datos e información (análisis)

²³ GUALTIERI, Mike. *The Pragmatic Definition Of Big Data*. Forrester. https://go.forrester.com/blogs/12-12-05-the_pragmatic_definition_of_big_data/. Consulta: 11 de noviembre de 2017.

²⁴ CARTER, Philip. *Big Data Analytics: Future Architectures, Skills and Roadmaps for the CIO*. p.1.

²⁵ DE MAURO, Andrea, GRECO, Marco y GRIMALDI, Michele. *A formal definition of Big Data based on its essential features*. p. 6.

²⁶ DEMCHENKO, Yuri, DE LAAT, Cees y MEMBREY, Peter. *Defining architecture components of the Big Data Ecosystem*. p. 4.

para mejorar el conocimiento, la toma de decisiones y el control de procesos; todos ellos exigen (deben ser respaldados por) nuevos modelos de datos (soportan todos los estados de datos y etapas durante todo el ciclo de vida de los datos) y nuevos servicios y herramientas de infraestructura que permiten obtener (y procesar) datos de una variedad de fuentes (incluyendo redes de sensores) y entregando datos en una variedad de formas a diferentes consumidores y dispositivos de datos e información".

Con base en lo anterior se define Big Data como la combinación de características, capacidades y tecnologías en torno a los activos de información empresariales que permiten la transformación de los datos en valor económico acorde a la visión y estrategia empresarial.

2.1.3. Características: las V de Big Data

Como se ha indicado anteriormente, la referencia de las características de Big Data está dada mayormente por las publicaciones de la empresa consultora y de investigación de TI Gartner Inc., siendo inicialmente volumen, velocidad y variedad, pero al darse con el tiempo una evolución de Big Data se pueden reconocer otros atributos que amplían su conceptualización, llegando hoy en día a encontrar que algunos expertos investigadores del tema han recopilado hasta 42 características diferentes pudiendo estar en camino de definir 100²⁷.

2.1.3.1. Volumen

Es sin duda la característica por la cual es más comprendido el Big Data, la cual hace referencia al tamaño de almacenamiento para una cantidad masiva

²⁷ SHAFER, Tom. *The 42 V's of Big Data and Data Science*. <https://www.elderresearch.com/company/blog/42-v-of-big-data>. Consulta: 11 de noviembre de 2017.

de datos. Como se puede observar esta característica es bastante subjetiva y dado los avances de la tecnología la comprensión de altos volúmenes de datos varía en el tiempo, asimismo el criterio de consideración para una cantidad masiva de datos también es acorde al sector de la industria o el alcance geográfico de la actividad que realice la empresa. El parámetro de referencia al momento es de cantidades de datos medidas en terabytes como mínimo, y en muchos casos se manejan cantidades desde *petabytes* hasta *zetabytes*.

2.1.3.2. Velocidad

Esta característica hace referencia a la velocidad del movimiento, proceso y captura de datos que a su vez presenta una creciente demanda de rapidez por la necesidad de contar con información en tiempo real. Hoy en día la necesidad de acceso a los datos en tiempo real requiere la mínima latencia, es decir, el mínimo de tiempo de espera entre el momento en el que se crean los datos, el momento en el que se captan y el momento en el que están accesibles.

2.1.3.3. Variedad

Esta característica hace referencia a la diversidad de tipos y fuentes de datos que son gestionados, entre los cuales se toma en cuenta los datos tradicionales que comúnmente están estructurados, los datos semiestructurados y los datos no estructurados, que a su vez pueden ser obtenidos de fuentes internas o externas a la empresa, pero definitivamente es necesario que sean integrados para su correspondiente análisis.

2.1.3.4. Variabilidad

Esta característica hace referencia al constante cambio que hoy en día tiene el significado de los datos, particularmente los que no son estructurados en los casos que su significado puede variar acorde al contexto y por lo mismo no poseen definiciones estáticas.

Considerando la tecnología que soporta Big Data, esta característica también hace referencia a la elasticidad que se posee para administrar al cambio constante del comportamiento en los flujos de datos, tanto en volumen como en variedad, dado que de un momento a otro pueden variar enormemente.

2.1.3.5. Veracidad

Esta característica hace referencia a la existencia de falta de exactitud, certeza y precisión de los datos, lo cual conlleva la correspondiente labor de generar datos confiables con un alto grado de fiabilidad para ser analizados.

Esta característica exige una particular adaptación a la incertidumbre, flexibilidad a trabajar con ella y obliga a contar con procesos de limpieza de datos y reducción de datos erróneos.

2.1.3.6. Valor

Esta característica hace referencia al costo de la tecnología utilizada y el valor económico obtenido del uso de dicha tecnología, lo cual se interpresa como transformación de los datos en rentabilidad para la empresa.

2.1.3.7. Visualización

Esta característica hace referencia al reto de representar la información de un modo claro, accesible y fácil de leer tomando en cuenta la variedad y volumen de los datos.

2.1.3.8. Vaguedad

Esta característica hace referencia a que a pesar de los altos volúmenes de datos que se pueda tener, es posible que el significado de los mismos no esté claro.

2.1.3.9. Validez

Esta característica se encuentra muy relacionada a la veracidad y hace referencia a los procesos que conlleva dar por aceptado los datos para ser utilizados en el correspondiente análisis que los transforme en información.

2.1.3.10. Vocabulario

Esta característica hace referencia a la diversidad de vocabulario que se requiere para afrontar una diversidad de problemas, acorde a diferentes contextos y diversos contenidos relacionados a los datos.

2.1.3.11. Lugar

Esta característica como traducción de la palabra inglesa *venue*, hace referencia a la diversidad de localidades donde se encuentran los datos teniendo en consecuencia el manejo de datos físicamente distribuidos.

2.1.3.12. Otras características

La cantidad de características que se van reconociendo en Big Data sigue en incremento como consecuencia del crecimiento acelerado de generación de datos que se da en cada momento dando como efecto tomar en cuenta las características de volatilidad, versionamiento, viscosidad, varifocalidad, visibilidad, virage de los datos.

2.1.4. Capacidades de Big Data

Big data posee fortaleza en sus capacidades, siendo más notorias y reconocidas las que corresponden a almacenamiento, procesamiento y análisis de datos.

2.1.4.1. Almacenamiento

Esta capacidad hace referencia a los recursos de hardware y software que permite el almacenamiento eficiente de altos volúmenes de datos, en búsqueda de costos más bajos para dicho objetivo.

2.1.4.2. Procesamiento

Esta capacidad hace referencia al conjunto de herramientas de procesamiento de los datos que permitan manipularlos eficientemente, en búsqueda de incrementar la velocidad del tratamiento de los mismos.

2.1.4.3. Análisis

Esta capacidad hace referencia a las diversas metodologías que se tienen para realizar el análisis de los datos que son convertidos en información, tomando en cuenta la variedad de los mismos.

2.2. Tendencias tecnológicas que demandan los clientes

El ritmo acelerado de evolución de Big Data conlleva una constante atención a diversidad de necesidades que se requiere atender como demanda del comportamiento en la generación de datos, teniendo hoy en día, que toda propuesta de generación de valor de los datos debe orientarse a considerar distintas corrientes, que si bien es cierto la velocidad de cambio es mayor a la velocidad de mantenerse al día, se ha de conocer los horizontes hacia donde se encamina el uso de la tecnología.

2.2.1. Movilidad

Hoy en día la diversidad de operaciones, acciones e interacciones son realizadas desde diversos lugares en todo momento mediante diversos dispositivos accesibles e interconectados desde afuera de la empresa especialmente desde dispositivos móviles.

2.2.2. Cloud

El paradigma actual de ubicación de los datos es literalmente en la nube, la cual está conformada por diversidad de servicios interconectados a través de la red que provee internet, teniendo en consecuencia que los datos se encuentran fuera de la empresa en dispositivos de los que se desconoce detalle de los

mismos, importando para la empresa únicamente que el servicio se encuentre en total funcionamiento todo el tiempo.

2.2.3. Social media

Dado que el ser humano es un ser social por naturaleza, es de reconocer que existe una clara relación entre la empresa y la interacción con los clientes a través de las redes sociales y a todos los mecanismos que se utilizan para dicho efecto es lo que se conoce como social media que no puede faltar en toda estrategia de marketing empresarial.

2.2.4. Análisis en tiempo real

La variedad y cantidad de datos existentes que se tienen al alcance sobrepasa los modelos tradicionales de análisis de datos, particularmente por la velocidad en que son generados, lo cual ha incidido en la existencia de nuevos modelos de análisis de datos que incluyen el trabajar con datos no estructurados, así como el análisis en tiempo real mientras los datos se encuentran en movimiento o con algoritmos predefinidos que realizan análisis mediante la ejecución de procesos específicos.

2.3. Big Data 2.0

Desde el 2014 y 2015, expertos en el tema de Big Data han identificado que existe una nueva generación de Big Data a la cual se le ha llamado Big Data 2.0, la cual es respuesta a la tendencia de análisis en tiempo real que permite

tomar decisiones optimas de forma inmediata²⁸.

La primera generación de Big Data en sus capacidades iniciales se vio afectada por la pobre integración de sistemas tradicionales que son inflexibles y difíciles de administrar y dado que el comportamiento de los datos continúa cambiando hacia cargas de trabajo más sofisticadas que requieren una analítica sobre la marcha de procesos operacionales requirió un cambio en las arquitecturas tecnológicas como mejora a las arquitecturas tradicionales existentes.

La nueva generación de Big Data contempla dar respuesta a las necesidades de procesamiento casi inmediato, tanto como a las cargas de trabajo más complejas que generan los clientes en su forma de producir y demandar los datos mediante la incorporación de tecnologías como bases de datos columnares, bases de datos en memoria y ecosistemas compartidos de bases de datos no estructuradas y tradicionales, lo cual hoy en día es accesible económicamente por la disminución de costos en lo correspondiente a hardware y la existencia de software de código abierto para operar sobre dichas tecnologías.

