



FACULTAD DE POSGRADOS

ANÁLISIS DE MODELOS DE NEGOCIO BASADOS EN BIG DATA
PARA OPERADORES MÓVILES

AUTOR

Juan Esteban Gangotena Gomezjurado

AÑO

2018



FACULTAD DE POSGRADOS

**ANÁLISIS DE MODELOS DE NEGOCIO BASADOS EN BIG DATA PARA
OPERADORES MÓVILES**

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

Profesor Guía
MBA. Jaime Augusto Vinuesa Trujillo

Autor
Juan Esteban Gangotena Gomezjurado

Año
2018

DECLARACIÓN DEL PROFESOR GUÍA

“Declaro haber dirigido este trabajo, Análisis de modelos de negocio basados en big data para operadores móviles, a través de reuniones periódicas con el maestrante Juan Esteban Gangotena Gomezjurado, en el semestre 2018-2, orientando sus conocimientos y competencias para un eficiente desarrollo del tema escogido y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación”

Jaime Augusto Vinueza Trujillo
Master of Business Administration
C.I. 1716028509

DECLARACIÓN DEL PROFESOR CORRECTOR

“Declaro haber revisado este trabajo, Análisis de modelos de negocio basados en big data para operadores móviles, del estudiante Juan Esteban Gangotena Gomezjurado, en el semestre 2018-2, dando cumplimiento a todas las disposiciones vigentes que regulan los trabajos de titulación”

Leonardo Bolivar Cornejo Gonzalez
Master of Business Administration
C.I. 1704886678

DECLARACIÓN DE AUTORIA DEL ESTUDIANTE

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

Juan Esteban Gangotena Gomezjurado
C.I. 1717096794

AGRADECIMIENTO

Agradezco a Dios y a mi familia por el apoyo diario, pilar fundamental en la consecución de este objetivo, a todas las personas que me dedicaron tiempo y conocimiento para la realización de este trabajo y en especial al Ing. Jaime Vinueza, MBA por su apoyo como guía de este trabajo.

DEDICATORIA

El esfuerzo invertido en este trabajo lo dedico a Dios, a mis Padres por brindarme su sabiduría, sus valores y don del trabajo y a mis hermanas por el apoyo diario e incondicional fundamental para la culminación de este objetivo.

RESUMEN

El presente trabajo de titulación explica cuales son los beneficios de adoptar Big Data en operadores de telecomunicaciones moviles, propone un marco de referencia basado en las mejores practicas de la industria que proporciona una visión amplia, un esquema de elementos relacionados que apoya un enfoque particular de un determinado objetivo y sirve como guía para la adopción de Big Data en operadores de telecomunicaciones moviles.

La industria de telecomunicaciones se ha convertido en un ecosistema altamente competitivo con muchos desafíos, donde los operadores de telecomunicaciones moviles han venido teniendo varias dificultades para innovar e incursionar en nuevos modelos de negocio, han perdido fuerza con nuevas y más grandes empresas digitales como WhatsApp, Skype, Facebook. Incorporar Big Data incrementará la capacidad para establecer nuevos modelos de negocio que permitan recuperar competitividad frente a nuevos jugadores dentro del ecosistema de las telecomunicaciones.

Los operadores de telecomunicaciones moviles que han adoptado Big Data como concepto y estrategia y están incrementando la, lealtad de sus clientes, descubriendo nuevas oportunidades de negocio, realizando predicciones con gran exactitud, mejorando y optimizando sus procesos internos.

ABSTRACT

The present titration work explains what are the benefits of adopting Big Data in mobile telecommunications operators. It proposes a reference frame based on industry best practices, which provides a broad vision, a scheme of related elements, which supports a particular approach to a given objective and serves as a guide for the adoption of Big Data in mobile telecommunications operators.

The telecommunications industry has become an ecosystem of highly competitive companies with many challenges, where mobile telecommunications operators are presenting several difficulties at the moment of innovating and enter into new business models. These companies have lost strength against new and larger digital companies such as WhatsApp, Skype and Facebook. Through the incorporation of Big Data, it will be possible to increase the capacity to establish new business models that will allow recovering competitiveness against new players within the telecommunications ecosystem.

The mobile telecommunications operators that have adopted Big Data as a concept and strategy are increasing the loyalty received from their customers, discovering new business opportunities, making predictions with great accuracy, improving and optimizing their internal processes.

ÍNDICE

| | |
|--|----|
| 1. INTRODUCCIÓN..... | 1 |
| 1.1. Antecedentes | 1 |
| 1.2. Justificación | 4 |
| 1.3. Objetivos y Alcance | 6 |
| 2. MARCO TEÓRICO | 7 |
| 2.1. Big Data..... | 7 |
| 2.1.1. Definición | 7 |
| 2.1.2. Atributos..... | 8 |
| 2.1.3. Tipos de Datos..... | 11 |
| 2.1.4. Analítica Avanzada (<i>Advanced Analytics</i>)..... | 14 |
| 2.1.5. Minería de datos | 15 |
| 2.1.6. Big Data en tiempo real | 20 |
| 2.2. Industria de telecomunicaciones | 21 |
| 2.2.1. Evolución de las telecomunicaciones | 22 |
| 2.2.2. Modelo actual de negocio | 26 |
| 2.2.3. Cadena de valor..... | 27 |
| 2.2.4. Problemas de sustentabilidad del modelo actual..... | 29 |
| 2.3. Marco de referencia..... | 33 |
| 3. BIG DATA Y BIG DATA ANALYTICS..... | 33 |
| 3.1. <i>Big Data</i> & Operadores de telecomunicaciones móviles..... | 34 |
| 3.1.1. Oportunidades de Big Data..... | 37 |
| 3.1.2. Volumen..... | 38 |
| 3.1.3. Variedad | 41 |
| 3.1.4. Velocidad | 42 |
| 3.1.5. Nuevas estrategias de los operadores de telecomunicaciones móviles..... | 42 |
| 3.2. Repositorios de datos | 43 |
| 3.2.1. Business Intelligence Tradicional..... | 44 |
| 3.2.2. <i>Data Lake</i> | 46 |

| | |
|---|----|
| 3.3. Framework Mapreduce y Hadoop | 47 |
| 3.3.1. Ecosistema <i>Hadoop</i> | 49 |
| 3.3.2. <i>Hadoop Distributed File System</i> (HDFS) | 50 |
| 3.3.3. <i>Hbase</i> | 51 |
| 3.3.4. <i>Map Reduce</i> | 52 |
| 3.3.5. <i>Apache Pig</i> | 54 |
| 3.3.6. <i>Apache Hive</i> | 55 |
| 3.3.7. <i>Apache Avro</i> | 56 |
| 3.3.8. <i>Apache ZooKeeper</i> | 57 |
| 3.3.9. <i>Apache Chukwa</i> | 58 |
| 3.4. Almacenamiento NoSQL..... | 58 |
| 3.5. <i>Big Data</i> & Internet de las cosas (IOT) | 60 |
| 3.6. <i>Big Data</i> en tiempo real | 63 |
| 3.7. <i>Big Data</i> & <i>Cloud Computing</i> | 65 |
| 4. MODELOS DE NEGOCIO | 71 |
| 4.1. Casos de uso modelo actual | 71 |
| 4.2. Casos de uso del nuevo modelo | 72 |
| 4.2.1. Monetización de los datos | 72 |
| 4.2.2. Publicidad Móvil (<i>Mobile Advertising</i>) | 75 |
| 4.2.3. Exposición de datos del cliente..... | 77 |
| 5. MARCO DE REFERENCIA | 78 |
| 5.1. Evaluación del caso comercial | 80 |
| 5.1.1. Fase inicial..... | 81 |
| 5.1.2. Fase de planificación | 81 |
| 5.1.3. Fase de diseño | 81 |
| 5.2. Capa de Ingestión de datos..... | 81 |
| 5.2.1. Funcionalidades claves en la ingestión de datos..... | 82 |
| 5.3. Procesamiento de datos..... | 83 |
| 5.3.1. Transformación..... | 84 |
| 5.3.2. Correlación | 85 |
| 5.3.3. Enriquecimiento | 85 |

| | |
|---|-----|
| 5.3.4. Aseguramiento de la calida de datos | 85 |
| 5.4. Capa de analisis de datos | 86 |
| 5.4.1. Modelado descriptivo, predictivo, prescriptivo | 87 |
| 5.4.2. Procesamiento de eventos complejos (CEP)..... | 91 |
| 5.5. Visualizaci3n de datos | 92 |
| 5.6. Repositorio de datos vertical | 103 |
| 5.6.1. Datos no estructurados..... | 104 |
| 5.6.2. Datos semiestructurados | 104 |
| 5.6.3. Datos estructurados..... | 104 |
| 5.7. Gobernanza de datos..... | 105 |
| 5.7.1. gesti3n de la privacidad | 109 |
| 5.7.2. Cumplimiento | 113 |
| 5.7.3. gesti3n de la seguridad | 115 |
| 6. CONCLUSIONES Y RECOMENDACIONES..... | 117 |
| 6.1. Conclusiones..... | 117 |
| 6.2. Recomendaciones | 121 |
| REFERENCIAS | 124 |
| ANEXOS | 130 |

ÍNDICE DE TABLAS

| | |
|--|-----|
| Tabla 1. Atributos Big Data..... | 9 |
| Tabla 2. Procesos se coordinan mediante el uso de la DG "hoja de ruta" | 108 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1. Uso de datos en empresas para análisis. | 12 |
| Figura 2. Pasos del proceso de descubrimiento del conocimiento..... | 17 |
| Figura 3. Estructura jerárquica de un árbol | 18 |
| Figura 4. Red Neuronal feedforward de 4 capas, dos de ellas ocultas. | 19 |
| Figura 5. Reglas derivadas de un conjunto de elementos..... | 20 |
| Figura 6. Cadena de valor de internet. | 28 |
| Figura 7. Canibalización de ingresos OTT. | 31 |
| Figura 8. Disminución de ganancia total de la industria de telecomunicaciones. ... | 32 |
| Figura 9. Exabytes per Month of Mobile Data Traffic by 2021..... | 39 |
| Figura 10. Global Mobile Data Traffic Forecast by Region. | 40 |
| Figura 11. Global Mobile Traffic by Connection Type..... | 41 |
| Figura 12. Componentes de una solución BI..... | 45 |
| Figura 13. Framework MapReduce and Hadoop..... | 49 |
| Figura 14. Arquitectura de HDFS. | 51 |
| Figura 15. Aplicación de MapReduce para obtener palabras de un texto. | 54 |
| Figura 16. Expectativa de objetos conectados al 2020. | 61 |
| Figura 17. Hype Cycle del Internet de las cosas. | 62 |
| Figura 18. Overview of the Lambda Architecture. | 65 |
| Figura 19. Modelos de cloud computing..... | 70 |
| Figura 20. Casos de uso Big Data Analytics. | 73 |
| Figura 21. Telefónica Smart Steps. | 75 |
| Figura 22. Crecimiento del mercado global de segmentos seleccionados..... | 76 |
| Figura 23. Marco de referencia Big Data Analytics. | 79 |
| Figura 24. Fase de integración de datos. | 83 |
| Figura 25. Big Data Analytics Events, Streams, Feeds y CEP..... | 91 |
| Figura 26. Relación de una variable con otra y tendencias a lo largo del tiempo. 93 | |
| Figura 27. Aumento o disminución del valor inicial en una serie. | 94 |
| Figura 28. Diagrama de dispersión para visualizar la relación en los datos..... | 95 |
| Figura 29. Diagrama de burbujas..... | 95 |

| | |
|---|-----|
| Figura 30. Un gráfico de pastel para comparar los porcentajes. | 96 |
| Figura 31. Diagrama de caja que compara la distribución de puntos de datos. ... | 97 |
| Figura 32. Nube de palabras para representar la frecuencia alta o baja. | 98 |
| Figura 33. El diagrama de red para visualizar datos semi y no estructurados | 99 |
| Figura 34. Velocidad de procesamiento en las organizaciones..... | 100 |
| Figura 35. Visualización con gráficos relacionales. | 101 |
| Figura 36. Árbol de decisión, datos segmentados de acuerdo a ramificación.... | 102 |
| Figura 37. Gráfico de matriz. | 103 |
| Figura 38. Datos de gobernanza. | 107 |
| Figura 39. Big Data privacidad la banca ante nuevos desafíos éticos. | 111 |

1. INTRODUCCIÓN

1.1. Antecedentes

Actualmente las telecomunicaciones se han convertido en un factor fundamental en el desarrollo, económico, social y humano, es ahí en donde los operadores de telecomunicaciones móviles se han establecido como un papel fundamental para ello. Los operadores de telecomunicaciones móviles son empresas que impulsan las comunicaciones a través de redes inalámbricas y recursos tecnológicos sobre el cual el operador presta un servicio de telecomunicación. Los servicios más comunes de telefonía celular: servicio de llamadas de voz, servicios de mensajes cortos SMS (*Short Message Service*, Servicio de Mensajes Cortos). La telefonía celular es la transmisión de voz y SMS a través de redes inalámbricas, que permite que dos personas intercambien mensajes y voz de forma instantánea y a pesar de que la telefonía celular en un principio fue diseñada solo para la transmisión de voz, debido a las limitaciones tecnológicas de esa época, hoy en día es posible brindar muchos más servicios como datos, audio y video a través de la tecnología actual.

Gracias al largo camino recorrido desde hace varios años el sector de las telecomunicaciones se ha convertido en un ecosistema altamente competitivo con muchos desafíos, con esto los operadores de telecomunicaciones móviles están viviendo una auténtica revolución que implica una transformación hacia una empresa de telecomunicaciones digital que satisfaga todas las necesidades digitales de sus clientes, desde una conexión a internet hasta servicios digitales desde cualquier dispositivo, descubriendo y anticipando las necesidades de sus clientes. Una empresa de telecomunicaciones digital internamente apuesta por desarrollar el internet de las cosas, *Big Data*, digitalización, desarrollo de contenidos, *cloud*, *hardware*, *software*, conectividad, servicios, etc.

Google, *Samsung*, *Apple* se han convertido en verdaderos revolucionarios en la industria ya que han demostrado grandes capacidades para el desarrollo en este

ecosistema. Con esto los operadores de telecomunicaciones móviles se han visto afectados con la aparición de nuevos modelos de negocio y nuevos competidores.

Según lo establecido por *World Economic Forum & Accenture* (2017), desde hace varios años que los operadores de telecomunicaciones móviles no han tenido éxito en la protección de sus tradicionales fuentes de ingresos como servicios de voz y mensajería. Estimaciones han demostrado que las aplicaciones que utilizan red de datos de un tercero para comercializar sus servicios, Aplicaciones OTT (*Over-The-Top*, Capa superior) generan un 50% a un 90% menos de ingresos para los operadores. Si bien el aumento exponencial del consumo de datos ha proporcionado cierto alivio, esto no ha sido suficiente para superar la disminución constante del ARPU (*Average Revenue Per User*, Ingreso promedio por usuario). La presión competitiva de los precios y la interrupción del OTT han ido acompañadas de un rápido aumento de los costos, ya que las demandas de ancho de banda y velocidad han seguido creciendo exponencialmente.

Los operadores de telecomunicaciones móviles han venido teniendo varias dificultades para innovar e incursionar nuevos modelos de negocio. Los Proveedores de servicios de comunicaciones CSP (*Communications Services Providers*) han concentrado todos sus esfuerzos para abastecerse de la mejor tecnología, estándares e interoperabilidad, pero han perdido fuerza con nuevas y más grandes empresas digitales como WhatsApp, Skype, Facebook, que cada vez más están logrando agresivamente capacidades de comunicación digital. La participación de los operadores en el sector de la industria se ha reducido del 58% en 2010 al 47% en 2015 y se prevé que caiga aún más, hasta el 45% en 2018 según el *World Economic Forum & Accenture* en el 2017.

Las organizaciones de alto rendimiento como China Mobile, Vodafone o Telefónica ya están utilizando la información en tiempo real desde la visión de *Big Data* para mejorar el compromiso con el cliente. Estos operadores de telecomunicaciones móviles se han permitido comprender el comportamiento de sus clientes y anticipar

sus necesidades de comunicación para convertirlas en ventas a través del uso de soluciones en *Big Data*.

El término *Big Data* es considerado por varios autores como un nuevo paradigma en el uso estratégico de los datos para explotar grandes volúmenes de datos en busca de obtener información valiosa para la toma de decisiones. Los operadores de telecomunicaciones móviles pueden empezar a utilizar *Big Data* para comprender el comportamiento de clientes para crear ofertas y experiencias personalizadas a través de mejores canales para individuos, anticipar y satisfacer las necesidades de clientes y convertirlas en ventas.

Big Data tiene la capacidad de convertir datos de cualquier tipo de fuente, sea estructurada, semi estructurada o no estructurada, en información valiosa que permita transformar el negocio como por ejemplo utilizar los datos de navegación de clientes para poder adaptar y personalizar sus servicios, innovar en productos y servicios, así como tomar mejores decisiones en tiempo real para eliminar competidores en el mercado. *Big Data* también ha sido definida por la clasificación de cuatro características, más conocidas como las cuatro “V”: volumen, velocidad, variedad y valor. La presencia de una o varias de estas características son los que determinan y justifican la adopción de un proyecto *Big Data* en una empresa.

Al incorporar *Big Data* los operadores de telecomunicaciones móviles incrementarán significativamente en cuanto a su capacidad de entender a sus consumidores y permitirá abrir nuevas fuentes de ingresos, como la venta de datos agregados o análisis de datos, publicidad e investigación de mercado, modelos de negocio electrónicos como *e-Commerce*, *e-Advertising* entre otros.

Con esto, *Big Data* se convierte en una alternativa; operadores de telecomunicaciones móviles como Telefónica, Verizon, T-Mobile, Vodafone ya lo han adoptado como concepto y estrategia y están incrementando la lealtad de sus

clientes, descubriendo nuevas oportunidades de negocio, realizando predicciones con gran exactitud, mejorando y optimizando sus procesos internos.

Esto representa un gran reto para los operadores móviles, *Big Data* abre grandes posibilidades para proteger su rentabilidad y hacer crecer sus ingresos con una presión competitiva mayor, mejorar la eficiencia operativa y adaptarse a marcos regulatorios cada vez más exigentes.

Establecer un marco de referencia permitirá consolidar las propuestas, teóricas y mejores prácticas relacionadas con la adopción e implementación de *Big Data* en empresas de telecomunicaciones de tal manera que se pueda aprovechar sus beneficios para incursionar en nuevos productos y servicios, y mejorar las ofertas aprovechando la potencialidad de la interacción con los clientes.

1.2. Justificación

Este trabajo de investigación permitirá explicar descriptivamente cuáles son los beneficios de adoptar soluciones *Big Data* en operadores móviles, lineamientos para aprovechar los datos puede maximizar el valor de los datos generados en los sistemas informáticos y de red, para que mediante soluciones de *Big Data* sean aprovechados y de esta manera establecer nuevos modelos de negocio que permitan recuperar la competitividad frente a nuevos jugadores dentro del ecosistema de las telecomunicaciones.

Según la guía de *Big Data Analytics del TM Forum* (junio, 2017), la adopción de *Big Data* en los operadores de telecomunicaciones móviles implicaría beneficios para cada una de estas áreas:

- Estrategia
 - Ofertas en tiempo real adaptadas a las necesidades del cliente.
 - Promover el uso de servicios adicionales.

- Cartera de productos
 - Mejoramiento de productos.
 - Analizar nuevos productos.
- Operaciones
 - Proporcionar atención proactiva.
 - Reparación de red, basadas en experiencia de cliente.
- Rentabilidad
 - Predecir Comportamiento de *churn* (Ver Anexo 2).
 - Asegurar ingresos.
 - Detectar Fraude.

TM Forum menciona “creación de nuevos productos y servicios con la obtención de nuevos conocimientos en varias ramas de la industria de las telecomunicaciones, ha permitido gestionar de una manera más eficiente la información, eliminar trámites burocráticos, gestión de información para promover promociones concretas a los clientes, darles un servicio más personalizado, reducir riesgos de fraude y de pérdida de ingresos, de esta manera las empresas de telecomunicaciones podría actuar como generador de cambio, ya que tiene la capacidad de generar nuevas oportunidades de crecimiento”.

El presente trabajo de investigación responderá las siguientes preguntas:

- ¿Cómo pueden los operadores de telecomunicaciones móviles apalancarse en soluciones de Big Data para mejorar su Modelo de negocio actual?
- ¿Qué ventajas competitivas ofrecen las soluciones basadas en Big Data a los operadores de telecomunicaciones móviles sobre sus nuevos competidores?

- ¿Cómo pueden los operadores de telecomunicaciones móviles adoptar nuevos modelos de negocio basados en *Big Data*?

Finalmente, el resultado de este trabajo llevará a presentar un marco de referencia que permita ser una guía para la adopción de conocimiento y mejores prácticas aplicadas a los nuevos modelos de negocio apalancados en *Big Data*.

1.3. Objetivos y Alcance

Objetivo General

Desarrollar un marco de referencia para la adopción de conocimiento y mejores prácticas de nuevos modelos de negocio para operadores de telecomunicaciones móviles apalancados en *Big Data*.

Objetivos Específicos

- Definir conceptos y soluciones basados en *Big Data*.
- Analizar los modelos de negocio actuales y sus desafíos para operadores de telecomunicaciones móviles.
- Analizar nuevos modelos de negocio basados en *Big Data* para operadores de telecomunicaciones móviles.
- Analizar casos de éxito y mejores prácticas en la industria de telecomunicaciones de soluciones apalancadas en *Big Data*.
- Describir los casos de uso de los nuevos modelos de negocio apalancados en *Big Data* para operadores de telecomunicaciones móviles.

- Desarrollar un marco de referencia basado en las mejores prácticas de la industria de telecomunicaciones para la adopción de nuevos modelos de negocio apalancados en *Big Data*.

2. MARCO TEÓRICO

2.1. Big Data

2.1.1. Definición

“Set de datos cuyo tamaño está más allá de las habilidades típicas del software de datos tradicionales para capturarla, almacenarla, administrarla y analizarla” (Manyika et al., 2011)

En otras palabras, el termino *Big Data* se utiliza para describir el notable crecimiento de datos, este término revela la cantidad y variedad de datos que se están generando hoy por hoy. Es por eso que a partir del 2012 la consultora Gartner agrega otros atributos a esta definición en los cuales habla de aquellos sistemas de información que manejan conjuntos de datos de gran volumen, variedad y velocidad.

“*Big Data* es gran volumen, gran velocidad, gran variedad de activos de información que demandan formas rentables e innovadoras de procesamiento de información para mejorar la comprensión y la toma de decisiones.” (Gartner, 2012).

Los autores (Sagiroglu & Sinanc, 2013) concluyen que “*Big Data* es un concepto para denominar a aquellos datos masivos, que se caracterizan por su calidad, variedad y complejidad, dificultando consecuentemente su almacenamiento, análisis y visualización para la consecución de resultados”.

Finalmente (Díaz & Zaki, 2015) postulan que “*Big Data* debe ser apreciado principalmente por el volumen de datos, su variedad, velocidad o flujo de creación y procesamiento, la veracidad o calidad de la información y, sobre todo, por el valor que entregan a las organizaciones, transformándose en activo estratégico en sí mismo. También los autores en su trabajo se refieren al tipo de datos que actualmente se generan, señalando que en su mayoría estos son del tipo estructurado, lo cual, impacta directamente a la forma en que las empresas los interpretan y analizan”.

En conclusión, *Big Data* no es solo una herramienta es un habilitador para la toma de decisiones basada en análisis predictivos, que viene marcando una nueva era de integración de información, que trata de la integración de un conjunto de tecnologías diseñadas para obtener el máximo rendimiento en grandes volúmenes de información. *Big Data* tiene la posibilidad de brindar nuevos conocimientos no solo del cliente, si no de disciplinas como la genética o la astronomía, de asuntos tan importantes como la sanidad pública o la seguridad, y de cuestiones de suma trascendencia para el futuro de la sociedad.

2.1.2. Atributos

Existen muchas definiciones de *Big Data* que están enfocadas en el volumen y el almacenamiento de datos, de esto se puede decir que el volumen es importante, pero también hay que tener en cuenta otros atributos relevantes: “velocidad, veracidad, variedad y valor” (UNIVERSITY OF AMSTERDAM, 2013). Estas características destruyen el paradigma de que *Big Data* es solo volumen.

Cada uno de estos atributos tienen las siguientes características como se muestra en la Tabla 1 (UNIVERSITY OF AMSTERDAM, 2013):

Tabla 1.
Atributos Big Data

| Volumen | Velocidad | Variedad | Veracidad | Valor |
|-----------------------------|----------------|-----------------|---------------------------|---------------|
| Almacenamiento En terabytes | Por lotes | Estructurado | Integridad y Autenticidad | Estadísticas |
| Registros | Tiempo Cercano | No estructurado | Origen y Reputación | Eventos |
| Transacciones | Tiempo Real | Multi-factor | Disponibilidad | Correlaciones |
| Tablas y Archivos | Procesos | Probabilística | Responsabilidad | Hipótesis |

Volumen se refiere a la cantidad de datos. Grandes volúmenes de datos se presentan en *terabytes*, *petabytes*, *exabytes*. Una encuesta realizada en el 2012 por IBM reveló que más de la mitad de los encuestados consideraron que un conjunto de datos de más de un *terabyte* de datos se considera “grande” (Schroeck, Shockley, inteligentes, Romero-Morales, y Tufano, 2012). One almacena *terabytes*, por lo cual todos los datos que encajarían en 1500 *CDs* o *DVDs* 220, suficiente para almacenar alrededor de 16 millones de fotografías de Facebook. Castor, Kumar, Li, Sobel, y Vajgel (2010) informan de que Facebook procesa hasta un millón de fotografías por segundo. Un *petabyte* equivale a 1.024 *terabytes*. Las estimaciones anteriores sugieren que Facebook almacena 260 mil millones de fotos utilizando el espacio de almacenamiento de más de 20 *petabytes*.

“El beneficio ganado de la capacidad de procesar grandes cantidades de información es el mayor atractivo del análisis con Big Data. Tener más datos supera a tener mejores modelos: matemáticas simples pueden ser tremendamente efectivas cuando se aplican a grupos de datos enormes” (Dumbill en O’Reilly: 2012: Ubicación 55 de 2321, Edición de Kindle).

Variedad se refiere a la heterogeneidad estructural en un conjunto de datos. Las empresas generan y utilizan varios tipos de datos estructurados, semi-estructurados y no estructurados. Los datos estructurados, que constituye sólo el 5% de todos los datos existentes (Cukier, 2010), se refiere a los datos que se encuentran en las

hojas electrónicas o bases de datos relacionales. Texto, imágenes, audio y vídeo son ejemplos de datos no estructurados, que a veces carecen de la organización estructural requerida por las máquinas para el análisis.

“Rara vez los datos se presentan de una manera ordenada y listos para su procesamiento. Un tema común en los sistemas de procesamiento de *Big Data* es que las fuentes de los datos son diversas y no caen en la categoría de estructuras relacionales. Podría ser texto proveniente de las redes sociales, información de imágenes, o los datos crudos de un sensor de movimiento. Ninguna de estas cosas viene lista para su integración en un proceso” (Dumbill en O’Reilly: 2012: Ubicación 110 de 2321 Kindle).

El termino velocidad describe la prontitud con que se generan los datos y la capacidad de analizar y actuar con la misma rapidez. La proliferación de dispositivos digitales, como teléfonos inteligentes y sensores ha dado lugar a un ritmo sin precedentes de creación de datos y está conduciendo una creciente necesidad de análisis en tiempo real y la planificación basada en la evidencia. Incluso los minoristas convencionales están generando datos de alta frecuencia. *Wal-Mart*, por ejemplo, procesa más de un millón de transacciones por hora (Cukier, 2010).

“La importancia de la velocidad de los datos, es decir la frecuencia exponencial con la que la información ingresa a una organización genera un “*loop de feedback*” que lleva los datos desde su ingreso hasta las tomas de decisión en mucho menor tiempo” (Dumbill en O’Reilly: 2012: Ubicación 88 de 2321 Kindle).

Veracidad. IBM acuñó veracidad como la cuarta V, que representa la falta de fiabilidad inherente a algunas fuentes de datos. Por ejemplo, los sentimientos de clientes en los medios sociales son inciertos en la naturaleza, ya que implican el juicio humano. Sin embargo, contienen información valiosa. Por lo tanto, Big Data

tiene la capacidad de tratar con datos imprecisos e inciertos y aborda el uso de herramientas y análisis desarrollados específicamente para el análisis y explotación de datos inciertos.

Valor. *Oracle* presentó Valor como un atributo importante en la definición de *Big Data*. Sobre la base de esta definición de *Oracle*, *Big Data* se caracterizan por una “baja densidad de valor”. Esto se debe a que la información que se recibe comúnmente en la forma original, tiene un valor bajo en relación con su volumen. Sin embargo, el verdadero valor de los datos se puede identificar mediante *Big Data* a través del procesamiento y análisis de este tipo de información.

El uso de *Big Data* tiene el potencial de transformar economías y sociedades con su aporte de nuevas formas de productividad, crecimiento y valor agregado para los clientes. La utilización de *Big Data* es un atributo clave para la competitividad y el crecimiento, para esto los líderes deben estar preparados para reconocer las ventajas y también sus amenazas. Los operadores de telecomunicaciones móviles no pueden estar ajenos a esta revolución de los datos.

2.1.3. Tipos de Datos

Dentro de una empresa se genera varios tipos de datos y es habitual dividir en dos grandes categorías de datos provenientes de diferentes fuentes de datos, estas categorías son: estructurados (datos tradicionales) y no estructurados (datos *Big Data*). Sin embargo, con el progreso de nuevas tecnologías y herramientas especializadas para *Big Data* se ha establecido una nueva categoría, ya aceptada y son los datos semiestructurados.

Forrester's Global Business indica que tan solo el 31% de la información no estructurada, 27% de la información semiestructurada y el 40% de la información estructurada es aprovechada dentro de una empresa (Figura 1)

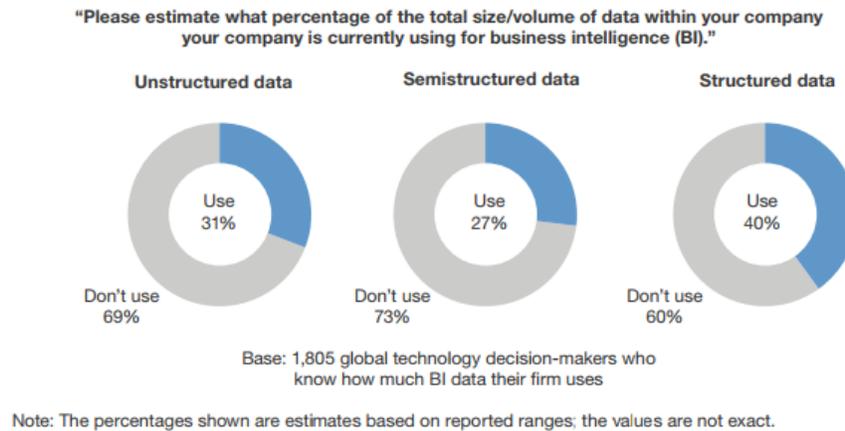


Figura 1. Uso de datos en empresas para análisis.

Tomado de: (Forrester's Global Business Technographics, 2015).

2.1.3.1 Datos estructurados

Los datos estructurados son datos más comunes que se encuentran en bases de datos relacionales y que son fáciles de procesar, y predecible, los datos no estructurados pueden ser texto y que se pueden expresar en filas y columnas con títulos e índices. Este tipo de datos tienen la característica que pueden ser analizados y procesados sencillamente por herramientas de análisis de datos. Estos datos pueden ser internos o externo, donde todo está identificado, etiquetado y tiene fácil acceso.

El dato estructurado interno puede contener información de clientes, recursos operaciones, perfiles web, inventario, finanzas, facturación y operación, que se obtienen de fuentes como: CRM (Client Relationship Management, Gestión sobre la Relación con los Consumidores), ERP (Enterprise Resource Planning, Planificación de recursos empresariales) y es consolidada en un DWH (Data Warehouse, Almacén de datos) con propósitos de reporte y análisis para toma de decisiones.

2.1.3.2. Datos semiestructurados

Los datos semiestructurados son estructuras de datos que pueden ser tratados como estructurados, tienen un flujo lógico y un formato definido pero que aún no es de fácil comprensión para el usuario.

Los datos semiestructurados tienen en cierta manera una estructura implícita pero no tan regular como para poder ser analizada y automatizada como la información estructurada, se refiere a datos que no suelen tener formato fijo, pero tienen etiquetas y otros separadores que permiten separar los datos.

Las herramientas de minería de datos requieren formas complejas para acceder a los datos, la lectura y escritura que se determinan después de acceder con la lectura. Un ejemplo típico de datos semiestructurados son los registros Web Logs de las conexiones a internet, un Web Log se compone de diferentes piezas de información. Etiquetas de lenguajes XML y HTML.

2.1.3.3. Datos no estructurados

Los datos no estructurados son tipos de datos sin estructura definida. Se almacenan como “documentos” u “objetos” sin estructura, y se tiene poco o ningún control sobre ellos. Datos de texto, video, audio, fotografía son datos no estructurados. Por ejemplo, las imágenes se clasifican por su resolución en píxeles. Este tipo de dato que no tiene filas o columnas fijas; ejemplos típicos son: audio, video, fotografía, documentos impresos, cartas, hojas electrónicas, imágenes digitales, formularios especiales, mensajes de correo electrónico y de texto, formatos de texto libre como correos electrónicos, mensajes instantáneos como SMS, artículos, libros, mensajes de mensajería instantánea tipo whatsapp, line, spotbros.

En una empresa el 80% de la información no habita en base de datos relacionales o archivos, sino que se encuentran dispersos a lo largo de toda la empresa. Unos

ejemplos claros de información no estructurada son, llamadas telefónicas, videos, correos electrónicos, etc. Un ejemplo que ilustra la importancia de analizar datos no estructurados son los correos electrónicos, aunque varias herramientas de minería de datos son capaces de analizar el contenido de los correos electrónicos, su recopilación y clasificación puede llegar a ser importante para la empresa, de igual manera el análisis de las llamadas de nuestros clientes podría permitir crear nuevos productos e impulsar la marca.

Evidentemente, los datos no estructurados son más complejos de analizar por los analistas y su continuo incremento ha inducido el apareamiento de nuevas herramientas para su manejo como es el caso de MapReduce, Hadoop o base de datos NoSQL.

2.1.4. Analítica Avanzada (*Advanced Analytics*)

Según Gartner es el examen autónomo o semi-autónomo de datos o contenido utilizando técnicas y herramientas sofisticadas, por lo general más allá de los tradicionales de inteligencia de negocio (BI), para descubrir una visión más profunda, hacer predicciones, o generar recomendaciones. Técnicas analíticas avanzadas incluyen aquellos tales como la minería de datos/texto, aprendizaje de máquina, la coincidencia de patrones, la previsión, la visualización, el análisis semántico, análisis de los sentimientos, la red y el análisis de conglomerados, estadística multivariante, el análisis gráfico, simulación, procesamiento de eventos complejos, las redes neuronales.

Según Zbinden la Analítica avanzada “es el proceso que permite la identificación de patrones y correlaciones significativas entre variables complejas, datos estructurados y no estructurados, históricos, predicción de eventos y evaluación de efectos futuros de las acciones tomadas de grandes volúmenes de datos para la resolución de problemas de negocio” (Geoffrey Zbinden, 2014).

El *advanced analytics* es una manera de descubrir qué ha cambiado y cómo reaccionar ante ese cambio. Es el mejor medio para descubrir nuevos segmentos de clientes, identificar mejores proveedores, asociar un producto por afinidad, comprender la estacionalidad de las ventas, etc. Está compuesta por una colección de técnicas relacionadas y herramientas que normalmente incluyen analítica predictiva, minería de datos, análisis estadísticos y programación compleja de SQL.

2.1.5. Minería de datos

Minería de datos, su traducción al inglés "*Data Mining*", es una técnica que pretende buscar valor en información almacenada en bases de datos, datawarehouse o data mart.

Data Mining según Leventhal "es el proceso de descubrir e interpretar patrones a partir de grandes volúmenes de datos y convertirlos en información con el fin de resolver problemas de negocio y tomar acciones" Leventhal (2010).

Dos operaciones existen en la minería de datos:

- Predecir tendencias y comportamientos
- Anticipar e identificar patrones desconocidos.

La minería de datos es un conjunto de técnicas que tiene como objetivo la búsqueda de conocimiento que utiliza métodos estadísticos, matemáticas, ciencias de la computación, inteligencia artificial y de aprendizaje de máquinas para descubrir patrones en grandes volúmenes de datos.

Esta información incluye patrones normalmente extraídos en un conjunto grande de datos. Estos patrones pueden ser reglas, afinidades, correlaciones, tendencias o modelos de predicción. Dentro de las categorías de minería de datos están:

- Minería web
- Minería de textos
- Descubre texto en archivos

La minería de datos tiene varias aplicaciones como el análisis de los sentimientos que, a través del análisis de mensajes de texto, en post de redes sociales y otras fuentes de datos, permite descubrir la opinión o el sentimiento con respecto a algo.

Los patrones y esquemas resultantes de la minería de datos son aplicados a algoritmos que provienen de la estadística y ciencias de la computación, que se ejecutan sobre un grupo de datos para obtener un resultado con valor.

De la necesidad de obtener información útil a partir de repositorios de datos como bases de datos, datawarehouse o datamarts nace la minería de datos y con el incremento de los volúmenes de datos, la inteligencia de negocio dio paso a que la minería de datos se aplique a soluciones empresariales y comerciales, pues esto permite descubrir información relevante del negocio a partir de multitudes de datos (Ver figura 2).

“En las ciencias y la ingeniería existe un amplio rango de problemas y dominios de aplicación para la minería de datos” (Grossman, Kamath, Kegelmeyer, Kumar, & Namburu, 2013).

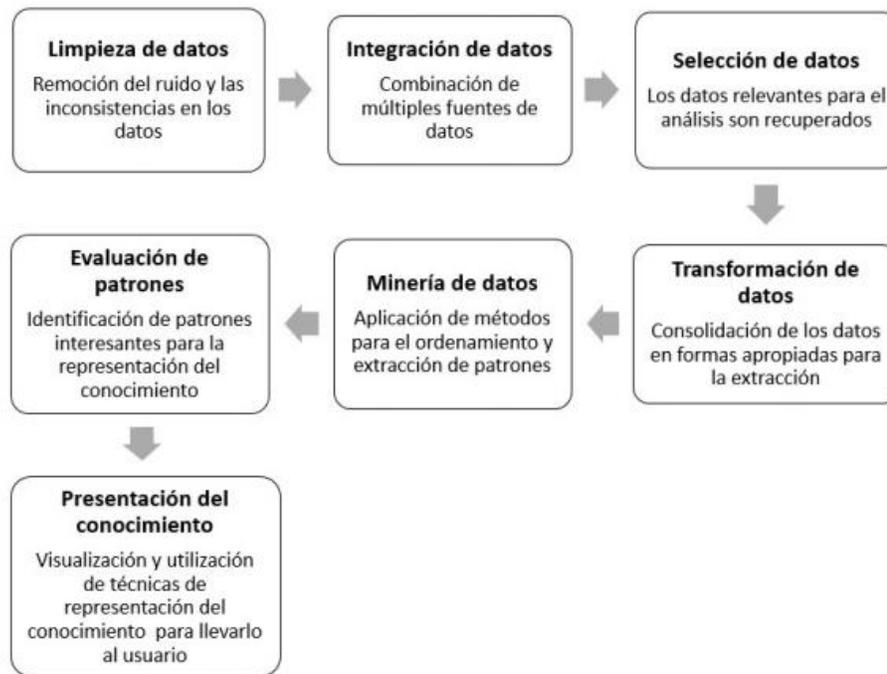


Figura 2. Pasos del proceso de descubrimiento del conocimiento

Tomado de: (Han et al., 2011).

“Un algoritmo de minería de datos es un conjunto de cálculos y reglas heurísticas que permite crear un modelo, el algoritmo analiza los datos proporcionados en busca de tipos de patrones o tendencias. El algoritmo usa los resultados de este análisis para definir los parámetros óptimos para la creación del modelo de minería de datos” (Cadenillas, Minería de Datos, 2011).

Existen algunas técnicas de minería de datos, entre las más representativas tenemos:

2.1.5.1. Árboles de clasificación

Los árboles de clasificación y regresión es la técnica más popular y se construye gracias a un análisis de un conjunto de datos en donde se va separando la población por subgrupos que difieren entre si según algún criterio definido.

Esta técnica parte de un nodo principal o nodo raíz que incluye a toda la población y se van realizando subgrupos dividiendo la población principal con aquella variable más discriminante para cada caso, y esta recibe el nombre de variable predictora, esta variable permite formar dos subgrupos más homogéneos internamente y más heterogéneos entre ellos.

Una rama del árbol deja de dividirse al no haber una variable lo suficientemente discriminante en el nodo (Ver Figura 3).

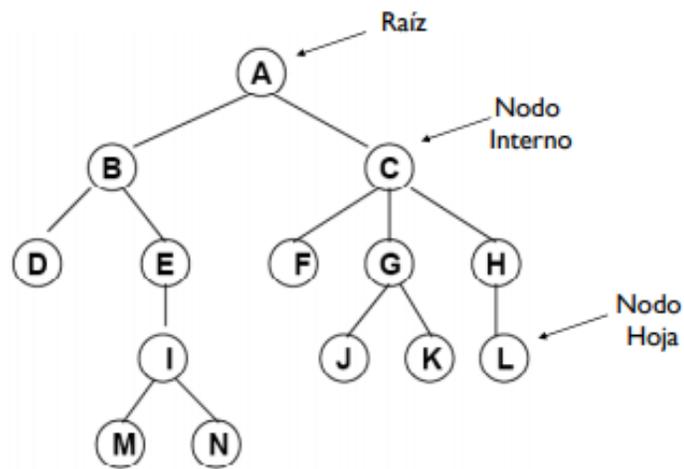


Figura 3. Estructura jerárquica de un árbol

Tomado de: (M.A. Weiss AddisonWesley, 2000).

De esta manera el árbol mínimo tiene dos subgrupos y para continuar con la partición uno de los nodos hasta que haya un número de divisiones con nodos terminales, es decir el árbol termina en el momento de que cada nodo terminal contiene elementos de una única clase, lo que puede llevar a árboles de gran tamaño hasta que no compensa un análisis tan grande por ello se puede eliminar las ramas que no sean necesarias.

2.1.5.2. Redes neuronales

Esta técnica es frecuentemente utilizada para tratar problemas altamente complejos y está formada por neuronas, que se unen unas con otras creando una red cuya extensión se fija a criterio de un experto (Ver Figura 4).

Las redes neuronales y el cerebro presentan varias semejanzas como:

- Entender y aprender tareas basadas en experiencia.
- Auto-organización, aprende y organiza la información que recibe.
- Tolerancia a fallos cuando se daña una parte de la red.
- Trabajos en tiempo real.

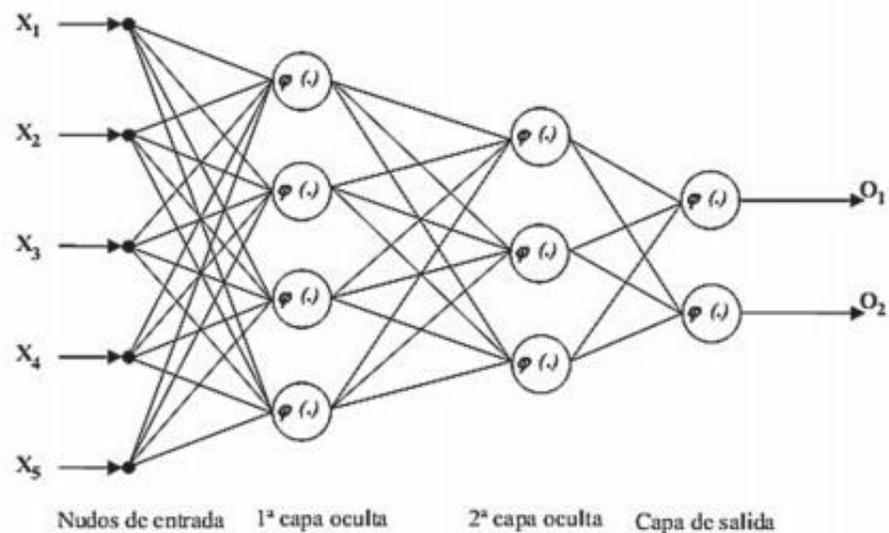


Figura 4. Red Neuronal feedforward de 4 capas, dos de ellas ocultas.