2.3.1. Tecnologías asociadas

En conjunto con la evolución de Big Data existen tecnologías que con parte del ecosistema que le va conformando por lo que es importante conocer las mismas.

²⁸ EUROPA PRESS. *Especialistas destacan que el Big Data 2.0 hace crecer más de un 30% a las empresas de comercio electrónico.* <https://www.20minutos.es/noticia/2487714/0/especialistas-destacan-que-big-data-2-0-hace-crecer-mas-30-empresas-comercio-electronico/>. Consulta: 11 de noviembre de 2017.

2.3.1.1. *Internet of things (IoT)*

Dada la capacidad que tienen los dispositivos electrónicos de interconectarse mediante a una red, la interconexión de los mismos brinda un nivel de inteligencia de forma que puedan comunicarse sin intervención humana a lo cual se conoce como internet de las cosas (*internet of things*), lo cual permite que los dispositivos de la vida cotidiana tengan cierta independencia de los seres humanos e interactúen con otros dispositivos de forma que automáticamente se realicen procesos.

2.3.1.2. *Machine learning*

Como una aplicación producto de la inteligencia artificial existen técnicas para que las máquinas puedan aprender y tomar decisiones por sí mismas mediante un aprendizaje automático (*machine learning*), lo cual es posible mediante la detección de patrones dentro de un conjunto de datos de manera que es el propio programa el que predice qué situaciones podrían darse o no, realizándose cálculos que permiten a las máquinas un aprendizaje para generar decisiones.

2.3.1.3. *Deep learning*

Para la automatización del análisis predictivo, la inteligencia artificial ofrece algoritmos que ofrecen un acercamiento a la percepción humana con la implementación de redes neuronales que componen el llamado aprendizaje profundo (*deep learning*), el cual utiliza estructuras lógicas que se asemejan en mayor medida a la organización del sistema nervioso de los mamíferos, teniendo capas de unidades de proceso (neuronas artificiales) que se especializan en detectar determinadas características existentes en los objetos percibidos.

2.3.1.4. *Complex event processing*

La necesidad de reconocer patrones de comportamiento en tiempo real, ha dado lugar a la implementación de tecnologías de software que permiten ejecutar operaciones sobre eventos complejos (*complex event processing*), dando la capacidad de leerlos, transformarlos e incluso abstraer un conjunto de eventos hacia un evento complejo que facilite su procesamiento.

2.3.2. Tecnologías en fase de crecimiento

La firma de investigación Forrester estableció 10 tecnologías relacionadas a Big Data que se encontraban para el primer trimestre del 2016 en fase de crecimiento²⁹.

2.3.2.1. Cifrado y enmascaramiento de datos

Dada la alta disponibilidad de acceso a la información, el tema de protección de la información sensible que corresponde a datos de identidad o información privada exige el uso de cifrado y enmascaramiento de datos en grandes volúmenes de datos a lo cual las tecnologías de Big Data conllevan ayudar a automatizar y simplificar dicha labor de seguridad y privacidad.

2.3.2.2. Integración de datos

Las empresas se están movilizandohacia arquitecturas de centros de datos distribuidos lo cual acrecienta la demanda de integración de datos mediante herramientas tecnológicas más sofisticadas, que hoy en día, incluyen la funcionalidad de operar datos en una arquitectura de procesamiento masivo

²⁹ NOEL YUHANNA, Brian Hopkins. *TechRadar™: Big Data, Q1 2016*. p.23.

en paralelo (*massively parallel processing* - MPP) para manejar rendimiento a escala.

2.3.2.3. Virtualización de datos

Dada la movilidad y el uso de tecnologías cloud la información empresarial puede tener varias fuentes de datos que se están dispersas e inconexas, incluyendo bases de datos, almacenes de datos empresariales y almacenes de datos distribuidos, a lo cual, la tecnología proporciona la capacidad de integrar datos de diversidad de fuentes, ubicados en la nube o locales, así como de software como servicio, poniéndolos a disposición como datos virtuales en un único punto de acceso a todos los datos corporativos, brindando agilidad en el consumo de los mismos.

2.3.2.4. Almacenes de archivos distribuidos

Para Big Data siempre ha estado presente la necesidad de almacenamiento y procesamiento de grandes conjuntos de datos lo cual se da mediante el uso de sistemas de archivos distribuidos lo cual permite acceder a grandes cantidades de datos en *clusters* que comprenden servidores en una arquitectura de escalamiento horizontal.

2.3.2.5. Datos en memoria

El aumento de la demanda de procesamiento y análisis de Big Data en tiempo real ha exigido que la tecnología ofrezca acceso a datos con baja latencia para grandes cantidades de datos, permitiendo su análisis predictivo en tiempo real, lo cual se da por medio de estructuras de datos en memoria con capacidad

de almacenar, procesar y acceder a archivos con grandes cantidades de datos en DRAM, flash y SSD.

2.3.2.6. Monitoreo y administración

La adopción empresarial de plataformas tecnológicas para manejo de grandes cantidades de datos conlleva la necesidad de una supervisión y un control más robustos, a lo cual la tecnología ofrece herramientas de monitoreo para la supervisión y administración de la infraestructura y tareas de supervisión del trabajo analítico, permitiendo realizar ajustes y distribución de recursos durante pruebas y operaciones.

2.3.2.7. Base de datos NoSQL

A medida que las organizaciones buscan apoyar nuevos conocimientos de negocios, análisis y análisis predictivo de más variedad de datos, los modelos de datos tradicionales van dando paso a la tecnología para bases de datos NoSQL, la cual proporciona una arquitectura de escalamiento horizontal y ofrece características más avanzadas de bases de datos tradicionales, tales como administración flexible de esquemas, procesamiento de datos distribuidos, automatización de *sharding database* o particionamiento horizontal, navegación de datos conectados, almacenamiento flexible de documentos y procesamiento de consultas complejas, incluyendo motores de recomendación, análisis de clientes, análisis de patrones y otras aplicaciones a escala de internet.

2.3.2.8. Análisis predictivo

Como se ha indicado previamente, el análisis predictivo es la tecnología indispensable a la generación de Big Data 2.0, la cual permite interpretar los miles

de datos que circulan por internet y crear algoritmos para conocer las tendencias de los clientes pudiendo identificar patrones y comportamientos de forma anticipada.

2.3.2.9. Búsqueda y descubrimiento de conocimiento

Ante la expansión de la cantidad y diversidad de datos que los usuarios pueden buscar rápidamente han surgido las tecnologías de búsqueda en Big Data y que a su vez permiten el descubrimiento de conocimiento, construidas en su mayoría con código abierto, ayudando a satisfacer la demanda de información a escala de internet.

2.3.2.10. Análisis de flujo

Streaming analytics es una tecnología que permite el análisis de flujos y comportamientos en tiempo real, permitiendo establecer alarmas de actividades que se detecten.

2.4. Smart data

Una vez comprendido el concepto de Big Data, existen tecnologías que permiten explotar en mayor medida la característica de velocidad en los datos, permitiendo que los sistemas de tratamientos de datos realicen el proceso de análisis en forma inmediata durante el procesamiento de los datos, de forma que minimiza la importancia del tratamiento de altos volúmenes de datos y dando mayor énfasis a la calidad de los datos que se han obtenido.

El *smart data* realiza un análisis en tiempo real y una extracción de información de estos datos a través de fórmulas matemáticas o algoritmos con el objetivo de resolver un problema, ofreciendo altas velocidades para este procesamiento, de forma que es utilizado para tomar decisiones de forma automática para entornos con gran número de sensores que proporcionan datos continuamente como ciudades inteligentes.

3. PROPUESTA DE METODOLOGÍA DE IMPLEMENTACIÓN DE BIG DATA EN PLANES ESTRATÉGICOS EMPRESARIALES

La realidad económica del país de Guatemala es que gran parte de la fuente de generación y sostenibilidad de empleo es aportada por pequeñas y medianas empresas conocidas como PYMES, que según el último informe del Sistema Nacional de Información MIPYME Guatemala³⁰, elaborado por la Oficina Coordinadora de Estadísticas de la Micro, Pequeña y Mediana Empresa -OCSE MIPYME-, se tiene que las PYMES en Guatemala representan el 35 % del Producto Interno Bruto (PIB) y el 85 % de la fuerza laboral del país. Definiendo según normativa del país (acuerdo gubernativo 211-2015) una pequeña empresa a toda unidad de producción, que realiza actividades de transformación, servicios o comercio, que tiene entre 11 y 80 trabajadores, y genera ventas anuales de entre 191 a 3 700 salarios mínimos, equivalentes a Q. 523 790,76 y Q. 10 146 732,00 respectivamente para el 2018, se considera mediana empresa a toda unidad de producción que realiza actividades de transformación, servicios o comercio, que tiene entre 81 y 200 trabajadores y ventas anuales de entre 3 701 a 15 420 salarios mínimos no agrícolas, equivalentes a Q. 10 149 474,36 y Q. 42 287 191,2 respectivamente para el 2018.