Tomado de: (F. Villada, N. Muñoz & E. García, 2012).

Agrupamiento o Clustering: es una técnica que permite agrupar un conjunto de vectores de acuerdo a su distancia, esta técnica permite ubicar los vectores cuya distancia sea la más cercana y con las mismas características. Esta técnica utiliza métodos estadísticos para realizar las agrupaciones en base a características similares, por ejemplo, segmentar clientes, sucursales, poblaciones para optimizar las campañas de marketing y crear estrategias más personalizadas.

2.1.5.3. Algoritmo de asociación

“Los modelos de asociación se generan basándose en conjuntos de datos que contienen identificadores para casos individuales y para los elementos que contienen los casos. Un grupo de elementos de un caso se denomina un conjunto de elementos. Un modelo de asociación se compone de una serie de conjuntos de elementos y de las reglas que describen cómo estos elementos se agrupan dentro de los casos. Las reglas que el algoritmo identifica pueden utilizarse para predecir las probables compras de un cliente en el futuro, basándose en los elementos existentes en la cesta de compra actual del cliente. El siguiente diagrama muestra una serie de reglas en un conjunto de elementos” (Microsoft SQL server, 2014). (Ver Figura 5)

| Regla |
|---|
| Road Bottle Cage = Existing, Cycling Cap = Existing -> Water Bottle = Existing |
| Mountain-200 = Existing, Mountain Tire Tube = Existing -> HL Mountain Tire = Existing |
| Mountain-200 = Existing, Water Bottle = Existing -> Mountain Bottle Cage = Existing |
| Touring-1000 = Existing, Water Bottle = Existing -> Road Bottle Cage = Existing |
| Road-750 = Existing, Water Bottle = Existing -> Road Bottle Cage = Existing |
| Touring Tire = Existing, Sport-100 = Existing -> Touring Tire Tube = Existing |

Figura 5. Reglas derivadas de un conjunto de elementos.

Tomado de: (Microsoft SQL server, 2014).

2.1.6. Big Data en tiempo real

En la actualidad tecnologías existentes como los, dispositivos móviles, máquinas M2M (Machine to Machine), sistemas de información geográfica, sensores, códigos QR, sistemas de medios de comunicación, etc., generan grandes volúmenes de datos que pueden ser aprovechados por las empresas para distintos fines.

Dada esa gran cantidad de información y su flujo creciente sobre todo en datos de texto, ubicación, video, fotografía, audio, el procesamiento en tiempo real es crucial para encontrar datos significativos de calidad.

El término en “tiempo real” explica que los datos se analizan y procesan en tiempo presente y no en el futuro, o la capacidad de analizar y procesar los datos tan pronto está en el sistema. El calificativo en “tiempo real” se puede definir como rapidez con que un usuario recibe una respuesta de un ordenador sea inmediata, este tipo de análisis también recibe el nombre de “integración de datos en tiempo real” o “inteligencia en tiempo real”.

El Análisis en tiempo real no es solo un proceso para almacenar grandes cantidades de información en un almacén de datos, más bien trata de la combinación de técnicas, análisis para que se pueda tomar decisiones en el momento oportuno y en el lugar correcto. El procesamiento de datos en tiempo real presume una ventaja y es apto para muchos sectores y casos de uso de *Big Data*, “la información derivada del análisis aporta a las empresas visibilidad de numerosos aspectos del negocio y de las actividades de los clientes, como el uso del servicio (para la medición/facturación), la actividad del servidor, los clics en un sitio web y la ubicación geográfica de dispositivos, personas y mercancías, y les permite responder con rapidez ante cualquier situación que surja. (Amazon Web Service, 2017)”.

2.2. Industria de telecomunicaciones

Los operadores de telecomunicaciones móviles están en un momento complejo, muchos expertos lo definen como la “tormenta perfecta”, una economía con poca expansión, los ingresos crecen de a poco, con un incremento excesivo en sus redes, nuevos competidores, transformación de la tecnología, cambios regulatorios y una inevitable aproximación entre las industrias de comunicaciones y entretenimiento (Madrid, 2013).

2.2.1. Evolución de las telecomunicaciones

El sector de las telecomunicaciones ha venido experimentando un gran crecimiento, a partir de la evolución de los dispositivos móviles, con esto ha surgido una explosión por la demanda de datos y aplicaciones móviles que permite a sus usuarios utilizar servicios móviles en cualquier parte del mundo. El GSM (Global System for Mobile communications, Sistema Global para Comunicaciones Móviles) es un estándar para empresas de telefonía móvil, acogida por más de una centena de países, que ha permitido que las economías de escala sean suficiente para que las inversiones en una nueva infraestructura sean bastantes menores de lo que habrían sido en otras circunstancias. Además, la mayor capacidad de tráfico de las redes de telecomunicación permite repartir los costes de la infraestructura entre la mayor cantidad de usuarios lo que ha permitido una baja continua de tarifas. Este desarrollo consensuado ha dotado de nuevas facilidades, a medida que se van desarrollando aplicaciones de usuario específicas que aumentan su valor como la personalización a través de una tarjeta SIM (Subscriber Identification Module, módulo de identificación de abonado), la capacidad de enviar mensajes cortos SMS, y la separación entre el abonado y el usuario o que es lo mismo entre el usuario y el terminal.

Esta combinación entre todos estos factores técnicos y económicos ha permitido transformar las ofertas de los servicios de comunicaciones móviles que la que existía con las redes analógicas, motivando a los usuarios a reevaluar el valor que se les puede aportar a las comunicaciones.

Tal como lo determina Voinea (2011), “Las redes de telefonía móvil han ido evolucionando y lo siguen haciendo a una velocidad vertiginosa” (Pág. 167). Las redes móviles se han categorizado en las siguientes generaciones:

- Primera Generación (1G)

La primera red móvil se construyó en Tokio, Japón fue impulsada por la empresa NTT (Nippon Telegraph and Telephone), posterior a esto en Europa se lanzó dos sistemas analógicos más importantes fueron el NMT (Nordic Mobile Telephones) y el TACS (Total Access Communication System). Esta red móvil ofreció por primera vez handover y roaming, sin embargo, estas redes móviles fueron incapaces de conectarse con otras redes en otros países y esto significó una de las principales desventajas de esta primera generación de redes móviles.

En 1982, en Estados Unidos, el sistema AMPS (*Advanced Mobile Phone System*) operaba en el ancho de banda de 40 MHz en el rango entre 800 y 900 MHz por la FCC (*Federal Communications Commission*) para AMPS.

Posterior a esto aparece un ancho de banda expandido asignado a AMPS en 1998 de 10 MHz, que permitió que las transmisiones de las estaciones bases a móviles usen frecuencias entre 869 y 894 MHz. El canal es usado para transmisiones de móviles a estación base, usando frecuencias entre 824 y 849 MHz. AMPS y TACS usan la técnica de modulación de frecuencia (FM) para radio transmisión. El tráfico es multiplexado sobre un sistema FDMA.

- Segunda generación (2G)

Las redes móviles de segunda generación se introdujeron al final de los 1980s, usan tecnología digital de múltiple acceso, como son TDMA y CDMA. Como consecuencia, y comparados con los sistemas de primera generación, mayor eficiencia del espectro, mejores servicios de datos y roaming más avanzado han sido ofrecidos por los sistemas 2G.

En la segunda generación de redes móviles hubo tres líneas de desarrollo en los sistemas celulares digitales: El primer sistema digital, introducido en 1991, fue el IS-54 (*North American TDMA Digital Celular*), del cual una nueva versión que soporta servicios adicionales fue introducida en 1996. Mientras tanto, la primera versión de CDMA fue desplegado en 1993. La comunicación 2G es generalmente asociada con los servicios GSM.

- Tercera generación (3G)

La tercera generación de redes móviles se caracteriza por un incremento de la claridad en la señal de transmisión. Estas redes de telecomunicación 3G soportan servicios que proveen un radio de transferencia de información de al menos 2 Mbps. La gran creciente de datos ha permitido que el paquete sobre el espectro radioeléctrico se comporte como un circuito de interruptores de llamada. Por lo tanto, parte de esta eficiencia de conexión de paquetes está perdida en el entorno de circuitos interruptores.

A esta tercera generación de redes móviles es un conjunto de estándares que habilitan a los operadores de red a ofrecer mejores y un rango más amplio servicios con una mayor capacidad de la red mediante eficiencia espectral mejorada. La primera red comercial 3G fue lanzada en Japón denominado FOMA, esta red de tercera generación brinda mayor velocidad en el acceso de datos multimedia, la conectividad permanente inalámbrica y una velocidad hasta siete veces más rápida que la conexión telefónica estándar.

- Cuarta generación (4G)

La cuarta generación 4G o también llamada LTE fue probada por primera vez en Tokio, y fue un éxito el alcanzar la transmisión de paquetes en tiempo real en 1 Gbps en la descarga a una velocidad de movimiento de cerca 20 km/h. Para usar

servicios 4G, terminales de usuario deben tener una tecnología adicional para poder soportar esta red.

Esta nueva generación permite a los clientes importantes asuntos en la movilidad del terminal, localización, el sistema rastrea y localiza un terminal móvil para una posible conexión.

La nueva versión IPv6 móviles es un nuevo protocolo que se soporta en IP estándar para sistemas de redes de telecomunicaciones, cada terminal tiene una dirección IPv6. Cuando quiera, el terminal se mueve fuera de la red local, la dirección de domicilio se vuelve inválida, y el terminal obtiene una nueva dirección IPv 6 ofreciendo mayor flexibilidad del uso del espectro.

Con el diseño y la optimización de próximas técnicas de acceso de radio y una nueva evolución del sistema existente, la 3GPP había establecido las bases de la futura LTE (Long Term Evolution) estándares avanzados el candidato 3GPP para 4G.

Este nuevo acceso al espectro de esta nueva tecnología se fijó a 30 bps/Hz y 15 bps/Hz en enlace de subida y enlace de bajada de transmisión, respectivamente. Aparte de las técnicas de multisesión para el acceso, el nuevo método y técnicas de transmisión de canales y la amplia coordinación entre múltiples sitios, a las cuales se denomina multipunto coordinado de transmisión/recepción clave para LTE.

- Quinta generación (5G)

La quinta generación de redes móviles 5G es la red de comunicación más veloz hasta la fecha, es una red de comunicación inalámbrica, que se compone de estándares de telecomunicaciones móviles más allá de los estándares del 4G.

La siguiente generación de redes móviles 5G no es un término oficial y se espera que salga al mercado en el 2020, pero se hace público por compañías de telecomunicaciones u organismos de estandarización como son 3GPP, WiMaxForum, o la UTI-R. Más allá de la alta velocidad se espera que esta nueva generación tenga nuevos casos de uso, mejorar aún más el rendimiento del sistema y añadir nuevas capacidades con nuevas áreas de aplicación para el internet de las cosas, servicios en la nube, entre otros. La mejora de rendimiento se verá en baja latencia, disminución de costos, mayor capacidad de transmisión y menos consumo de energía, se estima que alcanzará entre 5 y 10 gigabytes, 10 veces más que la anterior generación 4G.

2.2.2. Modelo actual de negocio

El modelo de negocio actual según el Consejero Delegado del grupo Telefónica, Julio Linares “este modelo no es sostenible y que el sector tiene que centrarse en equilibrar el crecimiento del tráfico y la inversión, con un nuevo modelo de Internet y la creación de valor”. El directivo explicó que si se mantiene el actual modelo en el que el tráfico crece y disminuyen los ingresos quebrará el sistema, pero precisó que eso no le interesa a ningún agente del sector y se va a actuar, pero que hay que darse cuenta de que hay que hacerlo cuanto antes. Linares también recalcó la necesidad de hallar una respuesta entre todos al "imparable" aumento del tráfico debido a Internet, introduciendo medidas que optimicen la capacidad de la red, reduzcan los costos por megabit, descarga del tráfico a través de redes como Wi-Fi y el uso de las llamadas "Femtocélulas", y precisó que no sirven respuestas individuales.

EL modelo de negocio actual de los operadores de telecomunicaciones móviles se basa en servicios de voz, SMS e internet, este modelo tradicional basa sus ingresos principalmente en los servicios contratados por sus clientes (Consumidores, Pymes, Grandes Empresas, etc.) y el pago a proveedores para infraestructura, interconexión, terminales y espectro radioeléctrico.

2.2.3. Cadena de valor

La cadena de valor ayuda a determinar las actividades que permiten que una organización desarrolle una ventaja competitiva sobre otras que están en el mismo mercado. El modelo de la cadena de valor resalta las actividades específicas del negocio, en las que se pueden aplicar mejores estrategias competitivas y diferenciadoras, para lograr un mayor impacto dentro del sector en el que está inmersa la organización. El uso de esta herramienta y/o modelo al interior de una organización considera la comparación de sus procesos de negocios con los de sus competidores, o con otras organizaciones relacionadas, y ayuda a identificar las mejores prácticas de la industria; esto con el fin de poder determinar los factores que generan un valor diferenciador en los productos/servicios.

Kearney (2010) identificó los diferentes elementos de la industria de las telecomunicaciones e identificó las fuentes de innovación y desarrollo que permiten generar valor. Para el 2015 se ha actualizado para reflejar todos los actores involucrados en el servicio extremo a extremo como se muestra en la Figura 6 y los elementos son:

- *Content Rights* (Derechos de contenido): Derechos a varios contenidos que se comercializan a través de internet.
- *Online Services* (Servicios en línea): Existe una amplia gama de servicios en línea que se ofrecen a través de internet a través de navegadores, plataformas y aplicaciones. Abarca mucho de lo que la gente percibe como "Internet".
- *Enabling Technology Services* (Servicios tecnológicos habilitadores): Servicios que no son necesariamente visibles para los usuarios de internet, pero necesarios para el funcionamiento de la infraestructura global.
- *Connectivity* (Conectividad): Representa el medio por el cual los usuarios se conectan a internet.

- *User Interfaces* (Interfases de Usuario): Aplicaciones, dispositivos, sistemas, software que utiliza el usuario para acceder a internet.
- *Users* (Usuario): Los Usuarios y/o clientes.

En esta cadena de valor se combina B2B (Business to Business, Negocio entre empresas) y B2C (Business to Consumer, negocio al consumidor), por ejemplo: redes sociales como LinkedIn, servicios de comercio electrónico como Amazon y plataformas de comunicación como Skype sirven tanto en el mercado de B2B como B2C.

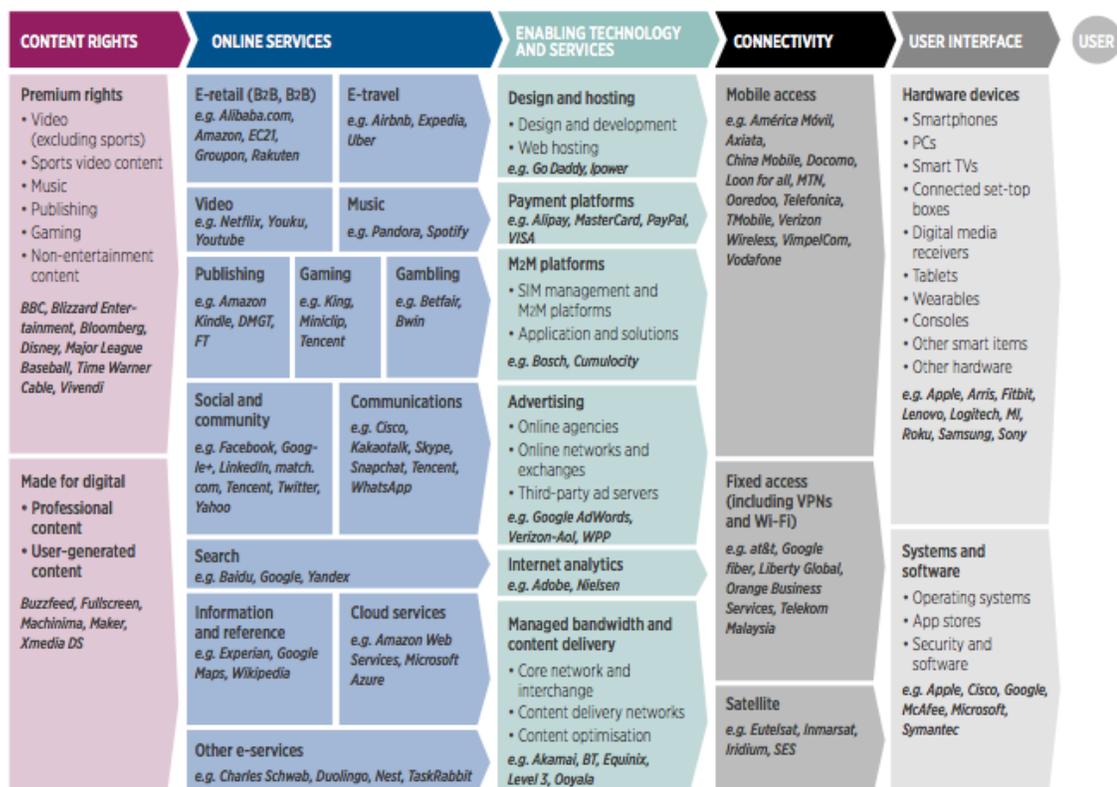


Figura 6. Cadena de valor de internet.

Tomado de: (A.T. Kearney Analysis, 2017).

El segmento de conectividad representa el medio por el cual los usuarios finales se conectan para hacer uso de los servicios voz, servicios de valor agregado como SMS, conectividad a internet y servicios de contenido. Es la responsabilidad

principal de los operadores de telecomunicaciones y representa la base sobre la cual se está construyendo la revolución digital., en 2015, el 43% de la población estaba conectada a internet y 1100 millones de habitantes lo hacía a alta velocidad, al 2017 el 48% de la población estuvieron conectados a internet según una estimación del ITU (About International Telecommunication Union, Unión Internacional de las telecomunicaciones). La propagación y acceso a internet están transformando por completo el ecosistema.

Con el modelo actual los operadores de telecomunicaciones móviles se enfrentan ante un gran desafío en revertir la situación de los próximos años proteger su rentabilidad y hacer crecer sus ingresos con una presión competitiva y regulatoria cada vez mayor que obliga un cambio de estrategia y mejorar la oferta al cliente.

Los operadores de telecomunicaciones móviles han fracasado en gran medida para lanzar y escalar nuevos modelos de negocio. Los Proveedores de servicios de comunicaciones (CSPs) pueden haber avanzado mucho en la tecnología, los estándares y la interoperabilidad, pero han perdido tanto con nuevas y más ágiles empresas digitales como WhatsApp, Skype, Facebook y los operadores de tecnología en activo que están adquiriendo agresivamente capacidades de comunicación digital. La participación de los operadores en el sector de la industria se ha reducido del 58% en 2010 al 47% en 2015 y se prevé que caiga aún más, hasta el 45% en 2018 según el (*World Economic Forum & Accenture, 2017*).

2.2.4. Problemas de sustentabilidad del modelo actual

A medida que el mercado digital crece un adicional de mil millones de clientes en promedio para los operadores de telecomunicaciones móviles y sobre todo en mercados emergentes para el 2025 según la consultora (*McKinsey & Company, 2017*), la puerta para los nuevos proveedores over-the-top (OTT) se está abriendo. Estos nuevos proveedores digitales están ofreciendo los mismos servicios que los operadores de telecomunicaciones móviles, servicios básicos de llamada,

mensajería y video que suelen ser el dominio de los operadores de telecomunicaciones móviles convencionales. En particular los proveedores OTT como FaceTime de Apple, Google, Skype, WhatsApp, WeChat están amenazando ya con canibalizar las principales ofertas de los operadores de telecomunicaciones móviles, incluso con servicios más atractivos de mensajería y comunicación.

Según (*McKinsey & Company*, 2017) en su artículo *Overwhelming OTT* menciona que hace pocos años la mensajería, voz fija y servicios de voz móviles de proveedores OTT representaron el 9, 11, 2 por ciento respectivamente de sus ingresos con la posibilidad de aumento para cada uno de estos servicios. En el escenario más agresivo la proporción de mensajería, voz fija y voz móvil proporcionado por los proveedores OTT podrían estar en un 60, 50 y 25 por ciento, respectivamente para el 2018 en un entorno totalmente IP.

Poco a poco estas ofertas se están constituyendo en modelos de negocios innovadores que estarán a un precio más bajo para los consumidores que los que puedan ofrecer los operadores de telecomunicaciones móviles tradicionales.

De acuerdo con Ovum, esto probablemente resultará en una caída en el gasto en servicios de comunicación tradicionales en un 36 por ciento en los próximos diez años, empujando aún más las compañías predominantes a los márgenes de la prestación de voz y datos (Ver Figura 7).

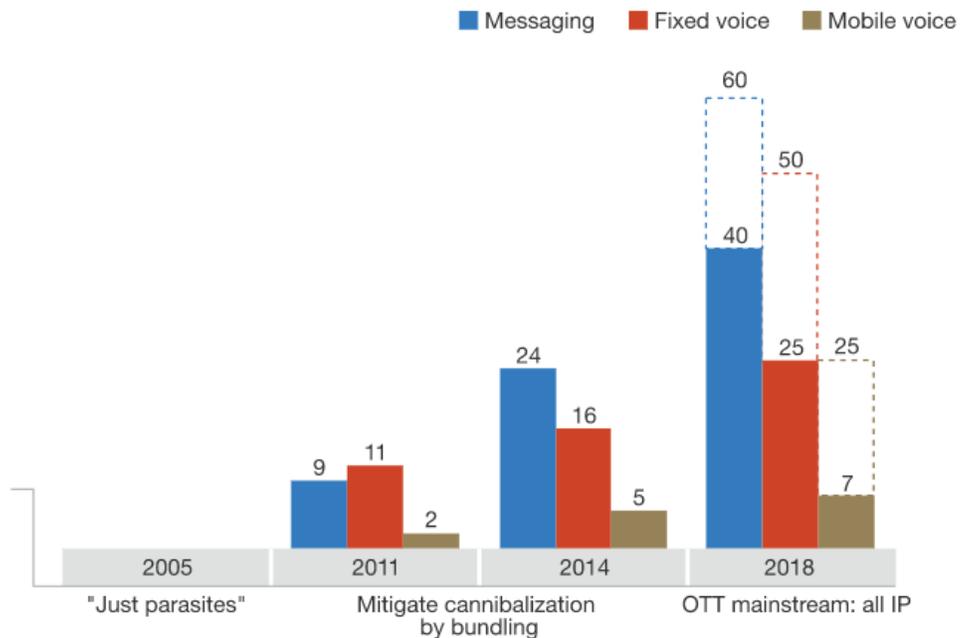


Figura 7. Canibalización de ingresos OTT.

Tomado de: (Mckinsey analysis Ovum, s.f.).

Según (Accenture, 2017) menciona que los operadores de telecomunicaciones móviles no han tenido éxito en la protección de sus principales fuentes de ingresos y las estimaciones han demostrado que los proveedores OTT generan 50 a 90% menos de ingreso para los CSPs. Si bien el aumento exponencial del consumo de datos ha proporcionado cierto alivio, no ha sido suficiente para superar la disminución constante del ingreso promedio por usuario (ARPU) del servicio de voz móvil.

Los operadores de telecomunicaciones móviles han fracasado en gran medida para lanzar y escalar nuevos modelos de negocio, los CSPs han avanzado mucho en tecnología de interoperabilidad para la conectividad inalámbrica, pero han perdido tanto en relación a empresas digitales como Whastapp que han ido adquiriendo nuevas capacidades de comunicación digital. Las presiones de colisión de los ingresos y costos de los operadores de telecomunicaciones móviles acompañado a la presión de monetizar estas oportunidades digitales han dado lugar a que la creación de valor se vea afectada. La participación de los operadores de

telecomunicaciones móviles según (Accenture, 2017) en ganancias por industria ha disminuido de 58% en 2010 a 47% en 2015 y se prevé que caiga a 45% en el 2018 (Ver Figura 8).

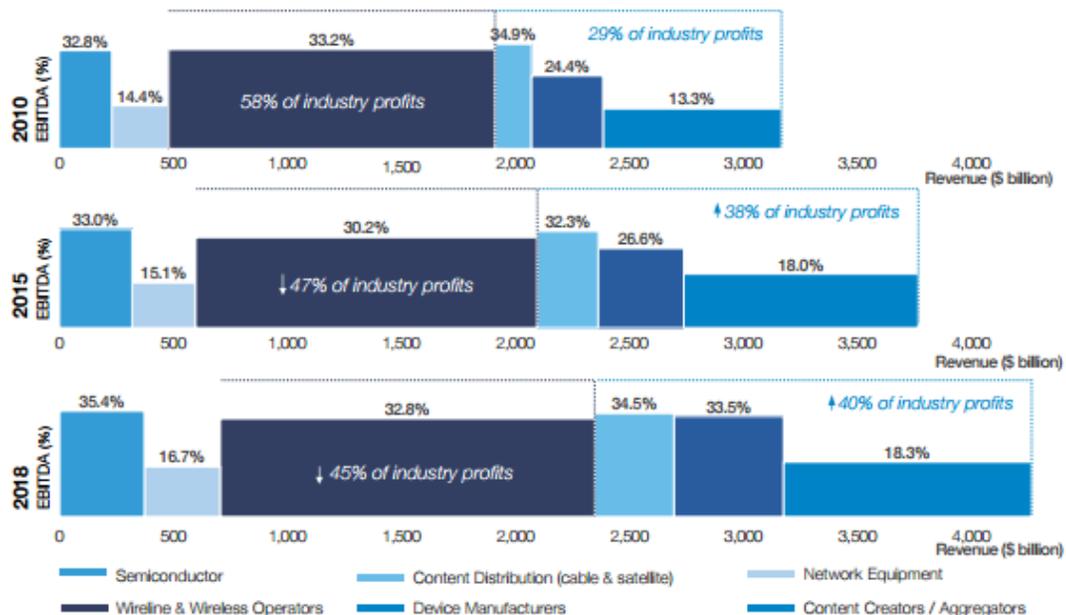


Figura 8. Disminución de ganancia total de la industria de telecomunicaciones.

Tomado de: (World Economic Forum/Accenture analysis, 2017).

Sin embargo, no todos los segmentos del ecosistema de telecomunicaciones han luchado para crear valor. Las empresas de creación, distribución y agregación de contenidos digitales han experimentado grandes aumentos en los beneficios y en el valor para los accionistas durante los últimos 10 años. Junto con los fabricantes de dispositivos, la participación combinada de los beneficios de la industria de estos segmentos se espera que aumente a 40% en 2018, frente al 29% en 2010. Los creadores de contenido digital y agregadores, incluyendo Google, Netflix y Facebook, han alcanzado escala y valor capturado aprovechando efectivamente la penetración mundial de Internet para ofrecer servicios y experiencias digitales únicos a los consumidores finales.

2.3. Marco de referencia

En general, un marco es una estructura real o conceptual destinado a servir como un soporte o guía que permitirá la construcción de un proyecto útil que complemente en el ámbito académico.

Una primera definición de Framework, extraída de Erich Gamma es: "A framework is a set of classes that embodies an abstract design for solutions to family related problems (Un marco de referencia es un conjunto de clases que incorpora un diseño abstracto para problemas familiares)" (JohnsonandFoote, 1988).

Zachman framework, una estructura lógica destinada a proporcionar una representación integral de una empresa de tecnología de la información que es independiente de las herramientas y métodos utilizados en cualquier negocio de TI en particular.

En conclusión, un marco de referencia proporciona una visión amplia, un esquema de elementos relacionados que apoya un enfoque particular de un determinado objetivo, sirve como guía y puede ser modificado de acuerdo a las necesidades adicionando o eliminando elementos.

3. BIG DATA Y BIG DATA ANALYTICS

El análisis de *Big Data* reside en el análisis de grandes volúmenes de información de una amplia variedad de tipos y sobre todo de gran valor, para descubrir encontrar información oculta, patrones recurrentes, nuevas correlaciones y otra información útil de modo que el resultado del análisis permita a las empresas aprovechar sus datos y utilizarlos para identificar nuevas oportunidades. Eso, a su vez, conduce a movimientos de negocios más inteligentes, operaciones más eficientes, mayores ganancias y clientes más felices.

Las grandes tecnologías de datos, como *Hadoop* y el análisis basado en la nube, aportan importantes ventajas en términos de costes cuando se trata de almacenar grandes cantidades de datos, además de identificar maneras más eficientes de hacer negocios. Con la velocidad de *Hadoop* y la analítica en memoria, combinada con la capacidad de analizar nuevas fuentes de datos, las empresas pueden analizar la información inmediatamente y tomar decisiones basadas en lo que han aprendido.

Con la capacidad de medir las necesidades de clientes y la satisfacción a través de análisis viene el poder de dar a los clientes lo que quieren. Con la analítica de *Big Data*, más empresas están creando nuevos productos para satisfacer las necesidades de clientes.

3.1. *Big Data* & Operadores de telecomunicaciones móviles

Gracias al *Big Data* se puede exprimir al máximo la información que se almacena y plantear preguntas complejas o saber mucho más sobre los usuarios o las operaciones de una empresa. Los datos que generan los operadores de telecomunicaciones móviles se cuentan como activos económicos con un inmenso valor y que generan interés para las empresas, poder explotar estos datos y maximizar el valor de estos activos. Para los operadores de telecomunicaciones móviles, *Big Data* supone nuevas inversiones, pues aumentará el tráfico de las redes y la demanda de sistemas en la nube. También son necesarios nuevos conocimientos, plataformas hardware, herramientas software y procesos operativos y comerciales. Sin embargo, *Big Data* es también una poderosa tecnología que pueden explotar para ganar cuota de mercado y mejorar la percepción del cliente, crecer en ingresos y rentabilidad, y optimizar operaciones. Así, *Big Data* supone grandes retos para las operadoras, pero también una gran oportunidad.

Los operadores pueden obtener información muy importante sobre sus clientes, desde localización a incluso intereses personales, pero aún por diversos motivos, no han podido extraer todo el valor estratégico de dichos datos. Los datos son tanto

estructurados (perfil del cliente, peticiones de servicios, tarificación, incidencias técnicas generadas, etc.), no estructurados (documentos, vídeos, imágenes, contenido Web, localización, presencia, DPI, señalización SIP/Diameter/SS7, logs, grabaciones del “contact center”, etc.) y parcialmente estructurados (perfil del cliente enriquecido con CDRs o “call data records” e información externa como blogs, foros, redes sociales, etc.). Mediante DPI (Deep Packet Inspection), pueden saber cuánto ancho de banda utiliza el usuario, cuándo se conecta, qué sitios Web visita, qué aplicaciones utiliza, etc. Incluso podrían obtener información adicional en tiempo real sobre los gustos e intereses del cliente, si bien actualmente se encuentran limitados por cuestiones legislativas sobre privacidad y confidencialidad. Toda esta información que se obtiene de distintas fuentes tiene que ser organizada y luego analizada para dar soporte a la toma de decisiones.

Esta tecnología permite a las operadoras, por ejemplo, predecir y reducir la rotación de abandono de clientes, además de impulsar la fidelidad de éstos mediante ofertas especiales que combinen diversos productos y proporcionen servicios personalizados. No obstante, se trata de una tecnología que aún debe implantarse en mayor medida, si las operadoras quieren rentabilizar económicamente los datos de sus clientes. De hecho, según señala Ovum, uno de los principales retos de las operadoras es renovar su tradicional infraestructura de tecnologías de la información, y ser más flexibles, imitando el modelo de sus competidores OTTs (Dintel, 2012), que pueden ofrecer nuevos servicios sin grandes inversiones y en muy poco tiempo.

Un claro ejemplo es la empresa española Telefónica que ha sido una de las operadoras con mayor futuro, la cual a través de su nuevo plan estratégico denota que la aproximación a *Big Data* cada vez es mayor. Telefónica Digital en octubre del 2012, creó una nueva unidad de negocio global, *Telefónica Dynamic Insights*, para el desarrollo de ofertas comerciales para empresas privadas y organismos públicos, basadas en el propio *Big Data* del Grupo Telefónica. El primer producto en ser lanzado al mercado, en concreto en Reino Unido, es “*Smart Steps*”, que utiliza

datos agregados y totalmente anónimos de la red móvil. Estos datos permitirán a las empresas privadas y organismos públicos medir, comparar y comprender qué factores influyen en el número de personas que visitan un determinado lugar en cualquier momento.

De este modo, los establecimientos comerciales pueden realizar ofertas puntuales adaptadas a cada una de sus tiendas y determinar las mejores localizaciones y los formatos más apropiados para la apertura de nuevas tiendas. “*Smart Steps*” podrá ayudar asimismo a los ayuntamientos a valorar los resultados en materia de afluencia de público de diferentes iniciativas como, por ejemplo, la respuesta de los ciudadanos a la disponibilidad de zonas de parking gratuito en diferentes zonas de la ciudad o la afluencia de personas a ferias y mercadillos. *Telefónica Dynamic Insights* está desarrollando también productos analíticos destinados a empresas de sectores muy variados, como la protección contra el fraude y la tecnología “*smart city*”, incluida la gestión de tráfico.

Big Data tiene la capacidad obtener y procesar información procedente de diversas fuentes, obtener patrones de comportamiento para diseñar soluciones y hacer más eficientes las ciudades, es así que varios autores definen a una “*smart city*” como un sistema interconectado que aplica tecnologías de la información para mejorar la calidad de vida de los ciudadanos gestionando de manera eficiente la movilidad urbana, residuos urbanos, seguridad ciudadana, gestión de agua, energía y eficiencia energética.

“Una ciudad inteligente detecta las necesidades de sus ciudadanos, y reacciona a estas demandas transformando las interacciones de los ciudadanos con los sistemas y elementos de servicio público en conocimiento. Así, la ciudad basa sus acciones y su gestión en dicho conocimiento, idealmente en tiempo real, o incluso anticipándose a lo que pueda acaecer” (BBVA,2017), explica Juan Murillo, responsable de Análisis Territoriales de BBVA Data & Analytics.

3.1.1. Oportunidades de Big Data

Big Data abarca el procesamiento, el tratamiento de grandes volúmenes de datos, supera las capacidades de las herramientas de bases de datos actuales. Estos datos son procesados con diferentes técnicas a fin de convertir esta en información que permita descubrir qué oferta es la más adecuada para el cliente. Por lo tanto, un número creciente de organizaciones les hacen frente desplegando herramientas especializadas como bases de datos de procesamiento masivamente paralelo, sistemas de archivos distribuidos Hadoop, algoritmos MapReduce, computación en la nube. La integración de datos permite a la organización la combinación de *Big Data* con los datos transaccionales tradicionales para generar valor y conseguir la mayor eficacia posible. Por esta razón uno de los aspectos más interesantes no es tanto lo que harán ellos mismos por el negocio, sino qué se podrá conseguir para el negocio cuando se combinan con otros datos de la organización.

El mayor valor de los *Big Data* puede producirse cuando se los combinan con otros datos corporativos. Colocándolos en un contexto más grande para conseguir la mayor cantidad de conocimiento sobre el negocio y se incremente exponencialmente con el tiempo.

Frank (2012) considera que es muy importante que la organización no desarrolle una estrategia de *Big Data* distinta a su estrategia tradicional de datos ya que en ese caso fallará toda la estrategia del negocio. *Big Data* y datos tradicionales son ambas partes de la estrategia global. Para que las organizaciones tengan éxito se necesita desarrollar una estrategia cohesiva donde los *Big Data* no sean un concepto distinto y autónomo. De igual forma insiste en: “la necesidad desde el comienzo de pensar en un plan que no solo captura y analice los grandes datos por sí mismo, sino que también considera como utilizarlos en combinación con otros datos corporativos y como un componente de un enfoque holístico a los datos corporativos” (p. 22).

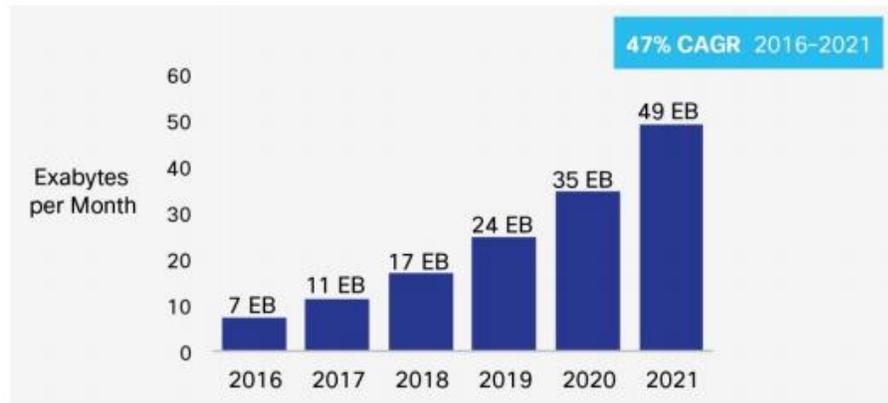
Todas las empresas que han tomado la decisión de aprovechar sus datos con *Big Data*, conseguirán una ventaja competitiva, ya que crear valor a través de sus datos les permitirá mejorar el conocimiento de sus productos, servicios, clientes, fortaleciendo sus oportunidades de crecimiento y expansión.

Un caso de éxito es *T-Mobile* que consiguió disminuir al 50% el número de portabilidades a otras operadoras gracias a la adopción de *Big Data*. *T-Mobile* analizó 3 puntos importantes como: sistemas de facturación, herramientas de monitorización social y herramientas de inteligencia de negocios, descubrieron que la portabilidad a otras operadoras en Estados Unidos podía determinarse por tres factores: facturas, llamadas no terminadas y conversaciones de clientes (positivas, negativas o neutras). Con todos estos factores *T-Mobile* logró calcular el valor que tiene el servicio para cada cliente en tiempo real y ofrecer servicios más personalizados. *T-Mobile* menciona en un informe que de 100.000 portabilidades en los primeros tres meses del 2011 lograron reducir a tan sólo 50.000 en el segundo trimestre, un descenso del 50% gracias la ayuda del *Big Data* en tiempo real.

Con todo esto *Big Data* se convierte no solo en una capacidad para analizar y procesar grandes volúmenes de datos si no en una capacidad para crear nuevos productos, conocer mejor el comportamiento de sus clientes, incrementar las ventas, maximizar sus ingresos, tener una visión 360° del cliente, extender su negocio, conseguir resultados basados en el cliente.

3.1.2. Volumen

Se espera que el tráfico total de datos móviles crezca a 49 exabytes al mes para el 2021, un aumento siete veces superior a 2016. El tráfico de datos móviles crecerá a una tasa de crecimiento interanual del 47 por ciento entre 2016 y 2021 (Ver Figura 9).

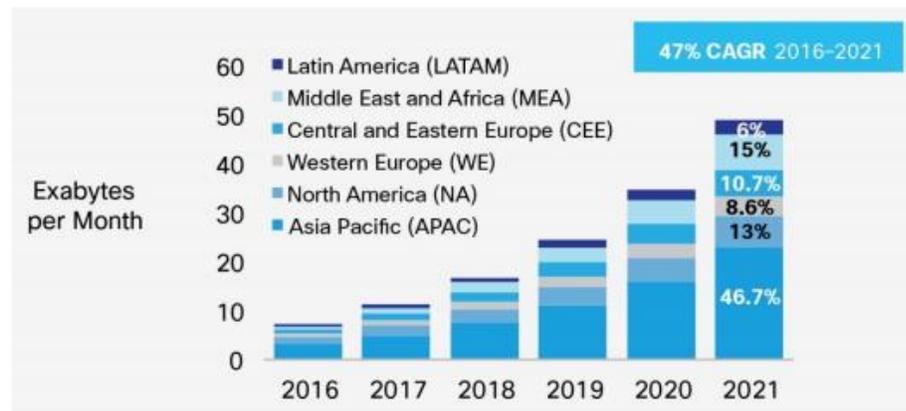


Source: Cisco VNI Mobile, 2017

Figura 9. Exabytes per Month of Mobile Data Traffic by 2021.

Tomado de: (Cisco VNI Mobile, 2017).

El crecimiento de datos en Asia representará el 47 por ciento del tráfico móvil mundial en 2021, la mayor proporción de tráfico de cualquier región, como se muestra en la Figura 9. Norteamérica, que tuvo la segunda mayor participación de tráfico en 2016, sólo tendrá la cuarta mayor participación en 2021, superada por Europa Central y Oriental y Oriente Medio y África. Oriente Medio y África experimentarán el CAGR más alto del 65 por ciento, aumentando 12 veces durante el período de pronóstico. Asia Pacífico tendrá el segundo mayor CAGR de 49 por ciento, aumentando 7 veces durante el período de pronóstico (Ver Figura 10).



Source: Cisco VNI Mobile, 2017

Figura 10. Global Mobile Data Traffic Forecast by Region.

Tomado de: (Cisco VNI Mobile, 2017).

Los crecientes volúmenes de datos móviles han alcanzado niveles nunca antes vistos, y más aun con el apareamiento de redes LTE / 4G, el volumen de datos operativos generados con cada llamada o sesión se está multiplicando por diez. El uso ampliado de GPS, servicios basados en la localización, y medios sociales está añadiendo al torrente de datos. La inclusión de IPv6 creará tantas direcciones IP como hay granos de arena en la Tierra, permitiendo que el número de dispositivos conectados a Internet crezca de forma exponencial. Este volumen de datos requiere nuevas capacidades operacionales en tiempo real para funciones tales como la carga en tiempo real y el marketing basado en eventos, lo que a su vez exige un mayor almacenamiento de datos para el cumplimiento y posibles usos futuros, así como nuevas herramientas para mediar, archivar datos dentro de marcos de tiempo disponibles.

Según el informe *Cisco VNI Mobile* (2017) en 2016, 4G ya transportó el 69 por ciento del tráfico móvil total y representó la mayor proporción de tráfico de datos móviles por tipo de red. Seguirá creciendo más rápido que otras redes para representar el 79 por ciento de todo el tráfico de datos móviles en 2021 (Ver Figura 11). En 2021, 5G soportará el 1,5 por ciento del tráfico móvil. Se espera que la conectividad 5G

con su ancho de banda muy alto (100 Mbps) y latencia ultra baja (1 milisegundo) conduzca volúmenes de tráfico muy altos.

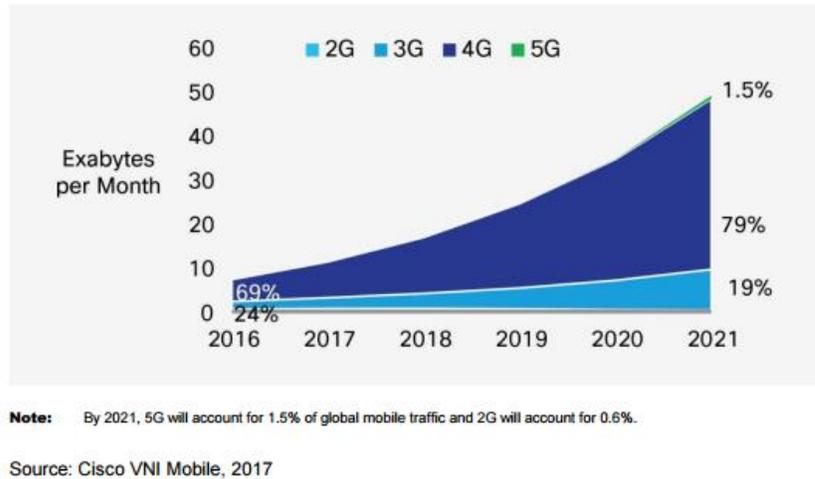


Figura 11. Global Mobile Traffic by Connection Type.

Tomado de: (Cisco VNI Mobile, 2017).

3.1.3. Variedad

Los medios sociales, los dispositivos móviles y los sensores que supervisan todo, desde el uso de servicios públicos hasta el cumplimiento médico, inundan las infraestructuras de telecomunicaciones con datos en innumerables formatos. Los operadores de telecomunicaciones móviles deben enriquecer sus datos de CDR con servicios basados en la localización, información financiera y otros datos no estructurados, luego estandarizarlos para plataformas de inteligencia de negocios antes de que puedan analizarlos para tener una mayor perspectiva de suscriptores y nuevas oportunidades de negocio.