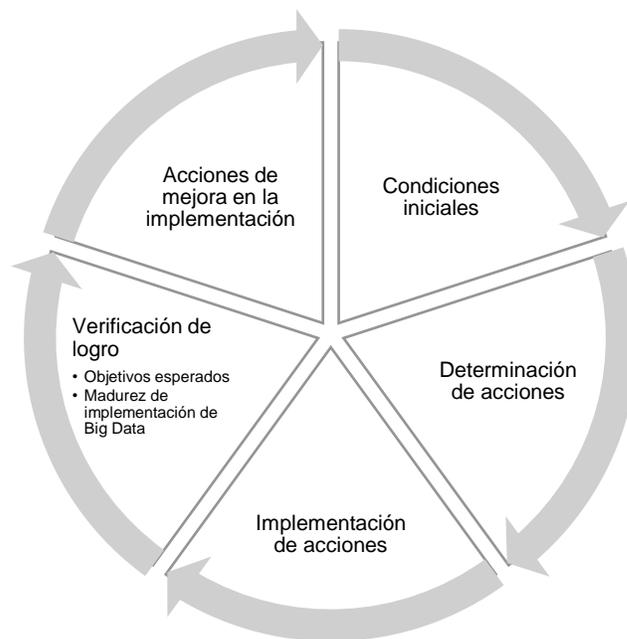
Dada la representatividad que tienen las PYMES en la economía del país de Guatemala, se ofrece a los gerentes de PYMES en Guatemala una propuesta de metodología de implementación de Big Data en planes estratégicos empresariales, acorde a las capacidades tecnológicas y recursos a la mano con el cual generalmente pueden contar.

³⁰ OCSE MIPYME Guatemala. *Sistema nacional de información. Ciudad de Guatemala: Ministerio de Economía de Guatemala, año base 2015.* p.20

Dicha propuesta consiste en la realización de 5 pasos dentro de un proceso gradual y cíclico:

- Condiciones iniciales
- Determinación de acciones
- Implementación de acciones
- Verificación de logro
 - Objetivos esperados
 - Madurez de implementación de Big Data
- Acciones de mejora en la implementación

Figura 2. **Proceso implementación Big Data en planes estratégicos**



Fuente: elaboración propia.

3.1. Condiciones iniciales

El primer paso consiste en establecer el punto de partida para comenzar a implementar una estrategia de Big Data en planes estratégicos, a lo cual conviene realizar un diagnóstico general de la situación de la empresa en cuanto a dirección estratégica, características de los datos utilizados, capacidades y tecnología utilizada para transformar datos en valor empresarial, y para el efecto se procede a realizar una evaluación que indique el nivel de madurez de implementación de Big Data en que se encuentra la empresa en diversos ámbitos relacionados con la transformación de los datos en valor, considerando como prioridad mayor el valor estratégico que se está dando a los datos.

Figura 3. **Niveles de madurez en ámbitos relacionados con la transformación de los datos en valor**



Fuente: elaboración propia.

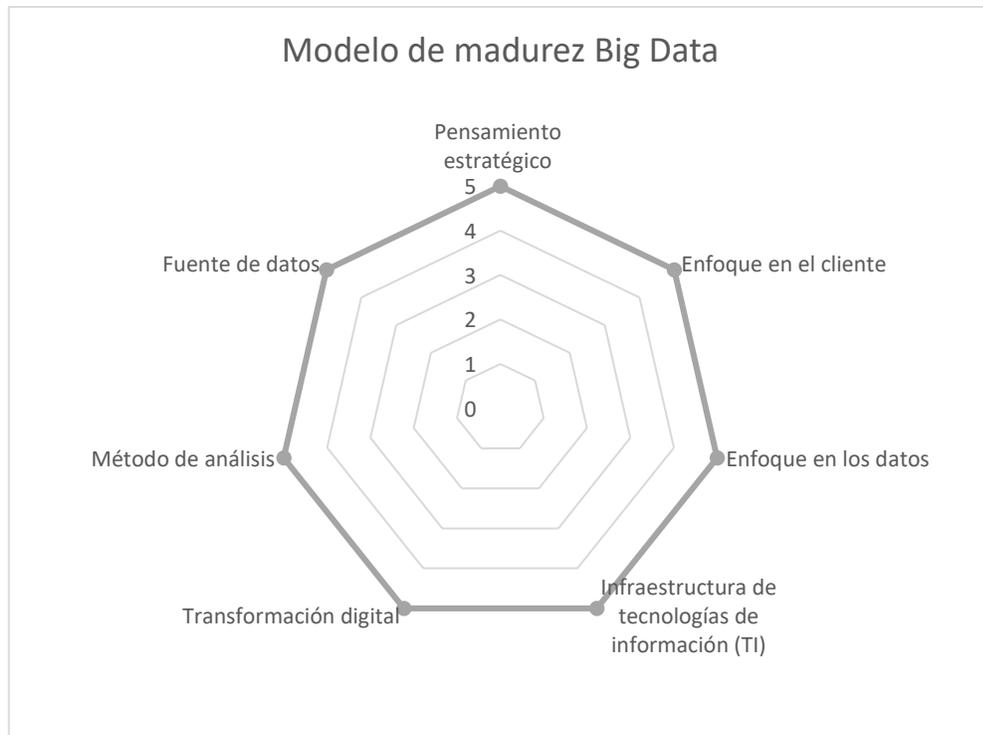
Para contar con una herramienta de decisión estratégica se obtiene una calificación ponderada entre los valores de 1 a 5, la cual indica el nivel de madurez de Big Data en que se encuentra la empresa, acorde a la identificación de la empresa en los distintos niveles de madurez que mejor describen su situación en las distintas áreas de revisión, pudiendo obtener como resultado un gráfico descriptivo de la condición en la que se encuentra la empresa, estableciendo un punto inicial para la implementación de Big Data en la misma.

Tabla I. **Identificación de distintos niveles de madurez en distintas áreas de revisión en modelo de madurez de Big Data**

Áreas de madurez Big Data	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel Establecido
Pensamiento estratégico					x	5
Enfoque en el cliente					x	5
Enfoque en los datos					x	5
Infraestructura de tecnologías de información (TI)					x	5
Transformación digital					x	5
Método de análisis					x	5
Fuente de datos					x	5
PROMEDIO						5

Fuente: elaboración propia.

Figura 4. **Modelo de madurez Big Data**



Fuente: elaboración propia.

3.1.1. **Pensamiento estratégico**

La evaluación de una cultura de pensamiento estratégico en los gerentes y directivos, brinda el panorama sobre el cual se está trabajando por lograr el alcance de la realización de la visión empresarial, a lo cual debe existir un plan estratégico actualizado y vigente que a su vez esté siendo debidamente puesto en operación por los mandos medios.

3.1.1.1. **Nivel 1: a la deriva**

Acorde a si a la mayoría de las siguientes preguntas la respuesta es afirmativa: ¿No se tiene una visión y misión de la empresa?, ¿No se cuenta con

plan estratégico empresarial?, ¿Se tiene más de un año de no evaluar el avance de cumplimiento del plan estratégico empresarial?, ¿Nunca o muy pocas veces se han hechos esfuerzos concretos de crear valor estratégico en la empresa?, ¿Existen incoherencias entre lo que se pretende sea la empresa y lo que está siendo?, ¿Se puede afirmar que realmente no se está compitiendo en el mercado?

3.1.1.2. Nivel 2: desenfoque

Acorde a si existiendo un plan estratégico empresarial, a la mayoría de las siguientes preguntas la respuesta es afirmativa: ¿No existe claridad en la diferenciación de los productos?, ¿Existen múltiples direcciones en el curso de la empresa?, ¿Existen múltiples formas de comprender el producto?, Cuando se logra alguna ventaja sobre la competencia ¿dura poco?, ¿No se contempla realizar mejoras en competitividad?

3.1.1.3. Nivel 3: aspiración

Acorde a si existiendo un plan estratégico empresarial, a la mayoría de las siguientes preguntas la respuesta es afirmativa: ¿Se ha llegado a un punto de contar con muchos productos, pero no se han creado líneas de productos?, ¿No se ha identificado una priorización de estrategias?, ¿No se tiene identificado un producto líder?, ¿No se cuenta con metas de mejora competitiva a largo plazo?

3.1.1.4. Nivel 4: limitación

Acorde a si existiendo un plan estratégico empresarial, y adicionalmente se han identificado estrategias priorizadas para la mejora en competitividad, a la mayoría de las siguientes preguntas la respuesta es afirmativa: ¿Se realizan

acciones y se toman decisiones que no están alineadas a las estrategias priorizadas?, para llevar a cabo las estrategias priorizadas ¿Se considera que no se cuenta con los recursos?, ¿Se considera que no se cuenta con la tecnología?, ¿Se considera que no se cuenta con el conocimiento suficiente?, de una u otra forma ¿Se consideran limitantes o barreras?

3.1.1.5. Nivel 5: propuesta de valor consistente

Acorde a que existe un plan estratégico empresarial, y el mismo se está ejecutando con éxito, evidenciando resultados en el reconocimiento de una propuesta de valor, con una diferenciación bien definida y cuenta con las capacidades y recursos para llevarla a cabo.

3.1.2. Enfoque en el cliente

La evaluación de la importancia que tiene el cliente para los gerentes y directivos, brinda el panorama sobre el enfoque que tienen las operaciones en relación con la atención de las necesidades de los clientes y si se está dando respuesta oportuna a dichas necesidades.

3.1.2.1. Nivel 1: indiferencia

Acorde a si a la mayoría de las siguientes preguntas la respuesta es afirmativa: ¿Se da mayor atención al producto que al cliente?, ¿Procura tener mayores volúmenes de venta con muchos descuentos?, ¿Los mayores esfuerzos están en la reducción de costos?, ¿Nunca o pocas veces se pregunta sobre las necesidades de los clientes?, ¿Se puede afirmar que se tienen de los precios más bajos?

3.1.2.2. Nivel 2: diferenciación

Acorde a si a la mayoría de las siguientes preguntas la respuesta es afirmativa: se tiene un producto único, pero ¿Nunca o pocas veces se pregunta sobre las necesidades de los clientes?, ¿Nuestros clientes nos buscan por tener productos únicos?, ¿Identificamos claramente que nos diferencia de la competencia?, ¿La competencia regularmente iguala las características de nuestros productos?, ¿No contamos con información clasificada de los clientes?, ¿Hay poco retorno de los clientes?