Los operadores de telecomunicaciones móviles cuentan con información muy importante sobre sus consumidores, desde localización hasta sus intereses personales, los datos son tanto estructurados (perfil del cliente, peticiones de servicios, tarificación, incidencias técnicas generadas, etc.), no estructurados (documentos, vídeos, imágenes, contenido Web, localización, presencia, DPI,

señalización *SIP/Diameter/SS7*, *logs*, grabaciones del “*contact center*”, etc.) y parcialmente estructurados (perfil del cliente enriquecido con *CDRs* o “*call data records*” e información externa como blogs, foros, redes sociales, etc.).

Mediante *DPI (Deep Packet Inspection)*, pueden saber cuánto ancho de banda utiliza el usuario, cuándo se conecta, qué sitios Web visita, qué aplicaciones utiliza, etc. Incluso podrían obtener información adicional en tiempo real sobre los gustos e intereses del cliente, si bien actualmente se encuentran limitados por cuestiones legislativas sobre privacidad y confidencialidad.

3.1.4. Velocidad

Para los operadores de telecomunicaciones móviles cuanto más rápido es el análisis de los datos y se acerca al tiempo real más valor tienen los datos. Por ejemplo, no tiene el mismo valor saber que un cliente tuvo una mala experiencia de usuario (Ver Anexo 2) antes o después que dejó la compañía, ya que en la primera opción se podría haber tomado acciones de forma proactiva para retener y fidelizar al cliente impactando directamente en el *churn* de la compañía.

3.1.5. Nuevas estrategias de los operadores de telecomunicaciones móviles

La industria de las telecomunicaciones desde hace varios años ha venido evolucionando y hasta hace poco los operadores de telecomunicaciones móviles basaban sus ingresos en modelos de negocio basados en llamadas de voz, mensajería e internet, mientras que ahora existe la necesidad de generar nuevas fuentes de ingresos. Amin y Feizi (2014) realizan un análisis sobre la incorporación de una estrategia de la explotación de datos en los operadores de telecomunicaciones móviles identificando oportunidades potenciales para mejorar la calidad de los servicios y ayudar a la empresa a entender mejor el comportamiento de sus clientes, adaptar la infraestructura en función del negocio en forma continua, eficiente y reducir riesgos.

"El *Big Data* es esencial en nuestra transformación hacia una empresa digital. Uno de los activos más valiosos que tenemos son nuestros datos. Antes no podíamos explotarlos suficientemente, pero con el *Big Data* se abre un mundo de posibilidades. Nuestro objetivo es convertirnos en una empresa enfocada en los datos, en la que todas nuestras decisiones estén basadas en esos datos", explica Elena Gil, directora *Global de Big Data* B2B y CEO de LUCA (Unidad de servicios *Big Data* para clientes corporativos) de Telefónica dice: "El *Big Data* ha pasado de ser una palabra de moda a una realidad por la que hay que apostar para tener éxito, El *Big Data* vertebró la estrategia de la operadora en su camino hacia una Telco digital. Además del uso transformador interno, la compañía comercializa servicios en esta área".

3.2. Repositorios de datos

Al interior de una empresa son algunos los factores que influyen en el buen funcionamiento de los procesos, estrategias, productos, servicios, etc. Conforme el mercado se globaliza los retos del business intelligence son cada vez mayores, y es necesaria la implementación de nuevas herramientas que permitan gestionar y monitorizar toda la información.

Big Data trae algunos problemas al business intelligence tradicional ya que sus soluciones no son suficientes para sacar provecho a la gran cantidad de datos que ha producido la explosión digital, que para muchos es una evolución del *Big Data*. *Business Intelligence Tradicional* se basa en el principio de agrupar todos los datos empresariales en un servidor central, los datos se estructuran en una base de datos relacional convencional con un conjunto adicional de índices y formas de acceso a las tablas.

En este contexto las soluciones de *Big Data* traen algunas complicaciones para el business intelligence tradicional:

- Costos de almacenamiento para grandes volúmenes de datos.
- Manejo de datos no estructurados
- Análisis en tiempo real
- Procesamiento en paralelo de grandes volúmenes de datos.

En un entorno Big Data se requiere de nuevas capacidades y un nuevo enfoque con multiples repositorios, un BDL (*Business Data Lake*, Lago de datos de negocio), es un repositorio de datos que contiene una gran cantidad de datos estructurados, semi estructurados y no estructurados sin procesar que utiliza una arquitectura plana para almacenar datos sin ser tabulada o clasificada.

3.2.1. Business Intelligence Tradicional

El concepto de BI (*Business Intelligence*, Inteligencia de negocio), es facilitar la toma de decisiones mediante el uso de los datos que se generan en una organización, con el objetivo de respaldar las decisiones con la construcción de conocimiento.

“BI es un proceso interactivo para explotar y analizar información estructurada sobre un área (normalmente almacenada en *datawarehouse*), para descubrir tendencias o patrones, a partir de los cuales se deriva ideas y extraer conclusiones. El proceso de BI incluye la comunicación de los descubrimientos y efectuar cambios. Las áreas Incluyen clientes, proveedores, productos, servicios y competidores.” (*Business Intelligence*, Competir con la información, 2012). (Ver Figura 12)

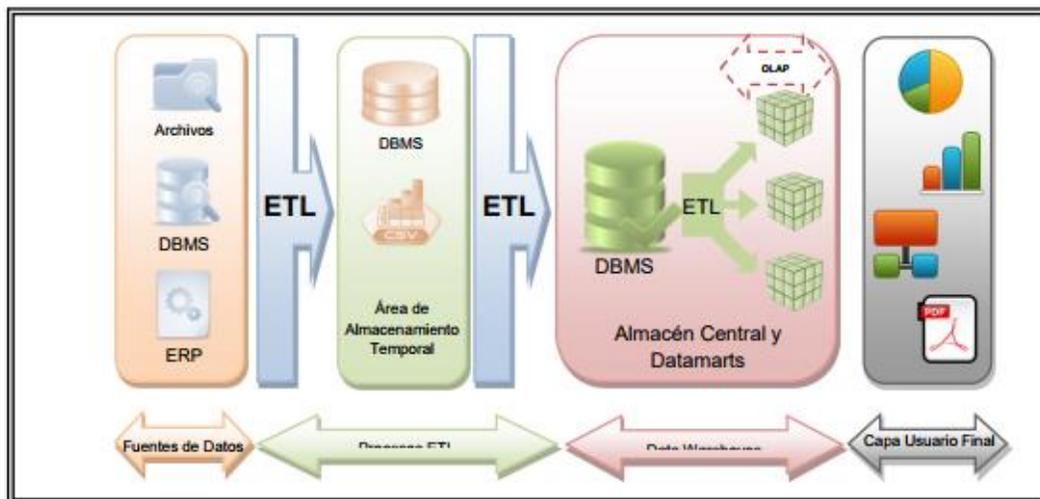


Figura 12. Componentes de una solución BI.

Tomado de: (Roland Bouman, 2009).

- ETL: Los datos de origen para las aplicaciones BI provienen de varias plataformas, que son gestionadas por una variedad de sistemas operacionales y aplicaciones. El propósito de los procesos ETL (Extracción, Transformación y Carga) es unir los datos de estas plataformas heterogéneas y transformar a un formato estándar para llevarlos a otras Bases de Datos, Datamart o DataWarehouse. (Rodríguez, 2009).
- *Data Warehouse* (DWH): Un DWH es un almacén de datos en donde se unifican los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso. (Powerdata, 2016). Un DWH está preparado para concentrar e integrar todos los datos estructurados de una organización independientemente de cuál sea su uso, normalmente se utiliza para guardar información histórica que cubre un amplio periodo de tiempo y el proceso encargado de alimentar el DWH es el ETL.

- *Data Mart*: Un *Data Mart* es una porción de un almacén de datos pensado para cubrir los requerimientos de un área específica dentro de la organización, es un almacén para los datos por departamentos.

3.2.2. *Data Lake*

Data Lake o lago de datos ha sido pensado para enfrentar los desafíos de *Big Data*, su principal característica es poder albergar datos estructurados, semi estructurados y no estructurados, permitiendo utilizarlos de acuerdo a las necesidades del negocio y sin necesidad de saber cuáles serán los análisis aplicados, un Lago de datos permite conservar todos los atributos, especialmente cuando no se conoce al alcance. Conservar, por ejemplo, los logs de navegación, estados de redes sociales, fotos, todo esto sin saber previamente cómo estos datos serán cruzados en el futuro, he aquí un buen ejemplo de *Data Lake*.

"En términos generales, los lagos de datos se comercializan como plataformas de gestión de datos para toda la empresa para analizar fuentes dispares de datos en su formato nativo", dijo Nick Heudecker, director de investigación de Gartner. "La idea es simple: en lugar de colocar los datos en un almacén de datos construido específicamente, se mueve a un lago de datos en su formato original, lo que elimina los costos iniciales de la ingesta de datos, como la transformación. Una vez que los datos se colocan en el lago, está disponible para análisis por todos en la organización".

Un *Data Lake* tiene como principal característica almacenar gran cantidad de datos en su formato original y utiliza una arquitectura plana para guardar los datos y cada uno de estos elementos tiene un identificador único, y es en el momento de la fase de análisis que los datos son reagrupados y que una eventual estructura es creada.

Andrew White, vicepresidente y distinguido analista de Gartner menciona que los *Data Lake* espera resolver dos problemas: Los silos de información y en lugar de tener bases de datos independientes, puede combinar estas fuentes en el *Data Lake*, reduce los costos mediante la reducción de servidores, licencias. El otro problema que resuelve, es que aborda las iniciativas de *Big Data* ya que estos proyectos requieren de una gran cantidad de información y variada que no puede ser soportado en un sistema tradicional. "Abordar estos dos problemas con un lago de datos ciertamente beneficia a TI en el corto plazo ya que la TI ya no tiene que pasar tiempo entendiendo cómo se usa la información - los datos simplemente se vierten en el lago de datos".

Un *Data Lake* también representa algunos retos para las empresas, como determinar la calidad de los datos ya que por definición un *Data Lake* acepta cualquier tipo de dato, sin supervisión y gobierno. Si no existe un gobierno para mantenerlo se puede convertir en un pantano de datos. Un *Data Lake* representa también algunos riesgos como seguridad y control de acceso, ya que los datos pueden ser colocados sin supervisión, los datos que se pueden almacenar pueden ser de carácter privado y representan una exposición de riesgo. Las capacidades de seguridad aún son básicas.

Actualmente *Hadoop* es la tecnología más común donde se utiliza el concepto de *Data Lake*. En este escenario todos los datos se cargan en *Hadoop* para posterior aplicar herramientas de análisis y minería de datos. Este tipo de prácticas incrementan el valor de los datos en una organización, otras fuentes de datos y datos empresariales, permiten una visión con un gran alcance para resolver problemas de negocio.

3.3. Framework Mapreduce y Hadoop

El análisis de *Big Data* requiere modelos de programación que cumplan con todas las dimensiones de *Big Data* (3Vs), estos modelos de programación se basan

principalmente en el paralelismo masivo, adaptándose a distintos ambientes de acuerdo a las necesidades de procesamiento. Entre ellas se pueden destacar los modelos DAG (*Directed Acyclic Graph*), BSP (*Bulk Synchronous Parallel*), *Stream processing*, y uno de los más utilizados, *MapReduce*.

Hadoop se ha convertido en una de las implementaciones más populares para procesar datos de gran tamaño en el orden de terabytes y petabytes y actualmente existen proyectos alrededor de Hadoop para distintos tipos de conjuntos de datos como:

- Avro: es un sistema de socialización de datos.
- Chukwa: es un sistema de recopilación de datos para la gestión de grandes sistemas distribuidos. HBase: es una base de datos distribuida y escalable que admite el almacenamiento estructurado de datos de grandes tablas.
- Hive: es una infraestructura de datos warehouse que proporciona resúmenes de datos y consultas ad hoc.
- Pig: es un lenguaje de alto nivel de flujo de datos y un framework de ejecución para el cálculo paralelo.
- ZooKeeper: es un servicio de coordinación de alto rendimiento para aplicaciones distribuidas

La arquitectura del framework se detalla a continuación en la Figura 13:

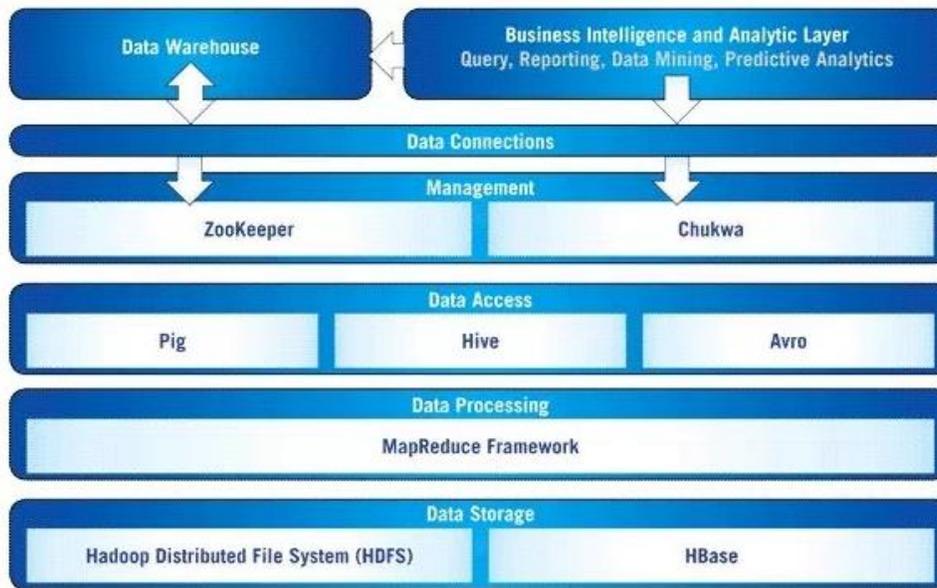


Figura 13. Framework MapReduce and Hadoop.

Tomado de: (Krazytech.com, 2017).

3.3.1. Ecosistema *Hadoop*

Hadoop es un proyecto de código abierto que fue fundado originalmente por *Yahoo* en el 2006 y creado por Dong Cutting y Mike Cafarella, actualmente este proyecto está liderado por la Fundación Apache, que soporta aplicaciones distribuidas bajo una licencia libre, y trae herramientas para el análisis de grandes volúmenes de información en el orden de terabytes y petabytes. *Hadoop* es un marco de trabajo de código abierto que permite almacenar y procesar grandes datos en un entorno distribuido en clústeres de ordenadores utilizando modelos de programación simple. Esta aplicación *Hadoop* trabaja en un entorno que proporciona almacenamiento distribuido y el cálculo a través de grupos de ordenadores. *Hadoop* está diseñado para escalar desde un único servidor a miles de máquinas, cada una ofreciendo computación y almacenamiento local, con un muy alto grado de tolerancia a las fallas.

En lugar de depender del hardware de alta gama, la fortaleza de estos clusters se debe a la capacidad que tiene el software para detectar y manejar fallas al nivel de

las aplicaciones (IBM, 2014). El sitio oficial del proyecto *Apache Hadoop* considera que este proyecto tiene 4 módulos:

- *Hadoop Common.*
- *Hadoop Distributed File System (HDFS).*
- *Hadoop YARN.*
- *Hadoop Map Reduce.*

3.3.2. *Hadoop Distributed File System (HDFS)*

Hadoop Distributed File System (HDFS) es el principal sistema de ficheros distribuidos de hadoop, que tiene mucha semejanza con otros sistemas de ficheros sin embargo existen pocas características que hacen la diferencia. HDFS tiene alta tolerancia a errores y tiene la capacidad de funcionar en hardware de bajo costo. HDFS provee un gran rendimiento para el acceso a los datos de aplicación y es adecuado para aplicaciones que tienen grandes conjuntos de datos. HDFS otorga facilidades con algunos requisitos POSIX (Portable Operating System Interface, Interfaz de Sistema Operativo Portable) que permitir el acceso de transmisión a los datos del sistema de archivos desde otras aplicaciones en tiempo real. HDFS inicialmente fue concebido como una solución de infraestructura para el proyecto de motor de búsqueda web *Apache Nutch*. HDFS es parte del proyecto *Apache Hadoop Core* (hadoop.apache.org, 2016).

HDFS tiene una arquitectura maestra/esclavo, con un servidor maestro (NameNode) que gestiona el espacio de nombres del sistema de ficheros y regula el acceso a los archivos por parte del cliente, mientras que los esclavos (DataNodes) se encargarán de almacenar datos y ejecutar cálculos (Ver Figura 14).

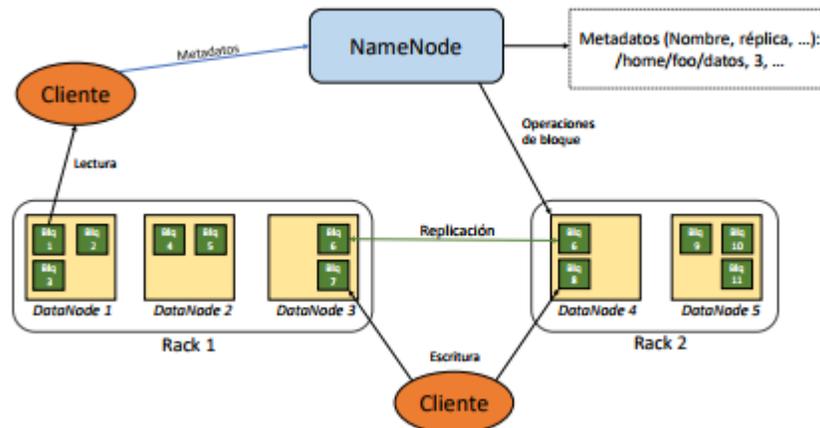


Figura 14. Arquitectura de HDFS.

Tomado de: (Fundación Apache, 2016).

3.3.3. Hbase

Hbase es una base de datos *NoSQL* de código abierto, su principal particularidad es proporcionar acceso en tiempo real de lectura y escritura a grandes volúmenes millones de filas y columnas que utiliza variedad de estructuras y esquemas diferentes.

“*Utilice Apache HBase* cuando necesite acceso aleatorio, en tiempo real de lectura / escritura a sus datos grandes. El objetivo de este proyecto es el alojamiento de tablas muy grandes - miles de millones de filas de X millones de columnas - encima de los clústeres de hardware de las materias primas. *Apache HBase* es una base de datos de código abierto, distribuida, versionada, no relacional modelada después de *Bigtable* de *Google*. De acuerdo con lo expuesto por Chang, *et al.*, (2014), un sistema de almacenamiento distribuido para datos estructurados como *Bigtable* aprovecha el almacenamiento de datos distribuidos proporcionado por el sistema de archivos de *Google*. En el caso de *Apache HBase* proporciona

capacidades similares a *Bigtable* en la parte superior de *Hadoop* y HDFS” (Fundación Apache, 2017).

Fundación Apache destaca como las principales características y beneficios de HBase de la siguiente manera:

- Escalabilidad lineal y modular.
- Alta consistencia en lectura y escritura.
- Compartición y balanceo de tablas automático y configurable.
- Soporta trabajos de Hadoop MapReduce con tablas de HBase.
- Fácil acceso del cliente con API Java.
- Búsquedas en tiempo real.
- Modelo de datos acomoda una amplia gama de casos de uso.

Hbase maneja una arquitectura Maestro/esclavo, en donde el servidor maestro coordina los servidores de su región, los servidores de la región son los encargados de escribir, leer y ejecutar otras operaciones con la data.

3.3.4. *Map Reduce*

Según el sitio oficial de Apache Hadoop, “*MapReduce* es un *framework* de software para escribir fácilmente aplicaciones que procesan grandes cantidades de datos (conjuntos de datos de varios terabytes) en paralelo en grandes clusters (miles de nodos) de hardware de productos de forma fiable y tolerante a fallos. Un trabajo *MapReduce* normalmente divide el conjunto de datos de entrada en bloques independientes que son procesados por las tareas del mapa de una manera completamente paralela. El marco ordena las salidas de los mapas, que luego se introducen en las tareas de reducción. Normalmente tanto la entrada como la salida del trabajo se almacenan en un sistema de archivos. El marco se encarga de programar tareas, supervisarlas y volver a ejecutar las tareas fallidas” (Fundación Apache, 2017).

Se puede decir que *MapReduce* es un flujo de datos distribuido que contiene dos fases, aunque la segunda se divide en dos:

- Map
- Reduce (barajado de datos y reduce)

La función Map divide el trabajo y se encarga del mapeo que se aplica a cada elemento de la entrada de datos. Se agrupan todos los pares asociados con la misma clave creando un nodo para su ejecución y los pasa a la función Reduce. La función Reduce se ejecuta en paralelo por cada nodo de la función Map y un listado de datos relacionados con dicha clave, los cuales se ordenan, se simplifican y como salida un archivo por cada tarea *Reduce*.

En conclusión, una cantidad determinada de archivos de entrada generarán un delimitado número de mapas de tareas para ser ejecutados y cada mapa generará tantos archivos de salida como tareas *Reduce* hayan configuradas en el sistema. Un ejemplo claro de la aplicación de MapReduce es obtener palabras de un grupo de textos y el número de veces que aparecen. MapReduce recibe los textos y los separa en segmentos que se van a mapear para obtener las palabras que conforman el texto. En el proceso de combinado todas las palabras iguales se agrupan. Por último, se reordenan las agrupaciones y se combinan para luego hacer la reducción y devolver el resultado como se muestra en la Figura 15:

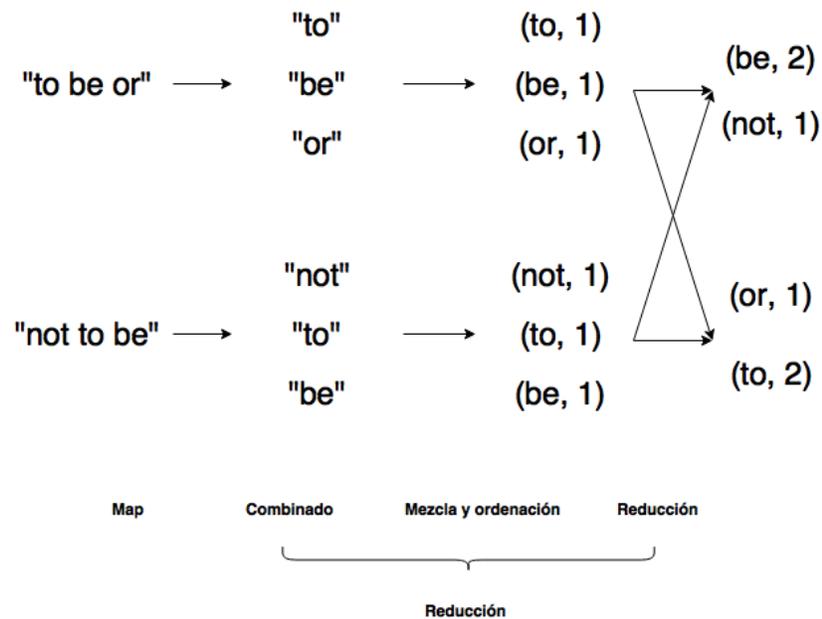


Figura 15. Aplicación de MapReduce para obtener palabras de un texto.