3.1.2.3. Nivel 3: relación establecida

Acorde a si a la mayoría de las siguientes preguntas la respuesta es afirmativa: ¿Los clientes reconocen algo diferente en los productos?, ¿Se tiene buen retorno de clientes?, ¿Se lleva registro de productos que los clientes buscan y no encuentran en nuestra empresa?, ¿No se cuenta con alianzas estratégicas?, ¿No se ofrece amplitud de productos que los clientes buscan?, ¿No se realiza *outsourcing*?, ¿Se tiene establecida una estrategia de atención al cliente?

3.1.2.4. Nivel 4: acogida

Acorde a si a la mayoría de las siguientes preguntas la respuesta es afirmativa: ¿Se ofrece variedad de productos alrededor de las necesidades de los clientes?, ¿Existe opción de tercerización para atender las demandas de los clientes?, ¿Se cuenta con aliados para atender las necesidades de los clientes?, ¿Se ofrece atender necesidades de los clientes en un único punto?, ¿No somos líderes en la participación de mercado?, ¿Se tiene competencia cercana?

3.1.2.5. Nivel 5: dominio

Acorde a si a la mayoría de las siguientes preguntas la respuesta es afirmativa: ¿Se es líder en la participación de mercado?, ¿Se tiene posibilidad de definir tendencia sobre nuestros productos?, ¿Se tiene canales de distribución únicos?, ¿Se identifica complementadores que enriquezcan la oferta?, ¿Se tiene control sobre barreras de entrada para la competencia?

3.1.3. Enfoque en los datos

La evaluación de la importancia que la empresa en todas sus áreas, brinda a los datos como recurso importante, establece un punto de partida para la conversión de dichos datos en valor estratégico y el nivel de aprovechamiento que se está dando de los mismos.

3.1.3.1. Nivel 1: resistencia

Acorde a si se puede describir que se tiene la idea de que siempre se han hecho las cosas de la misma manera y todo funciona bien, y se considera que no se tiene necesidad o capacidad de evolucionar por la misma razón.

3.1.3.2. Nivel 2: consciencia

Acorde a si se puede indicar que la empresa considera que cuenta con datos que son importantes para su actuar, pero se reconoce que no se ha generado valor con los mismos.

3.1.3.3. Nivel 3: análisis

Acorde a si se puede indicar que la empresa cuenta con datos que son importantes para la empresa y de alguna forma se realiza análisis de los mismos, pero no lo suficiente para dar respuesta a interrogantes que se quedan sin solución.

3.1.3.4. Nivel 4: comprensión estratégica

Acorde a si se puede indicar que la empresa presta atención a los datos de forma estratégica, descubriendo el porqué de estos datos mediante investigación interna, extrayendo *insights*.

3.1.3.5. Nivel 5: dirección por datos

Acorde a si se puede indicar que la empresa mediante el uso de los datos que posee, tiene la capacidad de contestar a preguntas sobre el futuro de sí misma en todos los niveles de su organización.

3.1.4. Infraestructura de tecnologías de información (TI)

La evaluación de la infraestructura de tecnologías de información (TI) existente en la empresa brinda información sobre las capacidades existentes de la misma y el nivel de tecnología implementada en la misma para agregar valor a la información. La misma se evalúa mediante la implementación de plataformas tecnológicas para la automatización de los procesos empresariales.

3.1.4.1. Nivel 1: inicial

Acorde a si no se cuenta con infraestructura de tecnologías de información para la operación de la empresa o se tiene una implementación de la misma muy limitada.

3.1.4.2. Nivel 2: aplicaciones de negocio

Acorde a si se cuenta con infraestructura de tecnologías de información para la operación de la empresa, la administración y planificación de los recursos (ERP), la gestión de la relación con los clientes (CRM), la gestión de la cadena de suministro (SCM), operaciones en línea (e-Business), pero de alguna forma dicha infraestructura tecnológica se encuentra independiente o desintegrada.

3.1.4.3. Nivel 3: gestión de procesos

Acorde a si se cuenta con infraestructura de tecnologías de información que automatizan la operación de la empresa de una forma integrada, ofrecen opción de modelado del negocio y monitoreo de las operaciones permitiendo reconocer flujos automáticos en los procesos.

3.1.4.4. Nivel 4: gestión del negocio

Acorde a si se cuenta con infraestructura de tecnologías de información que automatizan la operación de la empresa para los flujos en los procesos y permiten realizar inteligencia de negocios, monitorear y gestionar el rendimiento, gobernanza, riesgo y cumplimiento de forma automatizada.

3.1.4.5. Nivel 5: innovación

Acorde a si se cuenta con infraestructura de tecnologías de información completa que automatice operación, procesos, monitoreo, control, y ofrezca apoyo automatizado en la proactividad y anticipación a los escenarios futuros en los que se encuentre la empresa, permitiendo gestionar la innovación de forma automatizada.

3.1.5. Transformación digital

La evaluación del nivel de madurez en la implementación de transformación digital en la empresa brinda información sobre como la empresa toma en cuenta las oportunidades de mejora de experiencia del cliente con el aprovechamiento de las ventajas que brinda la tecnología.

3.1.5.1. Nivel 1: inmovilismo

Acorde a si la empresa opera bajo las condiciones de mejorar el tamaño y eficiencia más que la atención del cliente, no se corren riesgos de innovación tecnológica, no se realizan esfuerzos de emprendimiento hacia la mejora de la experiencia de cliente y no existe una urgencia de cambio a nivel general.

3.1.5.2. Nivel 2: activación

Acorde a si se identifica que dentro de la empresa existen iniciativas de digitalización, posiblemente aisladas o personales, las cuales corren riesgos en la experimentación de uso de la tecnología para la mejora enfocada a la experiencia de cliente, aun sin contar con el apoyo de los ejecutivos.

3.1.5.3. Nivel 3: proyecto

Acorde a si existen proyectos formales de implementación de acciones para el aprovechamiento de oportunidades en la mejora de la experiencia del cliente mediante el apoyo de la tecnología.

3.1.5.4. Nivel 4: estrategia

Acorde a si se considera que el cliente es una prioridad para la alta dirección, y la mejora de la experiencia del cliente mediante el uso de la tecnología está considerada en el plan estratégico y operativo con responsabilidades definidas en todas las áreas de la empresa para la implementación de la transformación digital.

3.1.5.5. Nivel 5: innovación

Acorde a si se puede reconocer en los diversos niveles de la empresa que se tiene una digitalización en su operación como apoyo a la propuesta de valor empresarial, teniendo definidos roles y procesos para la innovación constante enfocada a la mejor experiencia de cliente mediante la utilización de las ventajas que brinda la tecnología.

3.1.6. Método de análisis

La evaluación del método con el que se está o no analizando los datos brinda un panorama sobre la información que se está generando para la toma de decisiones estratégicas.

3.1.6.1. Nivel 1: carencia

Acorde a si se puede indicar que la empresa no realiza observaciones de los datos existentes, ni busca explicaciones de los sucesos en la operación de la empresa.

3.1.6.2. Nivel 2: descripción

Acorde a si se puede indicar que en el actuar de la empresa, con los datos existentes se realiza análisis que brinda respuestas descriptivas del pasado y del presente, pero no llega a determinar causalidad de los hechos.

3.1.6.3. Nivel 3: diagnóstico

Acorde a si se puede indicar que en el actuar de la empresa, con los datos existentes, se determina razones y causas por las cuales está ocurriendo el presente, sin analizar posibles escenarios futuros.

3.1.6.4. Nivel 4: predicción

Acorde a si se puede indicar que en el actuar de la empresa, con los datos existentes se realiza operaciones para determinar el resultado probable de eventos futuros, contando con identificación de patrones de comportamiento.

3.1.6.5. Nivel 5: análisis prescriptivo

Acorde a si se puede indicar que en el actuar de la empresa, con los datos existentes se realiza la identificación de oportunidades para incrementar la probabilidad de ocurrencia del escenario que tenemos previsto que brindará el

mayor beneficio y se determinan acciones estratégicas para lograr la realización de dicho escenario.

3.1.7. Fuente de datos

La evaluación del origen o fuente de información de la cual la empresa hace uso, brinda un indicador del aprovechamiento que está realizando de la diversidad de orígenes de datos que se encuentran en el mundo digital.

3.1.7.1. Nivel 1: datos registrados

Acorde a si se puede indicar que los únicos datos con los que cuenta la empresa, son los registrados de forma tradicional dentro de la organización en algún medio de almacenamiento digital de forma consolidada o dispersa.

3.1.7.2. Nivel 2: datos no registrados

Acorde a si se puede indicar que adicional a los datos registrados dentro de la empresa, se han identificado datos que no se registran ni almacenan y ya se ha establecido una digitalización de los mismos.

3.1.7.3. Nivel 3: datos de terceros

Acorde a si se puede indicar que adicional a que la empresa cuente con datos registrados propios y digitalizados los no registrados propios, cuenta con acuerdos de compartición con empresas aliadas o proveedores con los que se ha identificado el poder compartir dicha información.

3.1.7.4. Nivel 4: datos abiertos

Acorde a si se puede indicar que adicional a que la empresa cuente con datos registrados propios, digitalizados los no registrados propios, datos de terceros, se cuenta con identificación, acceso y se hace uso de datos abiertos o datos gratuitos que son relevantes para la empresa.

3.1.7.5. Nivel 5: datos no tradicionales

Acorde a si se puede indicar que se está utilizando toda fuente de datos interna como toda fuente de datos externa, incluyendo la obtención de autorizaciones y acuerdos precisos para recabarlos, así como el establecimiento de adquisición de todo tipo de datos necesarios.

3.2. Determinación de acciones

El segundo paso consiste en establecer concretamente una estrategia de implementación de Big Data dentro del plan estratégico de la empresa y su correspondiente plan de acción para la implementación con base en el resultado del diagnóstico realizado al evaluar el nivel de madurez propuesto. La estrategia de implementación a definir responde al objetivo de obtener mayor competitividad en la propuesta de valor empresarial mediante la obtención de valor estratégico de la información obtenida por los datos con que cuenta la empresa, por lo que se debe tener presente que la implementación de Big Data es el medio para lograr dicho objetivo y no el fin.

Las soluciones de Big Data siempre están alineadas con los requisitos de negocio y con base en dichos requisitos adaptan el uso de los datos, capacidades y tecnología existentes, a fin de respaldar la oportunidad de negocio para

incrementar la competitividad. La implementación de Big Data como parte una estrategia conlleva una actualización progresiva a lo largo del tiempo.

3.2.1. Características

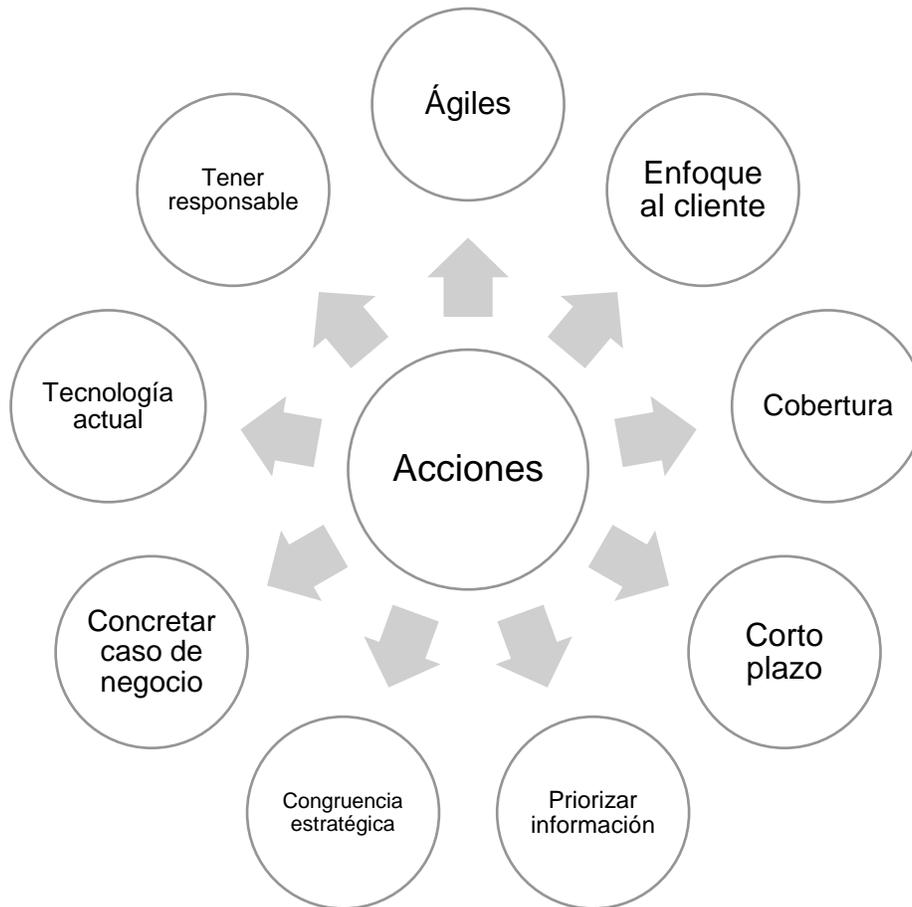
Acorde al nivel de madurez de Big Data en que se encuentre la empresa, para establecer acciones que le permitan mejorar dicho nivel, se recomienda tener presente:

- Se debe plantear acciones bajo el principio de agilidad en su implementación, aprovechando las condiciones de facilidad de implementación de cambios dado el tamaño de la empresa.
- Los esfuerzos a realizar siempre deben estar centrados en el cliente y en el enfoque de la propuesta de valor hacia la mejora de la experiencia del cliente, priorizando la escucha activa del mismo.
- Las acciones a incluir dentro del plan deben tener amplitud de cobertura a modo de concretarse en los diversos niveles de la empresa y en todas las secciones de la misma.
- Se debe buscar lograr resultados a corto plazo con el aprovechamiento de los datos, capacidades y tecnología existente a la vez que se va realizando mejora en dichas condiciones. Los resultados deben ser medibles mediante indicadores acorde al beneficio que se desea obtener.
- Se debe buscar el mejor beneficio y aprovechamiento de los datos existentes y accesibles de forma inmediata, bajo el principio de que menos

es más, priorizando la información que se puede incidir de forma más directa en la toma de decisiones estratégicas.

- El desarrollo de las acciones debe mantenerse acorde a las prioridades estratégicas de la empresa a manera de mantener congruencia estratégica.
- Establecer concretamente un caso de negocio sobre el cual se pueda cuantificar el beneficio del logro del mismo como caso de éxito.
- Se debe tomar en cuenta las tendencias tecnológicas teniendo presente el bajo costo que muchas de ellas ofrecen, y manteniendo el enfoque en la prioridad de los datos sobre la tecnología.
- Se debe nombrar un responsable de Big Data que lidere la implementación, acorde a las condiciones de la empresa, pudiendo ser una persona o un equipo de trabajo, tomando en cuenta que el principal líder en la implementación es el gerente general.

Figura 5. **Características de las acciones para implementar Big Data**



Fuente: elaboración propia.

3.2.2. Oportunidades

Para fortalecer la toma de decisiones en el planteamiento de la estrategia de implementación de Big Data y las acciones a realizar, se puede tomar en cuenta las oportunidades y beneficios que se pueden obtener con un mayor nivel de madurez en la implementación de Big Data que genere mayor valor a la información extraída de los datos, pudiendo focalizar la estrategia y acciones a implementar acorde al aprovechamiento de una de dichas oportunidades y de

forma paulatina en el tiempo, se puede ir ampliando la cobertura de aprovechamiento de oportunidad. Entre las oportunidades que ofrece Big Data se tiene:

- Obtener mayor conocimiento de los clientes para satisfacer de forma más adecuada su demanda con base en la anticipación de sus reacciones y comportamiento, conociendo de forma más específica respuestas a interrogantes como ¿Qué promociones los hacen reaccionar?, ¿Qué los hace comprar?, ¿Cómo prefieren comprar?, ¿Qué factores hacen que una persona recomiende una empresa?
- Determinar tendencias del mercado mediante la identificación de patrones y comportamiento del mismo, conociendo de forma más específica respuestas a interrogantes como ¿Cuándo cambiara la demanda de nuestros productos?, ¿Qué activa la variación de demanda de los productos que ofrecen?
- Conocer los movimientos de la competencia mediante la posibilidad de realizar *benchmarking* más minucioso, conociendo de forma más específica respuestas a interrogantes como ¿Qué está promocionando la competencia?, ¿Qué contenido está compartiendo la competencia en sus redes sociales?
- Optimizar las operaciones de empresa mediante la posibilidad de utilizar los datos recabados en los procesos de producción y operación e implementar mejoras en los mismos, conociendo de forma más específica respuestas a interrogantes como ¿En dónde se encuentran los cuellos de botella?, ¿En qué maquinaria de la línea de producción se tienen pérdidas

de material?, ¿En qué procedimientos de atención al cliente se tienen mayores demoras?

- Mejorar la gestión de talento humano mediante la identificación de mejores canales para la realización de atracción y retención de empleados, conociendo de forma más específica respuestas a interrogantes como ¿Qué personal posee los mejores índices de productividad?, ¿Qué beneficios brinda el desarrollo de personal?, ¿Qué indicadores de satisfacción del personal posee la empresa?

3.3. Implementación de acciones

El tercer paso consiste en la puesta en ejecución de las acciones planteadas tomando en cuenta que dentro de las recomendaciones realizadas es realizar un caso de negocio.

3.4. Verificación de logro

El cuarto paso consiste en realizar periódicamente acciones de verificación de que la implementación y ejecución de acciones definidas dentro de la estrategia de implementación de Big Data estén logrando los objetivos esperados, lo cual se realiza de forma más objetiva mediante la revisión de indicadores de cumplimiento del plan estratégico empresarial.

Es importante recordar que los indicadores de logro deben cumplir con características entre las cuales destacan:

- Deben estar enfocados a la estrategia empresarial.

- Deben cumplir con ser cuantitativos al expresarse a través de una cifra y ser cuantificables al incrementarse o reducirse.
- Deben reflejar resultados de impacto y relevancia para la toma de decisiones.
- Deben brindar información oportuna y requerida para la toma de decisiones, dado que pocos indicadores pueden no ser suficientes y muchos indicadores pueden incrementar el trabajo administrativo de control.
- Deben permitir medir el desempeño de la empresa en comparación con la competencia.

Los indicadores de logro a revisar con mayor atención van acorde a la oportunidad de negocio que se decidió aprovechar mediante la implementación de Big Data.

Tabla II. **Ejemplos de indicadores a evaluar oportunidad de mejora Big Data**

Oportunidad de mejora	Ejemplos de indicadores a evaluar mejora
Obtener mayor conocimiento de los clientes	<ul style="list-style-type: none"> - Porcentaje de clientes satisfechos - Porcentaje de personas que recuerdan la marca - Promedio de visitas por clientes - Promedio de días de estancia por pasajero - Porcentaje de productos comprados por nuevo uso con relación a compra total (en unidades o en valores monetarios)

Continuación de la tabla II.