Adaptado de: (Geeky theory, 2017).

3.3.5. Apache Pig

Apache Pig fue desarrollado por *Yahoo* en el 2006, para luego ser adoptado por la Fundación Apache en el año 2007, a partir del 2008 surge la versión inicial como parte del proyecto *Hadoop*. *Apache Pig* es una herramienta que permite el análisis de grandes volúmenes de datos que consta de un lenguaje de programación de alto nivel para desarrollar programas de análisis, muy sensibles a la paralelización lo que a su vez permite manejar grandes volúmenes de datos.

La infraestructura de Pig en la actualidad se compone de:

- **Compilador:** Permite escribir secuencias de programas *Map-Reduce*.
- **La capa lingüística de Pig:** se compone actualmente de un lenguaje textual llamado *Pig Latin*, que tiene las siguientes propiedades clave:

- Facilidad de programación. Permite crear tareas complejas compuestas que se codifican explícitamente como secuencias de flujo de datos que son fáciles de escribir, comprender y mantener.
- Oportunidades de optimización. Admite codificar tareas permitiendo una ejecución óptima de las mismas.
- Extensibilidad. Permite crear funciones para procesamiento.

De esta manera IBM menciona:

“Pig puede utilizarse para crear algoritmos *MapReduce* de una manera más simple y sencilla, generando líneas de código mucho más legibles y mantenibles en el tiempo. Pig proporciona una abstracción útil sobre *MapReduce*; es decir, permite escribir sentencias de manipulación de datos y queries en un lenguaje de alto nivel y gracias a su modelo *MapReduce* subyacente es capaz de paralelizar automáticamente y escalar las operaciones realizadas proporcionando un fuerte apoyo para el trabajo con conjuntos de datos muy grandes.” (IBM, 2012).

3.3.6. *Apache Hive*

Apache Hive es un *framework* creado por Facebook que permite trabajar con HDFS, facilita la lectura y escritura y administración de grandes conjuntos de datos. *Apache Hive* soporta grandes volúmenes de datos bajo HDFS *Hadoop* que ofrece un programa de consultas basado en SQL, que se lo denomina HiveQL y tiene algunas características:

- Indexación.
- Tipo de almacenamiento como texto, RCFile, HBase, otros.
- Operaciones sobre datos comprimidos.

- Funciones definidas por el usuario que permite trabajar con fechas, textos y otras herramientas de *Data Mining*.
- Consultas estilo SQL que son convertidas a *MapReduce*.

Las operaciones que se ejecutan sobre *Apache Hive* son muy similares a las operaciones SQL, sin embargo, hay muchas diferencias en la estructura y funcionamiento con una base de datos relacional. *Hive* en este caso no compara la estructura de los datos con la estructura de las tablas antes de insertar la información, en su lugar posteriormente se hacen comprobaciones cuando se leen los datos. *Hive* puede cargar datos de forma dinámica sin comprobar la estructura de las tablas, esto permite mejorar el tiempo de carga.

3.3.7. *Apache Avro*

Avro es un *framework* de serialización de datos creado por *Doug Cutting*, creador de *Hadoop* que forma parte del proyecto *Apache Hadoop*. En *Avro* se puede almacenar datos y acceder rápidamente desde varios lenguajes de programación, utiliza JSON (*JavaScript Object Notation* - Notación de Objetos de JavaScript) para definir datos y protocolos y serializa los datos en formato binario.

Avro tiene 3 ventajas sobre otras soluciones de serialización como *Thrift*, *Protocol Buffers*, que son:

- Escritura dinámica: *Avro* no requiere que se genere código para la construcción de sistemas de procesamiento de datos ya que siempre van acompañados de un esquema que permite el procesamiento completo.
- Datos no etiquetados: Como el esquema está presente cuando se leen los datos da como resultado un tamaño de serialización más pequeño con menos codificación.

- Cuando un esquema cambia, tanto el esquema antiguo como el nuevo están siempre presentes al procesar datos.

3.3.8. *Apache ZooKeeper*

- *Apache ZooKeeper* es un proyecto de la fundación *Apache* que permite la coordinación de procesos distribuidos cuya arquitectura soporta alta disponibilidad a través de servicios redundantes. *ZooKeeper* inicialmente fue construido por *Yahoo*. Para acceder a todas sus aplicaciones de manera más eficiente, luego se convirtió en un estándar para el proyecto Hadoop. Hbase y otros *frameworks* distribuidos. Por ejemplo, HBase utiliza *ZooKeeper* para realizar un seguimiento del estado de los datos distribuidos.

Apache ZooKeeper es un *framework* que se utiliza en un grupo de nodos para coordinar entre sí y mantener los datos compartidos con técnicas de sincronización utilizando el enfoque a prueba de fallos de sincronización.

ZooKeeper ofrece algunos beneficios (Fundación Apache, 2012):

- Proceso de coordinación distribuida simple
- Sincronización - La exclusión mutua y la cooperación entre los procesos del servidor.
- Los mensajes ordenados
- Serialización - Codificar los datos de acuerdo con reglas específicas.
- Confiabilidad
- La atomicidad

3.3.9. Apache Chukwa

Apache Chukwa es un *framework* de código abierto para análisis en tiempo real y captura de datos que trabaja con Hadoop que permite procesar grandes volúmenes de datos y se construye sobre el sistema de archivos distribuido *Hadoop* (HDFS) y *MapReduce*. *Chukwa* posee un grupo de potentes herramientas para visualizar, controlar y analizar los datos.

“*Chukwa* se construye sobre agentes funcionando en cada uno de los nodos que se han de supervisar, recolectando los archivos de registro de las aplicaciones ejecutadas en él. Estos archivos se envían a los llamados *collectors* y se almacenan en el HDFS” (Fundación Apache, 2012).

Apache Hadoop es un *framework* que a lo largo de los últimos años ha crecido en popularidad debido a que con *Hadoop* se pueden implementar para el procesamiento de grandes volúmenes de datos sin requerir una gran infraestructura. Empresas como *Adobe*, *AOL*, *eBay*, *Facebook*, *Google*, *IBM*, *LinkedIn*, *Twitter* o *Yahoo!*, han adoptado *Hadoop* por sus grandes beneficios para procesar gran variedad y cantidad de datos y con la posibilidad de almacenar datos en arquitecturas distribuidas sin ninguna dificultad y de procesarlas en paralelo, otras cualidades muy estimadas de este software de código abierto son su estabilidad, sus opciones de ampliación y el gran número de funciones.

3.4. Almacenamiento NoSQL

En el término NoSQL significaba No SQL, pero con el paso del tiempo el término No se ha cambiado por NOT ONLY, que reflejan las múltiples facetas de las bases de datos no relacionales. Las bases de datos NoSQL han sido diseñadas para manipular grandes volúmenes de datos de manera muy rápida, y no siguen el modelo de entidad relación típico de las bases de datos tradicionales.

Las bases de datos relacionales, aunque tienen unas características muy potentes para el manejo de tipo de datos mediante lenguaje SQL, su rendimiento disminuye cuando se aumenta el volumen de datos. Las bases de datos tradicionales basan su funcionamiento en tablas, juntas y transacciones, mientras que las bases de datos NoSQL no requieren una estructura de datos en forma de tablas y relaciones entre ellas, y no imponen un esquema prefijado de tablas. Las bases de datos NoSQL no utilizan un modelo entidad relación por ende tampoco utilizan lenguaje SQL y proporciona mayor flexibilidad ya que no requiere estructuras fijas. Otras de las ventajas de las bases de datos NoSQL son: manipulación de datos no estructurados, escalabilidad, rápido acceso, y distribución esto se debe a que no requieren de índices, funciona muy bien con hardware de bajo costo.

Las bases de datos NoSQL son las más idóneas para la lectura y escritura de grandes volúmenes de datos y cuando se requiere dar atención a millones de usuarios. Por estas razones grandes redes sociales como Facebook, Twitter, Instagram, LinkedIn, Google, utilizan base de datos NoSQL como soporte para su almacenamiento de datos.

Así, como tiene varias ventajas también tiene desventajas como: al ser de código abierto poseen un soporte diferente al soporte que ofrecen las compañías comerciales a sus productos. La comunidad open source, si bien podría responder a cualquier problema, posiblemente se puedan tardar más que una compañía comercial de código propietario, falta de experiencia, problemas de compatibilidad, a diferencia de las bases de datos relacionales, que comparten ciertos estándares, las bases de datos NoSQL tienen pocas normas en común. Cada base de datos NoSQL tiene su propia API, las interfaces de consultas son únicas y tienen peculiaridades.

Las bases de datos NoSQL son difíciles de clasificar ya que existen más de 150 sistemas de bases de datos NoSQL, existen varias aproximaciones, pero por lo general se suele agrupar en 4 grandes grupos:

- Orientadas a clave-valor(*Key-value*)
- Orientada a documentos
- Orientada a grafos
- Orientada a columnas y a *Bigtable*

En conclusión, las bases de datos NoSQL constituyen un ecosistema de información, y se están convirtiendo en alternativas viables a las bases de datos relacionales para muchas aplicaciones y empresas como los operadores de telecomunicaciones móviles que requieren sacar el máximo provecho de la información que se genera dentro.

3.5. *Big Data* & Internet de las cosas (IOT)

El concepto de IoT (*Internet of Things*, Internet de las cosas) fue acuñado por Kevin Ashton, quien formó parte de un equipo que descubrió como conectar dispositivos a Internet a través de RFID, desde ese momento se utilizó por primera vez la frase “Internet de las cosas”. Gartner define al Internet de las cosas como “es la red de objetos físicos que contienen tecnología integrada para comunicarse y sentir o interactuar con sus estados internos o el ambiente externo.”

El IOT es una tecnología que ha ido en crecimiento, cuenta con sensores conectados a internet de una amplia variedad de “cosas”, que puede tomar mucha complejidad ya que puede haber una infinidad de posibilidades de datos generados de conexiones cableadas, inalámbricas, sensores en movimiento, estáticos, velocidades de señal muy altas, las “cosas” pueden ser cualquier objeto (vivo o

inanimado), por ejemplo, un teléfono móvil se convierte en un IoT multi-sensor. El IoT tiene la posibilidad de cambiar el futuro.

Las estimaciones a futuro de las “cosas” conectadas dependen de varios factores, sin embargo, las cifras más utilizadas son las de Cisco, y se prevé que para el 2020, 50.000 millones sean los dispositivos conectados a Internet (Ver Figura 16).

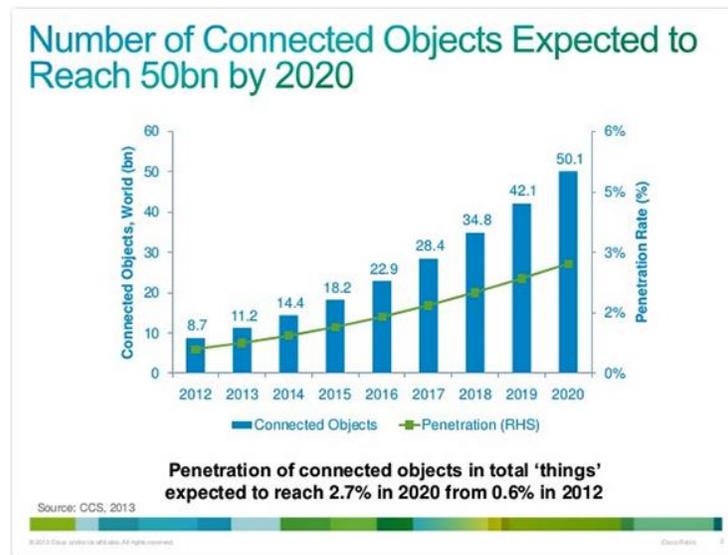


Figura 16. Expectativa de objetos conectados al 2020.

Tomado de: (CCS, 2013).

"El IOT en sí ayudará a las transformaciones digitales, pero tardará de 5 a 10 años en lograr la adopción de corriente", dijo Alfonso Velosa, vicepresidente de investigación de Gartner. Como se muestra en la Figura 17 el IOT ha tenido un crecimiento y se mantienen ya algunos años en la cúspide de tecnologías emergentes que más impacto tienen en las organizaciones.

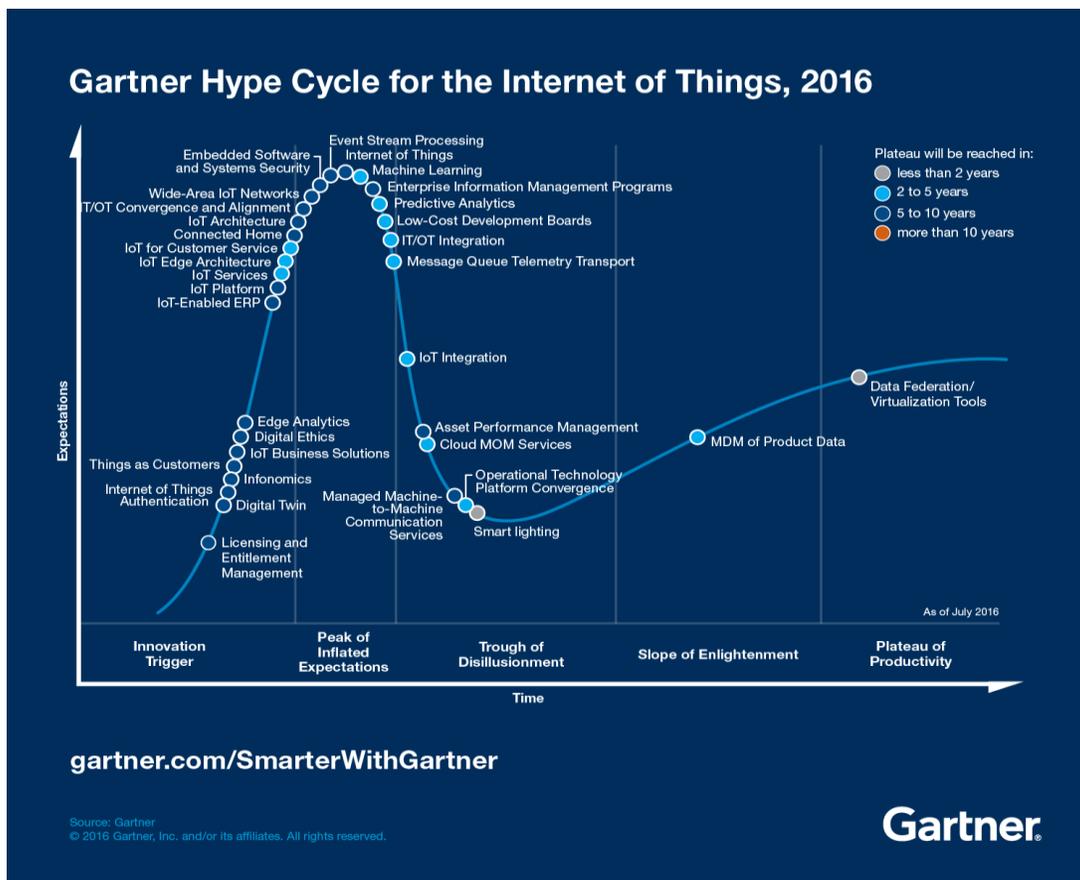


Figura 17. Hype Cycle del Internet de las cosas.

Tomado de: (Hype Cycle Gartner, 2016).

El IOT y Big Data están claramente interconectados, la expansión de las ciudades, miles de millones de “cosas” conectadas a internet que generan volúmenes de datos con una mezcla de información estructurada y no estructurada que llega en tiempo real. IOT y *Big Data* se convierten en una posibilidad de transformar muchas áreas de negocio, una nueva generación de aplicaciones IOT que permitirán atender las necesidades del negocio como:

- Mantenimiento predictivo
- Prevención de fraude
- Utilización de activos
- Planificación y recuperación de desastres

- Minimización del tiempo de inactividad
- Optimización del uso de la energía
- Gestión del rendimiento de la red

El IOT está jugando un rol muy importante en el crecimiento económico de muchas industrias, según Gartner predice que IoT generará 300 mil millones de dólares en ingresos para el año 2020, al mismo tiempo esto representa un reto para los directores de marketing, de acuerdo con el sitio web Salesforce, 2017, afirma que: “Los dispositivos conectados permiten a los directores de Marketing conocer cómo se evalúan y utilizan sus productos, en qué etapa del proceso están sus prospectos, y saber qué factores, potencialmente, podrían influir en el comportamiento de compra. El IOT no se trata solo de máquinas conectadas; se trata de productos conectados y marketing, también” (Salesforce.com, 2017).

3.6. *Big Data* en tiempo real

La analítica en tiempo real tiene muchos casos de uso, el avance de nuevas tecnologías como *Hadoop* ha permitido flexibilidad y escalabilidad para almacenar datos a gran velocidad y realizar consultas sin cambiar la estructura de los datos. Estas nuevas arquitecturas permiten reducir el tiempo de consultas a fracciones de segundo, consultas que antes podían llevar horas y en muchos casos hasta días, eso es prácticamente la velocidad del pensamiento.

“*Big Data* en tiempo real no es solo un proceso para almacenar *petabytes* o *exabytes* de datos en DWH,” dice Michael Minelli Co-autor de *Big Data, Big Analytics*. “Esto trata de la capacidad de tomar las mejores decisiones y tomar acciones significativas en el tiempo adecuado, se trata de detectar fraude mientras alguien está pasando una tarjeta de crédito o activando una oferta mientras el comprador está de pie en una línea de pago. Se trata de combinar y analizar datos para que se pueda tomar la acción correcta, en el momento adecuado y en el lugar correcto.”

La capacidad de almacenar datos en tiempo real no es nueva, lo nuevo es sacar provecho a estos datos de forma rápida y económica, marcos de referencia como *Hadoop* y *MapReduce* permiten a través de un procesamiento distribuido almacenar y recuperar datos en tiempo presente. *Hadoop* en un inicio no estaba diseñado para el procesamiento en tiempo real, sin embargo, ahora se ha integrado soluciones como *Apache Storm* que puede ser utilizado en la capa superior del *framework Hadoop* que ofrecen agregación de probabilidades en tiempo real. *Apache Storm* es un *Framework* creado por *twitter* que está compuesto por varios componentes de código abierto especialmente *ZooKeeper* para administración de clústeres, *ZeroMQ* para mensajes de multidifusión y *Kafka* para mensajes en cola, que sirve para recuperar streams de datos en tiempo real desde múltiples fuentes de manera distribuida, tolerante a fallos y en alta disponibilidad

Lambda es una nueva arquitectura que su objetivo es soportar el procesamiento en tiempo real, resolviendo el problema de calcular funciones en tiempo real mediante 3 componentes (Ver Figura 18) y son:

- Capa de carga (*Batch Layer*)
- Capa de servicio (*Serving Layer*)
- Capa de velocidad (*Speed Layer*)

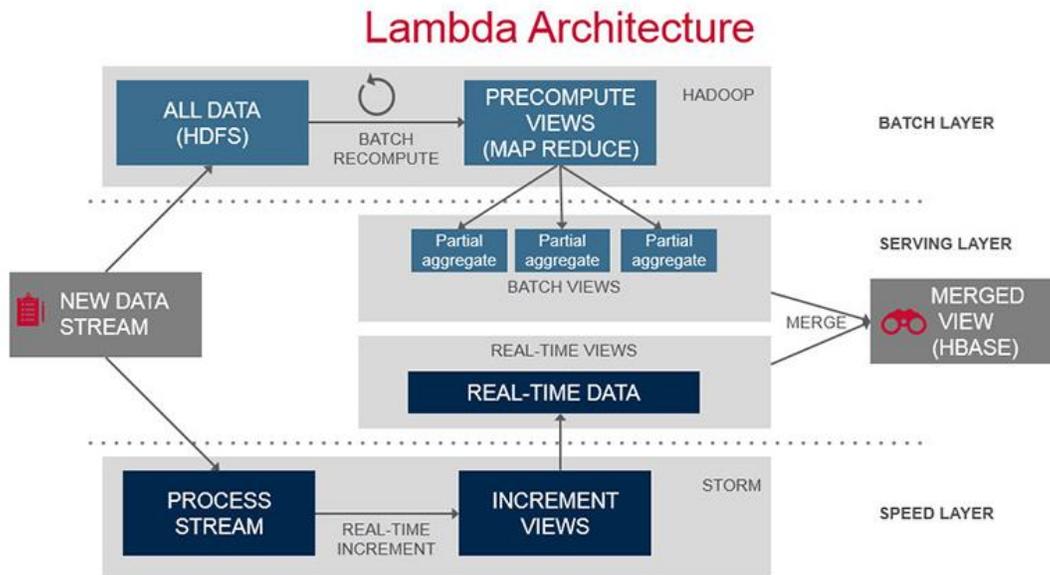


Figura 18. Overview of the Lambda Architecture.

Tomado de: (MapR.com, s.f).

John Akred, gerente senior Grupo de Innovaciones Tecnológicas Emergentes de *Accenture* dice "Normalmente cuando hablamos de sistemas en tiempo real o casi en tiempo real, lo que queremos decir son arquitecturas que le permiten responder a los datos tal como lo recibe sin necesariamente persistirlo en una base de datos en primer lugar".