<p>Determinar tendencias del mercado</p>	<p>Promedio de ventas del producto por mercado (en unidades o en valores monetarios) Promedio de ventas por canal de distribución Porcentaje de cobertura geográfica por región</p>
<p>Conocer los movimientos de la competencia</p>	<p>Porcentaje de clientes captados de la competencia Porcentaje de ventas de clientes leales Promedio de compras de clientes leales</p>
<p>Optimizar las operaciones de empresa</p>	<p>Tiempo promedio de atención de un pedido Tiempo promedio del ciclo de producción Porcentaje de uso de la capacidad instalada Número de productos no conformes Porcentaje de productos devueltos respecto de los productos entregados</p>
<p>Mejorar la gestión de talento humano</p>	<p>Índice de clima laboral Número de trabajadores con planes de carrera Porcentaje de trabajadores que reconocen las características de la cultura organizacional. Número de proyectos especiales para desarrollar empleados con alto potencial Porcentaje de mejoras implementadas por sugerencias de los trabajadores</p>

Fuente: elaboración propia.

Un segundo medio de verificación de la implementación de Big Data es conocer el incremento gradual del nivel de madurez de implementación, mediante la revisión del resultado de aplicación del modelo de diagnóstico propuesto en el primer paso.

3.5. Acciones de mejora en la implementación

El quinto paso consiste en el análisis e identificación de las causas que provocaron la variación positiva o negativa en la medición del alcance de los logros previstos, y acorde a las conclusiones obtenidas se elabora un plan de acción de mejora que debe incluir acciones dentro de las siguientes categorías:

- Acciones para mantener y reafirmar lo bueno en los casos que los logros obtenidos superaron las expectativas.
- Acciones para mejorar en los casos que los logros alcanzados estén por debajo de los objetivos propuestos, pero se detecta que existen ajustes que se pueden realizar para lograr alcanzarlos.
- Acciones para sustituir en los casos que se ha identificado que los objetivos no fueron alcanzados por causas difíciles de superar o de mayor costo en la implementación de su mejora que el beneficio a obtener.

3.6. Responsable de Big Data

Dentro del proceso de implementación de Big Data el rol que juega la persona o el equipo responsable es sumamente importante, por lo que debe tener la claridad en las funciones que debe realizar para proveer la mayor calidad de información a partir de los datos y el proceso de implementación, mencionando entre sus principales funciones:

- Mantener contacto constante con los clientes para aprender de sus experiencias.

- Buscar mecanismos de proactividad con aliados para ayudar a mantener viables y visibles las soluciones.
- Trabajar con los equipos de investigación y desarrollo para solucionar problemas que surjan en el camino.
- Asegurar que la empresa vaya adquiriendo la cultura de gestión, análisis y aprovechamiento de los datos.
- Extraer constantemente conocimiento e información de los datos, presentando resultados obtenidos de forma que puedan utilizarse en la toma de decisiones estratégicas.
- Utilizar mecanismos para la extracción de valor de los datos a través de la experiencia existente en la empresa.
- Desarrollar y aplicar técnicas, procesos, herramientas y métodos para la implementación de Big Data.
- Mantener la calidad, disponibilidad y seguridad de los datos.
- Crear herramientas visuales para ayudar al personal de la empresa a comprender datos complejos.

3.7. Ética empresarial

Un tema importante que no se debe perder de vista al momento de implementar Big Data es lo que atañe a los valores empresariales, a lo cual la utilización de los datos puede entrar en conflicto con la ética empresarial que se

tenga en la empresa, por lo que se debe mitigar el riesgo de aparición de problemas éticos complejos, donde los temas de mayor impacto son el conocimiento detallado de identidad y a la vez la privacidad de dicha información, la propiedad de los datos y el manejo de la reputación de la empresa en base a la confianza que se genere con el uso de los datos accedidos.

Se da por hecho que la empresa vela por permanecer alineada con los valores empresariales para mantenerse como una organización responsable con un código de ética, por lo que en todo el proceso de implementación se deben establecer criterios para el tratamiento de los datos, los cuales deben estar alineados a los valores empresariales, constituyendo un código de ética en la implementación de Big Data.

4. PROPUESTA DE ARQUITECTURA MÍNIMA PARA IMPLEMENTAR BIG DATA

4.1. Diseño de arquitectura

El proceso para extraer valor estratégico de los datos implica que los datos deben pasar por 3 estados:

- Datos crudos en su mínima unidad sin significado estratégico.
- Información extraída de los datos para la toma de decisiones estratégicas.
- Valor estratégico para la empresa luego de realizado un proceso de toma de decisiones en base a información analizada y conocimiento adquirido.

Figura 6. **Estados de datos en proceso para extraer valor estratégico**

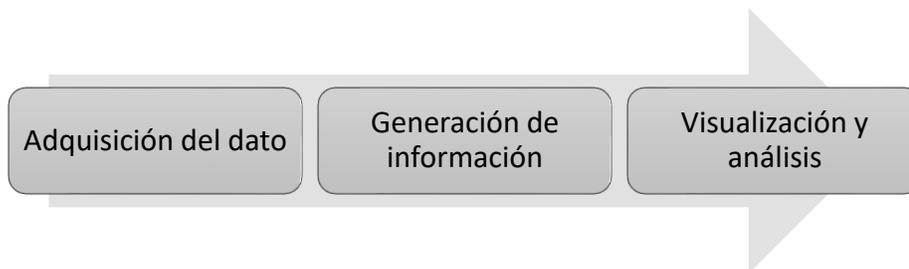


Fuente: elaboración propia.

Dada la transición de estados por las cuales deben pasar los datos, se tienen los procesos correspondientes para poder llevar a cabo las transiciones de estados, siendo:

- Adquisición de dato que corresponde a las formas en que los datos son adquiridos desde su origen, contemplando la variedad de tipos de datos (estructurados, semiestructurados o no estructurados), variedad de fuentes de información (interna o externa). Asimismo contempla el proceso en que estos son almacenados y dispuestos para su acceso y considerarse datos usables.
- Generación de información que corresponde a la forma en que a partir de los datos accesibles, estos son convertidos en información con significado estratégico para la empresa, regularmente mediante la aplicación de algoritmos y técnicas que hacen eficiente este proceso.
- Visualización y análisis que corresponde a la aplicación de herramientas para presentar y visualizar la información y de esta forma realizar el correspondiente análisis para obtener el valor estratégico de los datos, y a partir de ello tomar una decisión estratégica.

Figura 7. **Proceso para extraer valor estratégico de los datos**



Fuente: elaboración propia.

Los procesos descritos pueden darse en fases de procesamiento o en tiempo real, acorde a las necesidades definidas por la estrategia de la empresa.

Tomando como base el modelo de procesos que transforma los datos en valor, se proponen 3 niveles de arquitectura mínima para implementar Big Data en las PYMES en Guatemala, mediante la aplicación de la tecnología que permita realizar cada uno de los procesos del modelo.

La propuesta de arquitectura mínima mantiene los siguientes principios:

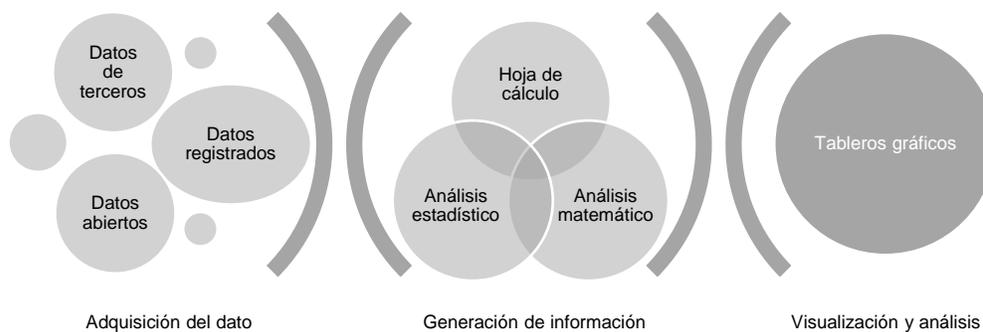
- Se orienta al uso de arquitecturas modulares debido al cambio constante de la tecnología, lo que permite no tener una dependencia en un único producto, aunque se hace mención a que existen productos que, aunque son de uso modular tienen una vinculación con otros productos limitando al uso de una versión determinada o incluir en la exclusión de otra herramienta, por lo que es importante tener presente la estrategia que tiene la empresa.
- Dada la capacidad limitada de inversión que regularmente poseen las PYMES, se mantiene un enfoque de inversión mínima a manera de priorizar la obtención del valor estratégico de los datos y obtener un mayor retorno de inversión a corto plazo.

4.1.1. Nivel Inicial: activación de Big Data

Las PYMES que se encuentran en un nivel nulo o mínimo de madurez de Big Data, es importante que realicen una implementación con el menor esfuerzo posible para obtener resultados que perciban sus beneficios en un corto plazo, por lo que la arquitectura propuesta se fundamenta en la priorización de la

característica de variedad de los datos, así como identificar los datos principales que se tienen disponibles de forma inmediata, los cuales brinden información valiosa que pueda transformarse en valor estratégico para la toma de decisiones en la empresa.

Figura 8. **Activación de Big Data**



Fuente: elaboración propia.

4.1.1.1. **Adquisición del dato**

El mayor esfuerzo a realizar conlleva la identificación de las fuentes de datos apropiadas y los mecanismos para acceder a dichas fuentes, a lo cual se contempla principalmente:

- Datos registrados que la empresa actualmente posee y regularmente es obligatorio su existencia en formato digital, tal como estados financieros, registros de compras y ventas, registros de inventarios, registros de costos de operación, entre otros.
- Datos de terceros en referencia a datos que poseen otras empresas dedicadas a la captación y comercialización de datos, considerando que

algunas proveen acceso a dichos datos de forma gratuita tal como Google mediante *Google trends*, *Google analytics* y *Google correlate*.

- Datos abiertos que son accesibles en el medio guatemalteco, tal como estadísticas en portal web del Instituto Nacional de Estadística, registros de variables macroeconómicas en el portal web del Banco de Guatemala, variables macroeconómicas internacionales en el portal web del Banco Mundial (relevantes en el sector de importación y exportación) entre otras fuentes.

4.1.1.2. Generación de información

Para la generación de información de los datos existentes es posible utilizar un programa de software de hoja de cálculo que seguramente se encuentra dentro de los paquetes de ofimática instalados en la sección administrativa de la empresa ya sea en una versión sin costo o una versión comercial, y en ausencia del mismo la inversión de adquisición de una versión profesional o completa es accesible de invertir en comparación con el valor competitivo que se requiere obtener con el análisis de los datos. Mediante el software de hoja de cálculo se debe aplicar análisis matemático y análisis estadístico sobre los datos adquiridos y esto se puede facilitar mediante la incorporación de componentes gratuitos en el programa de hoja de cálculo que se esté utilizando, generando información estratégica. Para el caso de utilizar microsoft excel se debe conocer que el mismo ofrece complementos como *power pivot*, herramientas de análisis de datos, *solver*, análisis de hipótesis y previsión que permite comparar escenarios, opciones de segmentación de datos y la posibilidad de agregar complementos gratuitos de análisis estadístico.

El análisis estadístico se puede ampliar con el uso de software específico para dicho análisis, existiendo versiones con licenciamiento sin cobro como GNU PSPP.

4.1.1.3. Visualización y análisis

Para la visualización y análisis para la toma de decisiones estratégica se debe crear tableros gráficos que permitan a los directivos visualizar de forma gráfica la información generada a partir de los datos y de esta forma realizar el correspondiente análisis para la toma de decisiones estratégica. Existe diversidad de herramientas accesibles con costo mínimo que pueden ser utilizadas:

- El software de hoja de cálculo utilizado para el análisis seguramente posee la capacidad de generación de gráficos de la información generada. Para el caso de utilizar microsoft excel se debe conocer que el mismo ofrece complementos como *power view* y *power map* adicional a la diversidad de opciones de tipos de gráficos para presentar la información.
- Es posible utilizar herramientas en línea que faciliten la presentación de la información en tableros gráficos, mencionando entre otras *zoho reports* en versión gratuita.

4.1.2. Nivel Básico: Big Data no administrado

Las PYMES que desean implementar Big Data pueden hacer uso de las ventajas que presenta hoy en día el universo del *cloud computing*, en cuanto que mediante la existencia de software como servicio (SaaS), infraestructura como servicio (IaaS) y plataforma como servicio (PaaS), se tiene también el Big Data como servicio (BDaaS), el cual consiste en que mediante todas las capacidades

que brinda el *cloud computing* se ofrece la implementación y ubicación de todos los recursos de Big Data en la nube, permitiendo acceder a sus capacidades como un servicio.

Para las PYMES se recomienda la implementación de Big Data como servicio (BDaaS) en una modalidad no administrada, entendiéndose que no se debe tener preocupación por la administración de los recursos e infraestructura tecnológica requerida por Big Data, lo cual permite orientar y enfocarse en mayor medida a los esfuerzos para la generación de valor estratégico de los datos y convertirlos en una ventaja competitiva, y de esta forma dedicarse esencialmente al aprovechamiento de las ventajas que ofrece.

Aunque existen varios proveedores para implementar Big Data como servicio, para implementar de una forma no administrada se recomienda la oferta de Google, dada la amigabilidad e integración de sus componentes.

Figura 9. **Big data no administrado**



Fuente: elaboración propia.

4.1.2.1. Adquisición del dato

El componente que permite realizar el almacenamiento en el uso de esta arquitectura es *Google bigquery* con una capacidad de almacenamiento de petabytes a un bajo costo, brindando una facilidad de uso para almacenamiento de variedad de datos, tanto en su forma estructurada, semiestructurada y no estructurada, como en su procedencia interna o externa a la empresa, de forma que los datos que son generados por sistemas informáticos en la empresa pueden consolidarse al ser enviados a dicho servicio tanto como los datos que sean obtenidos de forma externa a la empresa. En caso de requerir un movimiento de datos más avanzado, para traslado de datos y procesos de extracción y transformación es posible utilizar *Google dataflow*, el cual es un servicio de procesamiento de datos que facilita la realización de dichas tareas.

Se recomienda aprovechar la integración y compatibilidad de datos de terceros, en este caso Google como fuente de datos, mediante el uso de:

- *Google trends*: ofrece la posibilidad de comparar la popularidad de búsqueda de varias palabras o frases (*keyword*) durante un período de tiempo determinado, permitiendo la detección de tendencias de demanda de los clientes.
- *Google analytics*: ofrece información detallada del tráfico de clientes que acceden a un sitio web, teniendo gran relevancia información demográfica de los clientes que visitan el sitio web de la empresa, como mayores páginas web más visitadas de dicho portal web, a lo cual es posible saber de los clientes lo que les gusta y lo que no, dispositivo utilizado (smartphone, tablet o computadora de escritorio), procedencia desde

donde encontraron el sitio web de la empresa en internet, y lo que gusta y que no de lo la empresa ofrece en su sitio web.

- *Google correlate*: ofrece información sobre correlaciones de varias palabras o frases (*keyword*) ofreciendo la posibilidad de obtener información comparativa de cuántas veces se ha buscado una determinada palabra en el sitio web de la empresa y en internet en general, para saber cuáles son los gustos o intereses de los clientes o saber cuál es el sentimiento de ellos respecto a los productos de la empresa, así como identificar productos relacionados a los productos ofrecidos por la empresa.

4.1.2.2. Generación de información

Para la generación de información de los datos existentes es posible utilizar *Google cloud datalab*, la cual es permite explorar, analizar y visualizar datos para obtener información y aprendizaje a partir de los datos.

Si es necesario realizar un análisis más profundo se puede realizar consultas específicas mediante el uso de lenguaje de consulta de datos (SQL) con la información almacenada en *Google bigquery* de manera que se realiza un *bigquery analytics*.

El análisis estadístico que brinde información a partir de los datos se puede realizar mediante el uso del software libre y sin costo para estadística llamado proyecto R, que es posible conectar a *Google bigquery* mediante librerías para el efecto.

En caso de requerir realizar un análisis de datos predictivo es posible utilizar Google *prediction* API, la cual ofrece capacidades de aprendizaje automático para detección de patrones, de manera que es posible predecir productos que podrían gustarle a un usuario, valorar si datos no estructurados como comentarios publicados en la web son positivos o negativos, o estimar lo que podría gastar un usuario un día específico.

4.1.2.3. Visualización y análisis

Para la visualización y análisis para la toma de decisiones estratégica es posible crear tableros de mando e informes dinámicos que permitan a los directivos visualizar de forma gráfica la información generada a partir de los datos y de esta forma realizar el correspondiente análisis para la toma de decisiones estratégica, mediante el uso de Google Data Studio.

4.1.3. Nivel maduro: Big Data administrado

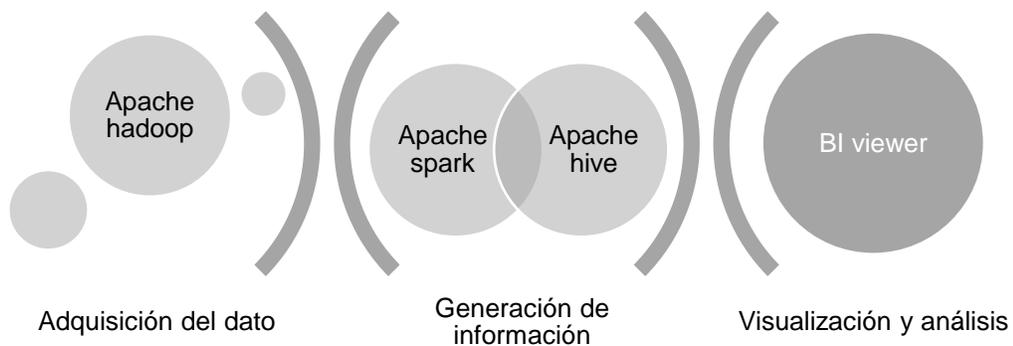
Las PYMES que tienen un plan estratégico que contempla mayor escalabilidad y proyectan a futuro que las capacidades de implementar Big Data como servicio (BDaaS) de una forma no administrada les ofrece una solución limitada en el uso de características particulares, pueden aventurarse a implementar un Big Data administrado mediante el cual pueden tener a su alcance la administración de características específicas que brinda la tecnología de Big Data a manera de personalizar su solución a la medida de las necesidades de la empresa, mediante el aprovechamiento de las ventajas del *cloud computing*, implementando de una forma más directa la característica de Infraestructura como servicio (IaaS) y plataforma como servicio (PaaS).

Implementar Big Data como servicio (BDaaS) de una forma administrada, requiere conocimiento específico de la tecnología y configuración que hay detrás de esta arquitectura tecnológica, existiendo a la fecha una configuración madura y mayormente utilizada, la cual se puede implementar dentro de una infraestructura como servicio (IaaS).

Los proveedores que permiten realizar esta implementación son Microsoft mediante su plataforma Azure, Google mediante su plataforma *Google Cloud Platform* y Amazon mediante su plataforma *Amazon Web Services*.

Una diferencia clave para concretar la arquitectura de Big Data a implementar es la característica de necesidad de manipulación de los datos en tiempo real, por lo que es importante tener claridad en la línea estratégica de la empresa, y si realmente es necesaria la implementación de Big Data en tiempo real.

Figura 10. **Big Data administrado**



Fuente: elaboración propia.