La implementación de una solución en tiempo real debe requerir un análisis previo de costo, recursos, tiempo, seguridad y privacidad, sin embargo, el principal reto en la analítica en tiempo real no es soportar la tecnología, si no está en la transformación de las personas y los procesos para tomar las decisiones correctas.

3.7. *Big Data & Cloud Computing*

La computación en la nube (*Cloud Computing*) es una tecnología que permite ofrecer servicios de computación, servidores, almacenamiento, bases de datos, redes, software, sin la necesidad de mantener *Hardware* informático local, utiliza una red de servidores conectados a través de la Internet.

Cloud Computing tiene algunas características y beneficios como:

- Costo: el *Cloud Computing* elimina el gasto de compra y mantenimiento de software y hardware, mantenimiento de centros de datos, electricidad y recurso humano para su administración.
- Confiabilidad: El Cloud Computing ofrece respaldo de datos, recuperación de desastres y continuidad de negocio.
- Elasticidad y Escalabilidad: Las aplicaciones cloud pueden aumentar a medida que aumentan las necesidades de computación, es decir, que podemos utilizar el 20% de la aplicación hoy y mañana podemos acceder al 80% de la misma.
- Pago por uso: Esta tecnología permite a los usuarios finales solo pagar por lo que usan, es decir que el usuario decide qué aplicaciones utilizar y escoger entre cuáles son gratuitas y las que no.
- Seguridad: La seguridad ha sido un tema muy discutido cuando se habla de Cloud Computing, sin embargo, los proveedores que ofrecen estos servicios ofrecen todas las seguridades necesarias, tanto físicas como de software, de forma que no exista una fuga de información, ni falla en la integridad de los datos.

De acuerdo con Ed Anderson, vicepresidente de investigación de Gartner, manifiesta que: "Las estrategias basadas en la nube son la base para mantenerse relevantes en un mundo acelerado, por lo cual, el mercado de servicios en la nube ha crecido hasta tal punto que ahora es un porcentaje notable del gasto total en TI, ayudando a la creación de empresas de nueva generación y proveedores nacidos en la nube".

Con el tiempo se han desarrollado 3 modelos de arquitectura en la nube, todos estos modelos comparten las mismas características y beneficios. Estos 3 modelos son:

- Nube Privada: La nube privada es un ambiente controlado no abierto al público, está enfocada al uso exclusivo de empresas y puede ubicarse físicamente en el centro de datos de la organización, los servicios y la infraestructura se mantienen en una red privada para el uso de empleados, socios y clientes. Esta nube privada puede, o no puede ser gestionada por un proveedor para el uso de la empresa. La nube privada está de alguna manera condicionada a la capacidad de TI de la empresa y permite aprovechar los recursos computacionales con más calidad, es decir, es posible disminuir el tiempo de despliegue de nuevos sistemas y aplicaciones y esta escalabilidad no solo aumenta la productividad de la organización si no también permite reducir costos.

Cuando se habla de nube privada se despierta un interés por la seguridad y privacidad, este modelo los niveles de seguridad dependerán de la organización y suelen ser más confiables ya que los servidores y data centers operan exclusivamente en el ambiente corporativo.

- Nube Pública: La nube pública es el modelo más común de *Cloud Computing* a la cual se puede acceder a través de cualquier dispositivo y cualquier navegador, el acceso es idéntico para todos los usuarios y no tienen costo en la mayoría de casos. Los recursos informáticos se entregan a través de internet, todo el *hardware*, *software* y otra infraestructura de soporte son propiedad y administrados por el proveedor de la nube.
- Nube Híbrida: El modelo de nube híbrida combina las mejores características de la nube privada y pública, proporciona la oportunidad de extender los servicios internos con orquestación y automatización entre las dos. La Nube Híbrida permite ejecutar cargas de trabajo críticas en la nube privada y la

nube pública para procesos de apoyo, ráfagas de carga de trabajo o picos de demanda.

La nube híbrida ofrece a las empresas mayor flexibilidad, obtener un mejor rendimiento del negocio en múltiples frentes de manera simultánea: costos, capacidad de administración, acceso a la información, implementación de nuevas funcionalidades, coordinación y colaboración, continuidad y seguridad de negocio, innovación empresarial y crecimiento capaz de adaptarse al tamaño y sector de la empresa.

La consultora Gartner indica: “el modelo híbrido cambiará el papel del profesional TI, ya que requiere profesionales TI externos e internos para apoyar las capacidades de negocio de la empresa creando estrategias entre la nube privada y la pública, con el objetivo de crear un traje a medida de la compañía”.

Los proveedores de Cloud Computing tienen algunos modelos para entregar los servicios:

- Software como Servicio (*Software as a Service - SaaS*)

Es un modelo de distribución de software donde permite a los usuarios conectarse a aplicaciones en la nube a través de Internet. Este modelo de distribución es clave porque cambia totalmente la forma de usar la tecnología, proveedores cloud alquilan el uso de una aplicación para una organización y sus usuarios se conectan a ella a través de Internet, normalmente con un explorador web.

- Infraestructura como Servicio (Infrastructure as a Service - IaaS)

En este modelo los proveedores arriendan sus servidores para que otras empresas las puedan usar de acuerdo a sus necesidades, proporciona acceso a recursos informáticos situados en un entorno virtualizado, la nube y permite desarrollar y ajustar las máquinas virtuales a las necesidades del equipo.

Según (*Microsoft, 2017*) IaaS debe cumplir con ciertas características:

- Debe existir la posibilidad de acceder a los recursos sin intervención humana
 - Una aplicación de misión crítica no puede depender de una sola conexión.
 - Garantía de red.
 - Ajustar los recursos a las necesidades del cliente.
- Plataforma como Servicio (*Platform as a Service - PaaS*)

Plataforma como servicio se puede catalogar como la combinación de los otros dos servicios que proporciona un entorno de desarrollo completo en la nube, provee a los desarrolladores un entorno de desarrollo de aplicaciones web sin tener que instalar ninguna herramienta adicional que pueden ir desde aplicaciones muy sencillas a aplicaciones muy sofisticadas. Este modelo también incluye middleware, herramientas de desarrollo, servicios de inteligencia empresarial (BI), sistemas de administración de bases de datos, etc.

Estos modelos de *cloud computing* permiten ofrecer servicios a través de internet que representan oportunidades para desarrollar nuevos productos y servicios, nuevas líneas de negocio, la entrada a nuevos mercados, ahorrar dinero y sacar el máximo provecho de la tecnología. (Ver figura 19).

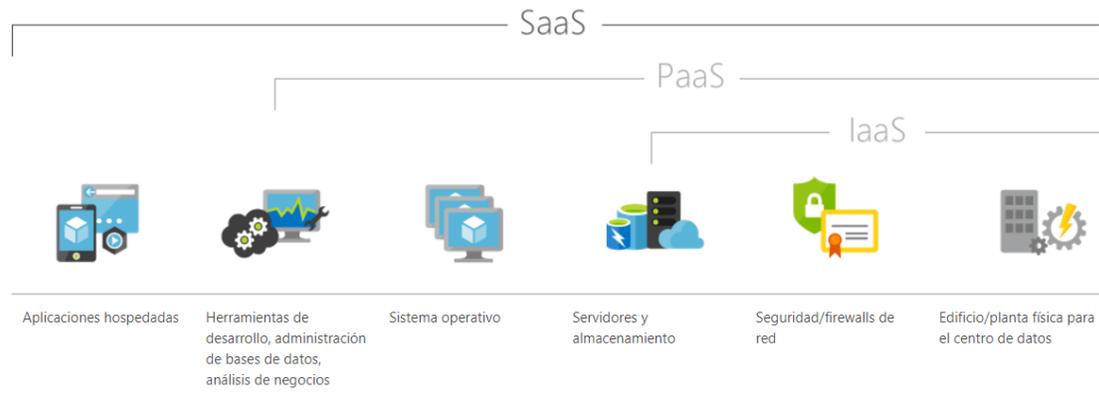


Figura 19. Modelos de cloud computing.

Tomado de: (Mapr architecture, 2017).

La computación en la nube (*Cloud Computing*) y los almacenes de datos en la nube se han convertido en un facilitador para *Big Data*, con un alcance cada vez mayor, una reducción significativa en el costo y una abundante variedad de aplicaciones disponibles.

BDaaS (*Big Data as a service*, *Big Data* como servicio) es un término que hace referencia a servicios que ofrece *Big Data* alojados en la nube. BDaaS puede ser visto como integración específica para IaaS, PaaS y SaaS, que ofrece distintos tipos de análisis. Como, por ejemplo: el monitoreo de contenidos web con interfaz de usuario final que permita un acceso a tablas, un tablero para visualizar y explotar tus datos, con garantía de disponibilidad, escalabilidad y redundancia y con soporte e integración directa a herramientas como *Hadoop*, *Map Reduce* o *Pig*, *Hive* en entornos distribuidos.

Otro buen ejemplo, es el servicio que proporciona IBM, este servicio incluye herramientas de analítica, almacenamiento y aplicaciones para el análisis de en

Twitter, 500 millones de tweets por día y 280 millones de usuarios activos mensuales, información que puede ser utilizada para cualquier tipo de propósito.

En conclusión, BDaaS ofrecerá entre los principales beneficios: procesamiento de datos estructurados, semi estructurados y no estructurados en la nube, escalabilidad bajo demanda, almacenamiento de datos distribuidos y virtualizados. Forbes (2012), menciona que BDaaS es la siguiente generación de *Big Data* que estará al alcance de todos, las empresas tendrán que preocuparse menos en asumir los complejos retos de la construcción de su propia arquitectura de datos y permitirá enfocarse en sacar el máximo provecho al análisis de los datos.

4. MODELOS DE NEGOCIO

4.1. Casos de uso modelo actual

El patrón de consumo en el sector de las telecomunicaciones ha cambiado y se está volviendo cada vez más complejo, el entorno empresarial ha evolucionado rápidamente y los operadores internamente soportan sus decisiones en sistemas tradicionales de BI y esto se está convirtiendo en un papel muy importante para que el modelo de negocios se vuelva obsoleto.

Según la guía de *Big Data Analytics* del *TM Forum* (junio, 2017), la adopción de *Big Data* en los operadores de telecomunicaciones móviles implicaría beneficios para cada una de las siguientes áreas:

- Estrategia
 - Ofertas en tiempo real adaptadas a las necesidades del cliente.
 - Promover el uso de servicios adicionales.
- Cartera de productos

- Mejoramiento de productos.
- Analizar nuevos productos.
- Operaciones
 - Proporcionar atención proactiva.
 - Reparación de red, basadas en experiencia de cliente.
- Rentabilidad
 - Predecir Comportamiento de *churn* (Ver Anexo 2).
 - Asegurar ingresos.
 - Detectar Fraude.

4.2. Casos de uso del nuevo modelo

La industria de telecomunicaciones en todo el mundo se está acercando a un punto de inflexión, la competencia se ha intensificado, los jugadores OTT han aumentado su dominio. Los operadores de telecomunicaciones móviles han invertido esfuerzos en innovar en mercados, canales e iniciativas de red, esto les permitirá un desarrollo exitoso hacia la nueva era digital.

4.2.1. Monetización de los datos

Big Data es una valiosa palanca estratégica que permite a los operadores de telecomunicaciones móviles aprovechar uno de sus activos más importantes, los datos. Monetizar los datos significa aprovechar el potencial de los datos para obtener beneficios adicionales, como dice Peter Sondergaard, vicepresidente de Gartner y su responsable global de investigación, “un dato no sirve para nada a no ser que sepas cómo utilizarlo y cómo actuar con él” (BVEX, 2015). De acuerdo con Ricardo Malhado (2017), analista de IDC "Los datos son activos con un potencial inmenso dentro de las empresas. De hecho, estimamos que, en 2019, el 40 por

ciento de las organizaciones creará nuevas líneas de ingresos para monetizarlos" (BVEX, 2015). Malhado explica que "Según los estudios que IDC hace, el 90 por ciento de las empresas comprenden el valor de los datos, pero muchas de ellas están aún en proceso de intentar aterrizar esos procesos en su organización" (BVEX, 2015).

TM Forum en su *Big Data Analytics Guidebook* (Guía para *Big Data Analytics*) que proporciona orientación a un proveedor de servicios de comunicación (CSP) sobre los componentes principales que se necesitan para la implementación de casos de uso de BDA (*Big Data Analytics*) reales y presenta un diagrama con los casos de uso para la monetización de datos (TM Forum, 2017). (Ver Figura 20).



Figura 20. Casos de uso Big Data Analytics.

Tomado de: (TM Forum, 2017).

Telefónica lleva a cabo algunos proyectos importantes como "Smart Step" es una herramienta que combina tecnologías móviles, geolocalización, sensores, *Big Data* y geo-procesos que proporciona conocimiento basado en el comportamiento de poblaciones basados en los datos móviles anónimos. Esta línea de servicios permite sacar provecho a los datos de clientes anonimizados y agregados en la red celular enfocados a la gestión de movilidad para una planificación más eficiente por ejemplo de los servicios de transporte y turísticos en las ciudades o, en el caso de una emergencia sanitaria.

“El *insight* que se toma es la señal que realizan los clientes de Telefónica a medida que se van moviendo. Cada vez que interactuamos con el celular, ya sea un smartphone o uno bien básico, se genera una señalización que se conecta a la antena Movistar más cercana”, explica Sabrina Muñoz, gerente regional para la Monetización de Grandes Volúmenes de Datos para Hispanoamérica en el Grupo Telefónica. “Luego, se toma esa información de las antenas, se agrega y se la extrapola para representar a un total de la población en determinada zona, más allá de clientes de Movistar”, sigue.

La señalización móvil se convierte en una gran oportunidad a las metodologías tradicionales, como las encuestas, pero con una muestra mucho mayor y más específica. Además, según Muñoz, tiene un mejor rendimiento y porcentaje menor de error y es notablemente más rápido. “Mientras que los datos recogidos en una encuesta tradicional pueden tardar años y años en ser procesados, con *SmartSteps* la información se procesa en sólo dos meses.”

Este es un servicio de analítica y consultoría externa que ofrece: Anonimizar, agregación y extrapolación de los datos y reportes de estacionalidad, frecuencia, origen y destino. (Ver figura 21)



Figura 21. Telefónica Smart Steps.

Tomado de: (Dynamic Insights Telefonica, 2017).

4.2.2. Publicidad Móvil (*Mobile Advertising*)

Ernst & Young en su publicación *The 2015 Global Telecommunications Study*, menciona que “El panorama de nuevas oportunidades de servicio evolucionará dinámicamente en los próximos cinco años. Algunas oportunidades de crecimiento digital, como la publicidad móvil, están establecidas para un fuerte crecimiento, pero los especialistas en servicios establecidos ya cuentan con altos niveles de cuota de mercado, lo que obligará a los operadores a considerar maneras más innovadoras de aumentar la cuota de cartera” (Ver Figura 22).

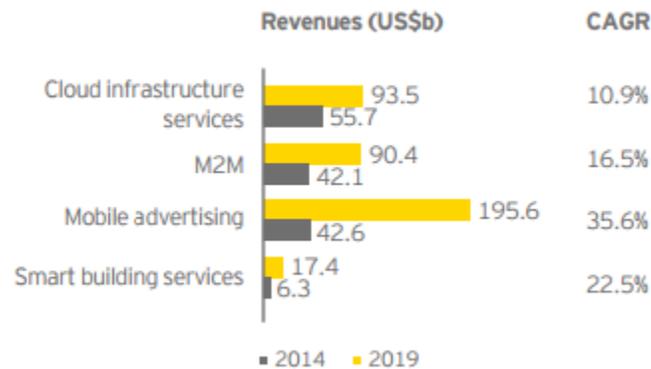


Figura 22. Crecimiento del mercado global de segmentos seleccionados.

Tomado de: (Ovum, Research & Markets, 2015).

La publicidad móvil o *mobile advertising* en inglés es un instrumento de marketing que utiliza dispositivos móviles (*smartphones* y *tablets*) para llegar con anuncios publicitarios a usuarios finales, esta técnica incluye anuncios a través de SMS o *Sat Push* es una alerta que se muestra en las pantallas de los celulares con un contenido publicitario o informativo, publicidad en facturas y pagina web móvil.

La publicidad móvil se produce a través de varios canales entre ellos:

- Web móvil
- Aplicaciones móviles
- Mensajería móvil
- Video / TV móvil
- Mensajes de texto (SMS / MMS)

Verizon Communications es un operador de telefonía móvil de Estados Unidos, la compañía resultó de una fusión entre varios operadores de telefonía móvil incluyendo *Bell Atlantic Mobile*, *AirTouch Communications*, *GTE Wireless* y *Alltel Wireless*. El 12 de mayo del 2015 se fusiona con AOL, AOL es el líder de contenido digital y publicitario. *Verizon Communications* en su reporte anual del 2016 indica: “El Modelo de negocio de AOL se alinea con nuestro enfoque, y creemos que la

combinación de contenidos digitales, la publicidad y la operadora, además de la plataforma de publicidad mejora nuestra capacidad para seguir desarrollando futuros flujos de ingresos”. (Wu, 2016)

Estos ingresos se generan con publicidad mediante anuncios gráficos en la pantalla de los dispositivos y publicidad en búsquedas cuando un consumidor hace clic en un anuncio basado en texto en su pantalla. Los contratos de publicidad pueden ser de 3 tipos, por tiempo, por rendimiento, o por impresión en pantalla.

Según la empresa Emarketer en su informe “*The eMarketer Forecast for 2017*” menciona que “Los ingresos publicitarios digitales de Estados Unidos para Verizon que incluyen los ingresos por visualización y búsqueda de *Millennial Media* y AOL aumentarán un 4,4% a 1,32 mil millones de dólares en 2017. La publicidad en pantalla, por 1,11 mil millones, representará más de cinco veces los ingresos de búsqueda (212,3 millones de dólares).

Big Data es un elemento clave en la publicidad y marketing móvil, la capacidad de captar y analizar grandes cantidades de datos proporciona a la industria de anuncios la capacidad de ofrecer publicidad en tiempo real y personalizada, los datos de las redes sociales junto con la ubicación ayudan a descubrir nuevas relaciones, a detectar tendencias y patrones para obtener información útil que conduce a una ventaja competitiva.

4.2.3. Exposición de datos del cliente

En la búsqueda de incrementar sus ingresos, los operadores de telecomunicaciones móviles crean valor a partir de los datos de sus clientes. Datos como: Ubicación, uso de red, navegación Web, interacción con canales, uso de redes sociales, uso de apps, pueden ser aprovechados por empresas de marketing para desarrollo de nuevos productos o servicios.

AT&T y *Verizon Wireless* han planteado la posibilidad de explotar los datos del cliente para generar nuevas fuentes de ingresos, para ello se requiere un cambio en la política de privacidad en los contratos con los clientes para que terceros puedan hacer uso de la información de forma anónima. *Verizon Wireless* lanzó su *Market Insights* cuyo objetivo es potenciar el alcance de sus iniciativas de marketing. basado en datos con mayor impacto para a las agencias, marcas y socios proporcionando un análisis basado en las preferencias específicas de perfil de cliente. (Verizon Wireless, 2016)

Verizon en su portal de negocios indica: “estamos comprometidos a proteger la privacidad de la información del cliente. La información que se comparte a las empresas se proporciona únicamente sobre una base agregada como un medio para mejorar la experiencia global de los medios de comunicación. Los datos de clientes que no decidan participar, no están incluidos”. (Verizon, 2017)

5. MARCO DE REFERENCIA

La gran cantidad de datos actualmente es un tema de relevancia para todas las industrias y la industria de telecomunicaciones no es la excepción y para sacar el mayor provecho de los datos estructurados, semi estructurados y no estructurados, *Big Data Analytics* es un camino claro para ayudar a implementar grandes proyectos de análisis de datos, es un marco de referencia desarrollado por el *TM Forum* que provee conceptos, prácticas y criterios para el diseño e implementación de procesos, estructuras, flujos de información o infraestructura tecnológica en la empresa. *Big Data Analytics* “proporciona una vista de alto nivel de los componentes funcionales en una gran plataforma de análisis de datos dentro de un gran ecosistema de soluciones de datos. Al separar las capas de responsabilidades entre los distintos componentes funcionales de la plataforma, podemos obtener una vista

más clara de las funciones y responsabilidades y sentar las bases para un entendimiento común del dominio *Big Data Analytics*.” (TM Forum, 2017)

Big Data Analytics tiene varias capas funcionales dentro de una plataforma de análisis de datos, al separar cada componente permite obtener una visión más clara de las funciones y responsabilidades para sentar las bases de un entendimiento global. (Ver Figura 23)

Cada uno de estos componentes funcionales de la Figura 23, cuentan con APIs externas e internas que permite comunicación con otras funciones de la capa externa y aplicaciones de terceros en los respectivos niveles. Los componentes funcionales del diagrama indican la funcionalidad total teórica, es decir, que si se tiene que prescindir de algún componente (a excepción de la capa de ingestión de datos) se puede considerar como opcional dependiendo el caso de uso de la solución.

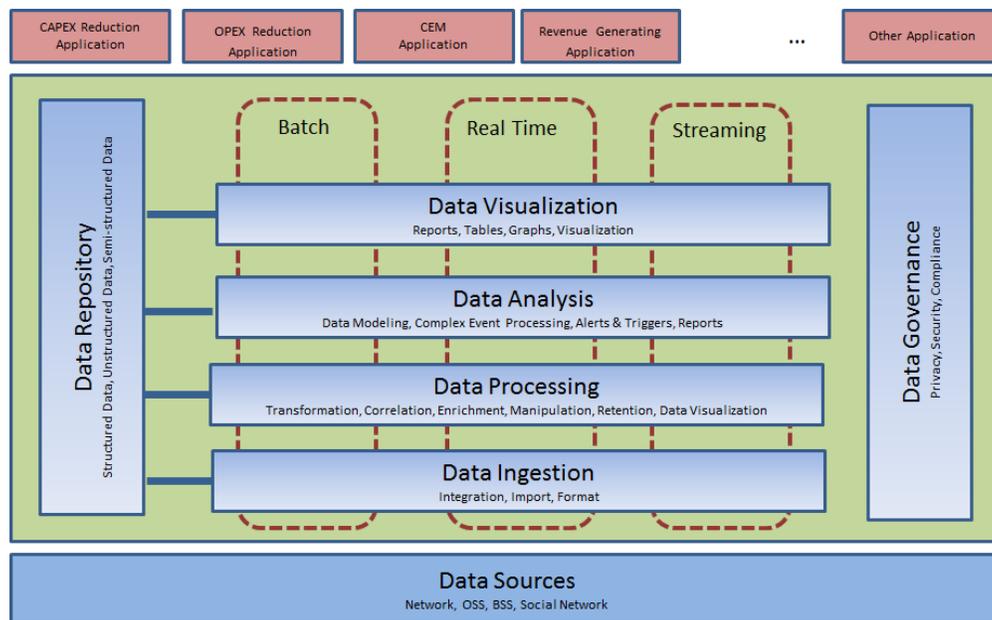


Figura 23. Marco de referencia Big Data Analytics.