4.1.3.1. Adquisición del dato

Para gestionar la adquisición y almacenamiento de los datos se utiliza *apache hadoop*, el cual es un sistema de código abierto que se utiliza para almacenar, procesar y analizar grandes volúmenes de datos, teniendo la capacidad de almacenar variedad de datos estructurados, semiestructurados y no estructurados, mediante el almacenamiento en un sistema de archivos distribuidos (HDFS).

4.1.3.2. Generación de información

Para generar información se utilizan diferentes herramientas:

- *Apache spark*: es un sistema de procesamiento distribuido de código abierto que utiliza almacenamiento caché en memoria y ejecución optimizada para ofrecer un desempeño rápido, a su vez compatible con procesamiento por lotes.
- *Apache hive*: permite realizar agrupación, consulta, y análisis de datos de *apache hadoop*.

4.1.3.3. Visualización y análisis

Para la visualización y análisis para la toma de decisiones estratégica es posible crear tableros de mando e informes dinámicos que permitan a los directivos visualizar de forma gráfica la información generada a partir de los datos y de esta forma realizar el correspondiente análisis para la toma de decisiones estratégica, mediante el uso de herramientas de inteligencia de negocios que se conecten a la infraestructura de datos creada, existiendo diversidad en el

mercado tanto en versión de pago como gratuitas, mencionando *qlink*, *tableau*, *pentaho*, entre otras.

CONCLUSIONES

1. Con base en la revisión de diversos conceptos existentes de Big Data se logró ofrecer un concepto propio que abarca la integración de características, capacidades y tecnologías existentes en la empresa.
2. El objetivo principal de Big Data más allá de brindar tecnologías para el tratamiento de altos volúmenes de datos, es convertir los datos en valor que pueda ser utilizado por las empresas para mejorar su competitividad.
3. Se estableció un recurso para alfabetización digital en tema de Big Data que puede ser utilizado como insumo para la concepción de proyectos de Big Data por ejecutivos en Guatemala.
4. Se creó una propuesta de metodología de implementación de Big Data en planes estratégicos empresariales que puede ser utilizada por los ejecutivos tomadores de decisiones en PYMES en Guatemala, dado que considera implementaciones desde 0.
5. Se definieron 3 propuestas de implementación de Big Data mediante la utilización de arquitectura tecnológica mínima, accesible para PYMES en Guatemala.

RECOMENDACIONES

1. Definir estrategias ágiles y concretas que se alineen a la estrategia empresarial sobre el uso de la tecnología más reciente, dada la amplitud de herramientas que existen para la implementación de Big Data.
2. Los ejecutivos en Guatemala deben prestar atención al valor de los datos para mejorar su competitividad y de esta forma incrementar la economía del país.
3. Los ejecutivos en Guatemala deben realizar esfuerzos necesarios para una adecuada transformación digital en las empresas que tome en cuenta la transformación de los datos en valor estratégico.
4. La ciudadanía debe brindar seguimiento a la implementación en el sector gobierno de mecanismos de acceso a la información por medios digitales y libres, mediante datos abiertos que puedan ser utilizados para el correspondiente análisis que extraiga valor a los datos que tiene bajo su resguardo.
5. Los expertos en tecnología en Guatemala deben realizar esfuerzos y establecer mecanismos para que las empresas en Guatemala puedan acceder a los beneficios de Big Data de una forma alcanzable acorde a sus posibilidades, permitiendo mejorar el nivel de competitividad a nivel regional e internacional.

BIBLIOGRAFÍA

1. ANDERSON, Chris. *The end of theory: the data deluge makes the scientific method obsolete*. Wired. [en línea] <<https://www.wired.com/2008/06/pb-theory/>> [Consulta: 11 de noviembre de 2017].
2. BARKER, Dan. *The Real Original Source of the Phrase Big Data*. Dan Barker. [en línea] <<http://barker.co.uk/bigdata>> [Consulta: 22 de noviembre de 2017].
3. BRYANT, KATZ y LAZOWSKA. *Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society*. Estados Unidos: Computing Community Consortium committee of the Computing Research Association, 2008. 7 p.
4. CARTER, Philip. *Big Data Analytics: Future Architectures, Skills and Roadmaps for the CIO*. Estados Unidos: IDC Go-to-Market Services, 2011. 15 p.
5. COX, Michael y ELLSWORH, David. *Application-Controlled Demand Paging for Out-of-Core Visualization*. Estados Unidos: NASA Ames Research Center, 1997. 12 p.
6. DANA BOYD, Kate Crawford. *Critical Questions for Big Data*. Estados Unidos: Information, Communication & Society, 2012. 18 p.

7. DE MAURO, Andrea, GRECO, Marco y GRIMALDI, Michele. *A formal definition of Big Data based on its essential features*. 2016. Library Review. 65. 122-135. 10.1108/LR-06-2015-0061.
8. Dell EMC. *Transforming IT Study Survey Results*. [en línea] <<https://www.emc.com/campaign/global/forum2013/survey.htm>> [Consulta: 8 de noviembre de 2017].
9. ————. *EMC Forums 2013 Transforming IT Global Study: Global Results Infographic*. [en línea] <<https://www.emc.com/infographics/forum2013-survey.htm>> [Consulta: 8 de noviembre de 2017].
10. ————. *Even Though 79 % Think Big Data Will Improve Decision Making, One-Third Have No Big Data Plans*. [en línea] <<https://www.emc.com/about/news/press/2013/20131212-01.htm>> [Consulta: 8 de noviembre de 2017].
11. DEMCHENKO, Yuri, DE LAAT, Cees y MEMBREY, Peter. *Defining architecture components of the Big Data Ecosystem*. 2014. 104-112. 10.1109/CTS.2014.6867550.
12. EUROPA PRESS. *Especialistas destacan que el Big Data 2.0 hace crecer más de un 30 % a las empresas de comercio electrónico*. 20 Minutos. [en línea] <<https://www.20minutos.es/noticia/2487714/0/especialistas-destacan-que-big-data-2-0-hace-crecer-mas-30-empresas-comercio-electronico/>> [Consulta: 11 de noviembre de 2017.]

13. Gartner. *Big Data*. Gartner. [en línea] <<https://www.gartner.com/it-glossary/big-data/>> [Consulta: 11 de noviembre de 2017].
14. GUALTIERI, Mike. *The Pragmatic Definition Of Big Data*. Forrester. [en línea] <<https://go.forrester.com/blogs/12-12-05-the-pragmatic-definition-of-big-data/>> [Consulta: 11 de noviembre de 2017].
15. IDG. *Data & Analytics Survey 2016*. [en línea] <<https://www.idg.com/tools-for-marketers/tech-2016-data-analytics-research/>> [Consulta: 8 de noviembre de 2017].
16. KPMG International Cooperative. *Going beyond the data turning data from insights into value*. Estados Unidos: KPMG International Cooperative, 2015. 32 p.
17. LANEY, Doug. *Application Delivery Strategies*. Estados Unidos: META Group Inc, 2001. 4 p.
18. LARSON, Erik. *How Did They Get Your Name? Direct-mail Firms Have Vast Intelligence Network Tracking Consumers*. [en línea] <http://articles.orlandosentinel.com/1989-07-26/lifestyle/8907254531_1_subscribe-to-magazines-subscriber-list-junk-mail> [Consulta: 22 de noviembre de 2017].
19. NewVantage Partners. *Big Data Executive Survey 2017. Executive Summary of Findings*. Estados Unidos: NewVantage Partners, 2017. 16 p.

20. NOEL YUHANNA, Brian Hopkins. *TechRadar™: Big Data, Q1 2016*. Estados Unidos: Forrester, 2016. 39 p.
21. OCSE MIPYME Guatemala. *Sistema nacional de información*. Ciudad de Guatemala: Ministerio de Economía de Guatemala, 2015. 30 p.
22. OECD. *Skills for a digital world*. Cancún, México: OECD, 2016. 58 p.
23. ORELLANA CÁMBARA, Maritryny. *Cumplimiento de las disposiciones y recomendaciones internacionales de la organización mundial de aduanas, caso: contrabando en la frontera la mesilla del departamento de Huehuetenango, Guatemala*. Trabajo de graduación de Licenciatura en Relaciones Internacionales. Universidad de San Carlos de Guatemala, Escuela de Ciencia Política, 2016. 77 p.
24. OSORES, Melisa. *Crece Big Data en AL en 2015, impulsada por software y servicios*. TechTarget. [en línea] <<http://searchdatacenter.techtarget.com/es/cronica/Crece-big-data-en-AL-en-2015-impulsada-por-software-y-servicios>> [Consulta: 8 de noviembre de 2017].
25. PENN, Christopher. *The Evolution of the Data-Driven Company*. [en línea] <<http://www.christopherspenn.com/2016/11/the-evolution-of-the-data-driven-company/>> [Consulta: 8 de noviembre de 2017].
26. PÉREZ ARBESÚ, Lizzette. *Prioridades de TI para las empresas de América Latina en 2017*. TechTarget. [en línea] <<http://searchdatacenter.techtarget.com/es/cronica/Prioridades-de->

TI-para-las-empresas-de-America-Latina-en-2017> [Consulta: 8 de noviembre de 2017].

27. RocaSalvatella. *Cultura Digital y Transformación de las Organizaciones*. Barcelona, España: RocaSalvatella, 2014. 50 p.
28. SHAFER, Tom. *The 42 V's of Big Data and Data Science. Elder Research*. [en línea] <<https://www.elderresearch.com/company/blog/42-v-of-big-data>> [Consulta: 11 de noviembre de 2017].
29. TYNAN, Dan. *The skills and traits of a next-generation CIO*. [en línea] <<https://www.cio.com/article/3207542/leadership-management/the-skills-and-traits-of-a-next-generation-cio.html>> [Consulta: 8 de noviembre de 2017].
30. World Economic Forum. *The Global Competitiveness Report 2017–2018*. Suiza: World Economic Forum, 2017. 381 p.