Tomado de: (TM Forum, 2017).

En esta vista de alto nivel del marco de referencia *Big Data Analytics* cada una de estas capas son consideradas como una funcionalidad y mas no como un componente arquitectónico.

- *Data Sources* (Fuentes de Datos) corresponde a los proveedores de datos, es distinto al repositorio de datos y no será definida hasta que se haya definido los casos de uso de la solución.
- En la capa superior en los recuadros de color rojo corresponde a los usuarios finales, aplicaciones que utilizan y que sacan provecho de la información.
- En los recuadros con líneas con espacio de color rojo corresponde al tipo de ingestión de datos: carga, gestión y análisis de los datos.
 1. Modo *Batch* o también considerado modo *off-line* (Fuera de tiempo) permite procesar un número finito de datos
 2. Modo *Streaming* o de secuencia permite el procesamiento se realice en ventanas de tiempo.
 3. Modo Tiempo Real corresponde al procesamiento en línea
- Repositorios de datos corresponde si se puede considerar un componente a parte de una arquitectura que además de almacenar datos puede facilitar el flujo de datos entre capas.
- La seguridad, privacidad y el cumplimiento pueden convertirse en un verdadero reto al momento de monetizar y analizar grandes cantidades de datos. Dentro de la gestión de datos se debe tomar en cuenta la seguridad, privacidad y cumplimiento legal y normativo desde el diseño del modelo para ello existen técnicas de anonimización, protección, cifrado, control de acceso que se deben aplicar a cada una de las capas de Big Data Analytics.

5.1. Evaluación del caso comercial

Esta es la primera fase y es imprescindible valorar la implementación de la solución *Big Data*, y de qué manera apoyará al negocio. Una implementación de *Big Data*

involucra a todas las partes interesadas del negocio y afecta en muchos aspectos a una organización, incluidos datos históricos, fuentes de negocio administradas por distintas unidades de negocio.

Un equipo de trabajo con líderes de la alta gerencia que establezcan la estrategia y dirección del negocio y representantes del negocio que puedan ver cómo se relacionan sus problemas de administración de datos.

5.1.1. Fase inicial

- Identificar las necesidades: Se debe considerar el volumen de datos que se va a almacenar, la variedad de datos, procesamiento, escalabilidad y velocidad.
- Valorar las limitaciones: Se debe considerar la capacidad, madurez tecnológica de la organización y sus recursos.

5.1.2. Fase de planificación

- Gestores: Patrocinadores, directores de proyecto, coordinadores y gestores de calidad.
- Diseñadores y arquitectos de datos
- Implementadores
- Operadores

5.1.3. Fase de diseño

- Se establece un diseño de acuerdo a las necesidades de la organización en Infraestructura (Redes, equipos, servidores) y Arquitectura (Protocolos, procedimientos, comunicaciones)

5.2. Capa de Ingestión de datos

La capa de ingestión de datos es la responsable de la integración, importación y formateo de los datos para ingresar en el sistema que se está construyendo o utilizando. Esta capa es de vital importancia ya que se debe tomar en cuenta la Variedad, Velocidad y Volumen.

En la capa de ingestión de datos existen algunas consideraciones de importancia, dudas que deben surgir en la construcción del modelo de referencia como ¿Cuántas fuentes de datos vamos a utilizar?, ¿Qué tamaño tienen los datos?, ¿Es probable que aumente el número de fuentes de datos?, ¿Qué hacemos con los datos que tienen errores?, ¿Qué hacemos cuando hay muy pocos datos o cuándo tenemos demasiados?, esto permitirá tener un mejor entendimiento de los módulos funcionales de esta capa.

5.2.1. Funcionalidades claves en la ingestión de datos

5.2.1.1. Integración

La fase de integración de datos es una combinación de procesos técnicos y de negocio que permite integrar los datos de distintas fuentes, establecer conexiones entre diferentes sistemas. Esta fase de integración permite combinar datos de muchas fuentes en la forma y estructura de una sola aplicación local, en la nube o ambas, como, por ejemplo, los datos completos de un usuario que puede incluir datos extraídos y combinados de marketing, ventas y operaciones, emails que se combinan para formar un informe completo.

En esta fase es fundamental para comprender los distintos tipos de datos con los que se va a trabajar, el rendimiento de la tecnología de integración de datos ya que si el procesamiento es intensivo o complejo el procesamiento se puede tornar muy lento y generar afectaciones en los casos que se planea integrar. (Ver Figura 24)

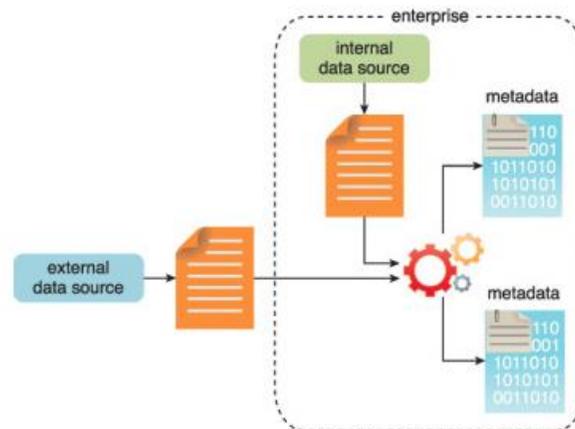


Figura 24. Fase de integración de datos.

Tomado de: (Power Data Fuente, s.f.).

5.2.1.2. Importación

En esta fase se importan todos los datos al sistema desde diversas fuentes externas, permite etiquetar los datos para determinar su origen.

5.2.1.3. Formateo

Esta fase es crucial para poder avanzar a la siguiente, cuyo objetivo es dar formato y estructura a diferentes datos provenientes de diferentes fuentes internas y externas que permita interactuar con todos los componentes de la solución. Los datos pueden llegar en un formato incompatible con la solución *Big Data* esto es más frecuente cuando los datos son de fuentes externas. Esta fase está dedicada a transformarlos en un formato que la solución de *Big Data* subyacente pueda usar para el análisis de datos a una mejor toma de decisiones de apoyo.

5.3. Procesamiento de datos

Los datos ingeridos en la plataforma de *Big Data*, pueden distribuirse en múltiples conjuntos de datos, haciendo necesaria la aplicación de técnicas y herramientas a los paquetes para la extracción de información requerida.

Por ejemplo, el conjunto de datos de tweets está en formato *JSON*. Para poder analizar los *tweets*, la identificación del usuario, la marca de tiempo y el texto del *tweet* se deben extraer y convertir a formato tabular. Además, el conjunto de datos meteorológicos llega en un formato jerárquico (*XML*) y los campos como marca de tiempo, pronóstico de temperatura, pronóstico de velocidad del viento, pronóstico de dirección del viento, pronóstico de nieve y pronóstico de inundación también se extraen y se guardan en forma tabular.

5.3.1. Transformación

Los datos en bruto que se obtienen de diferentes fuentes no necesariamente están en el formato que se requiere, el proceso de transformación tiene como objetivo hacer que los datos sean utilizables, convierte los datos extraídos en formato que se ajuste a los requisitos de la solución *Big Data*, manteniendo en la mayor medida posible la integridad y la información existente en los datos.

El tipo de transformación depende del tipo de análisis y las capacidades de la solución *Big Data*, se lleva a cabo mediante el uso de reglas o la fusión de datos, estas transformaciones pueden ser incluidas en las categorías:

- Comparación
- Fecha y Hora
- Lógico
- Matemáticas
- Estadísticas
- Texto
- Trigonometría
- Codificación
- Gestión de listas

- Gestión de direcciones URL

El crecimiento e importancia de los datos no estructurados en el proceso de transformación, convierten los datos no estructurados de tal manera en que pueden ser integrados de mejor manera con datos estructurados. Este finalmente es el reto más importante que este proceso que enfrenta para la adopción del *Big Data*.

5.3.2. Correlación

Los datos provenientes de diferentes fuentes de datos, requieren que estos conjuntos de datos se asocien entre sí, a través de campos comunes, como puede ser el ID o la fecha. En otros casos, los mismos campos de los datos puede estar en varios conjuntos de datos, de esta manera se requiere un proceso de asociación de los datos para representar el valor correcto, por lo tanto, esta etapa se encarga de integrar conjuntos de datos múltiples para llegar a una vista unificada.

5.3.3. Enriquecimiento

Esta fase pretende combinar varias fuentes de datos ya sean internas o externas sobre una unidad dentro de la empresa que tuviese información relevante, con la finalidad de tener una visión más amplia de dicha unidad. Para combinar esta información se debe identificar todas las aplicaciones y repositorios que contienen la información sobre esta unidad que permita mejorar la calidad de la información y mostrar una perspectiva más completa, como, por ejemplo, se puede enriquecer la información del cliente contenido en CRM, ECM, correo electrónico, entre otros., para incrementar el nivel de confianza del cliente por motivos de sexo, edad, nivel educativo, nivel de ingresos, etc.

5.3.4. Aseguramiento de la calidad de datos

En esta fase se requiere asegurar la calidad de los datos; ya que datos inválidos pueden influir en el resultado del análisis. A diferencia de los datos empresariales estructurados, donde su estructura está predefinida y los datos son validados, los

datos de entrada en una solución de Big Data pueden ser semi-estructurados y no estructurados, esta complejidad en los datos puede dificultar aún más la obtención de un conjunto de restricciones de validación adecuadas.

Esta fase de asegurar la calidad de datos está dedicada a la validación y limpieza de datos a través de reglas de validación a menudo complejas y eliminar datos inválidos. Las soluciones *Big Data* reciben datos redundantes de diferentes fuentes de datos y se pueden explotar para explorar conjuntos de datos interconectados con el fin de ensamblar los parámetros de validación y completar los datos válidos faltantes.

Cuando se habla de velocidad, los procesos de depuración de datos deben ser los adecuados ya que cuando se haya verificado su calidad, estos datos pueden ya haber perdido su valor para el negocio. Los procesos de control de calidad y monitorización en grandes volúmenes de datos se deben trabajar con aproximaciones e intervalos de confianza y no con valores absolutos como se lo hace en un *Data Warehouse* tradicional. Con la gran variedad de datos que vienen de distintas fuentes de datos es imposible poder aplicar una misma métrica de calidad a todos ellos, para ello se debe asegurar que los datos de las fuentes sean los adecuados y esperar que las fases anteriores hagan su trabajo seleccionando los campos útiles y descartar los que aportan ruido.

5.4. Capa de análisis de datos

Esta fase está dedicada a llevar a cabo tareas de análisis de datos, en donde intervienen uno o más análisis. Dependiendo del tipo de resultado que se requiera, esta etapa puede ser tan simple como consultar un conjunto de datos para calcular una agregación para la comparación o tan complejo como aprendizaje en tiempo real combinado con minería de datos y análisis estadísticos para descubrir patrones o generar un modelo matemático para representar relaciones entre variables.

En esta fase de análisis de datos se basa en una serie de técnicas que incluye:

- Detección de patrones de eventos
- El aprendizaje en tiempo real
- Abstracción de eventos
- Modelado de jerarquías de eventos
- Detectar relaciones (tales como causalidad o sincronización) entre eventos
- Abstractar los procesos basados en eventos
- Generación de alertas/activadores para la acción

Las funciones claves del análisis de datos son:

- Descriptivo, modelado predictivo y prescriptivo
- Procesamiento de eventos complejos
- Para calcular la métrica
- Generación de informes.

5.4.1. Modelado descriptivo, predictivo, prescriptivo

La mayoría de datos sin procesamiento, no ofrecen demasiado valor, pero aplicando técnicas y herramientas adecuadas podemos empezar a construir un modelo de cómo se relacionan esos datos y extraer valiosas ideas. El Modelado descriptivo, predictivo, prescriptivo (explicando el pasado, prediciendo el futuro y recomendando la siguiente mejor acción) mediante algoritmos de minería de datos, tales como:

- **Clasificación:** Consiste en identificar un elemento dentro de un grupo de datos, de acuerdo a una clase definida. Se identifica características que hacen que un elemento se vincule a un conjunto de datos.
- **Clustering:** Su objetivo principal es segmentar un conjunto de datos en grupos o clusters que contengan características similares. Por

ejemplo, permite identificar una lista de posibles clientes con perfil de riesgo en función de sus atributos.

- Minería de patrones de asociación: Se basa en conjuntos de datos que contienen identificadores para cada caso y para elementos de casos, busca relación entre los diferentes atributos de un conjunto de datos, por ejemplo, para el rediseño de un sitio de ventas web, se puede utilizar este algoritmo para identificar conjunto de productos que se adquieren juntos, así se puede predecir los productos adicionales en los cuales el cliente puede estar interesado basado en los productos que ya están en la cesta.
- Recommenders / Filtrado de colaboración: Algoritmo utilizado por sistemas de recomendación, permite hacer predicciones automáticas sobre intereses de un usuario o un grupo de usuarios, se basa en que si a un usuario A tiene la misma opinión sobre un usuario B es muy probable que A tenga la misma opinión de B en otro tema diferente que otro usuario escogido al azar.
- Redes neuronales: Es un algoritmo de procesamiento de datos, cuya estructura se basa en las redes neuronales dentro de un cerebro, compuesto por un set de nodos interconectados con un peso asociado y calcula las probabilidades de cada combinación según los datos del aprendizaje. Los tipos de aprendizaje disponibles se utilizan para tareas de predicción y clasificación.

Estos modelos se clasifican según el objetivo del análisis de la siguiente manera:

El modelado descriptivo pretende conocer si existen similitudes entre datos de cada una de las muestras, el propósito del modelado descriptivo es resumir lo que ha

pasado mediante técnicas como *Clustering* o Clasificación. Esta técnica se puede utilizar para estimar el consumo de datos de un cliente en base a información pasada de otros clientes con atributos similares. Los algoritmos más comunes de *clustering* se encuentra K-Means y mapas de Kohonen.

- Algoritmo K-means se basa en el análisis de varianzas entre datos y se usa cuando todas las variables son cuantitativas, empieza por dividir el conjunto inicial en k número de clusters que se pueden elegir de forma aleatoria o con algún conocimiento previo de los datos. Los clusters se forman asociando los objetos más cercanos a un pivote de acuerdo a una medida de proximidad que generalmente es la Euclidiana y permite minimizar la varianza dentro del mismo clúster. Por último, se calcula la media de los valores y se obtiene un nuevo pivote.
- Mapas de Kohonen permite descubrir rasgos comunes, correlaciones o categorías de datos organizados en mapas que contienen nodos o neuronas. Las neuronas deben auto-organizarse en función de los datos externos. Este algoritmo pretende que de un patrón de entrada solo exista una neurona de salida (o un grupo de vecinas). El procedimiento es buscar una neurona en el espacio de los datos en un mapa con la menor distancia, cada vector puede tener 8 vecinos. Cada una de estas neuronas va a tener una influencia sobre su vecina y mientras más cerca se encuentren se le asigna un peso mayor a la conexión que conducirá hacia la salida.

El modelado descriptivo resume el pasado el modelado predictivo permite a los analistas hacer predicciones sobre el futuro, este modelado plantea escenarios que pueden darse en el futuro ya que son probabilísticos basadas en técnicas de inteligencia artificial, que aprenden de datos históricos que ya los conocemos.

Entre las técnicas más comunes de inteligencia artificial tenemos:

- Aprendizaje de Maquina (*Machine Learning*) desarrolla algoritmos que permiten a las computadoras adquirir conocimiento a partir de reconocimiento de patrones complejos a partir de grandes volúmenes de datos, aprender de ellos y predecir comportamientos sin la necesidad de ser explícitamente programados. El aprendizaje de maquina tiene muchas aplicaciones en el sector de telecomunicaciones como control de fraude, por ejemplo, se puede utilizar para controlar el tráfico de datos y navegación gratuita y reventa de servicios de telecomunicaciones.
- Modelo predictivo (*Predictive modeling*) dispone de modelos matemáticos para predecir la mejor probabilidad de un resultado. Este modelo permite predecir la fuga de clientes, es decir predecir de forma anticipada si un cliente va a decidir cambiarse a otra compañía, dada la alta competitividad.

El modelo prescriptivo va más allá de los modelos descriptivos y predictivos, se centra en informar acerca de lo que debería suceder buscando mejorar el resultado esperado, es decir que es capaz de predecir consecuencias basándose en diferentes opciones de acción y puede recomendar el mejor camino para cualquier consecuencia preestablecida, este modelado es mucho menos teórico y bastante más práctico, más convincente, más real. Este modelo combina datos, reglas de negocio y modelos matemáticos, por ejemplo, este modelo puede beneficiar a la planificación del despliegue de la red aprovechando los datos operacionales, el uso de servicios, factores externos como tendencias demográficas, factores climáticos. El modelo prescriptivo requiere de un modelo predictivo con dos componentes adicionales: un sistema de retroalimentación para determinar las consecuencias sugeridas y datos probables.

5.4.2. Procesamiento de eventos complejos (CEP)

El procesamiento de datos completos CEP (*Complex Event Processing*) trata del procesamiento, análisis y correlación de grandes volúmenes de datos, permite solicitar un análisis una vez y a continuación evaluarlo en el tiempo de forma continua contra uno o varios flujos de eventos, por ejemplo: se requiere conocer el tiempo promedio de espera de atención de clientes para determinar producto, categoría del cliente y ciudad, los datos van llegando como una serie de eventos en secuencia continua donde es necesario calcular el tiempo promedio de espera de dichos eventos en tiempo real. (Ver figura 25)



Figura 25. Big Data Analytics Events, Streams, Feeds y CEP

Tomado de: (Network Analyst, 2016).

El CEP puede dividirse en dos categorías:

- La ejecución de algoritmos en línea como respuesta al evento de datos que entran en el sistema. Un ejemplo sencillo es calcular continuamente un promedio basado en los datos sobre los eventos entrantes.

- Detectar combinaciones de eventos llamados patrones de eventos o situaciones. Un ejemplo simple de detectar una situación es buscar una secuencia de eventos específicos.

El CEP es una tecnología clave que permite implementar escenarios para el procesamiento en tiempo real de volúmenes de eventos potencialmente elevados, controla el procesamiento, correlación de estos eventos y el cálculo de los KPI (Key Performance Indicator o Indicador Clave de Desempeño), en una base continua con posibilidad de agregación y composición de eventos de negocio para su visualización. Impulsados por las reglas de negocio proporcionada por el usuario, el CEP genera alertas destinados a seres humanos y desencadenadores para acciones posteriores por sistemas que procesen las notificaciones como paneles KPI, dispositivos de supervisión, bases de datos u otras aplicaciones.

Existen varias herramientas como BusinessEvent, RulePoint, IBM ISII que permiten generar informes de estos datos en tiempo real con frecuencia diaria, semanal o mensual, o bajo demanda para visualizar los grandes volúmenes de datos que muestran los resultados de forma más óptima y eficaz beneficiándose más rápido con estos resultados.

Se puede utilizar el procesamiento de eventos complejos para alarmar cuando un elemento de red no está disponible, enviar notificaciones y enrutar el tráfico por otro nodo, configurando umbrales de criticidad.

5.5. Visualización de datos

La visualización de datos a través de técnicas y herramientas de visualización de datos permite comunicar de forma clara y precisa los resultados del análisis para una interpretación efectiva por parte de los usuarios a través de gráficos

estadísticos, parcelas, información de gráficos, tablas, mapas de calor y la información geográfica.

La capacidad de analizar grandes volúmenes de datos y encontrar información útil tiene poco valor si los únicos que pueden interpretar los resultados son los analistas, una visualización efectiva ayuda a los usuarios a analizar y razonar sobre los datos y resultados, sin embargo, los usuarios también deben ser capaces de comprender los resultados para obtener valor del análisis y, posteriormente, tener la capacidad de proporcionar comentarios.

Los gráficos de líneas muestran la relación de una variable a otra. Se usan con mayor frecuencia para rastrear cambios o tendencias a lo largo del tiempo. Los gráficos de líneas también son útiles cuando se comparan varios elementos durante el mismo período de tiempo. Las líneas de apilamiento se usan para comparar la tendencia o los valores individuales para varias variables (Ver Figura 26).

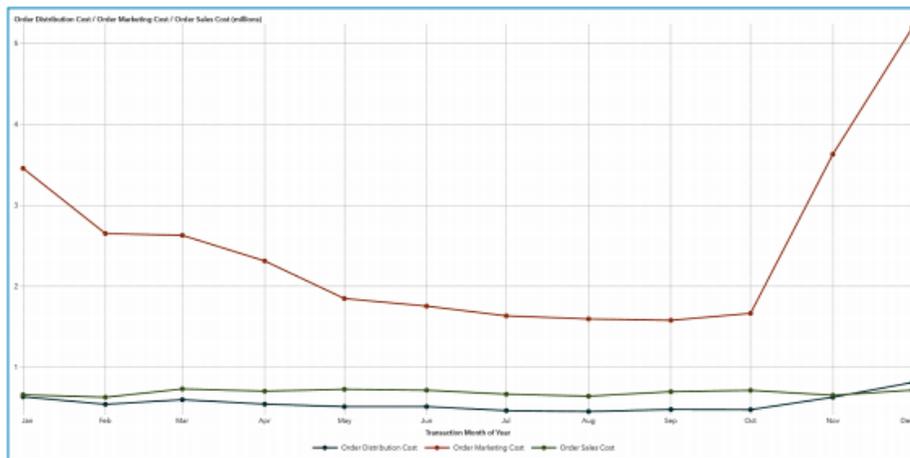


Figura 26. Relación de una variable con otra y tendencias a lo largo del tiempo.

Tomado de: (Network Analyst, 2016).

Los gráficos de barras son los más comúnmente utilizados para comparar las cantidades de diferentes categorías o grupos. Los valores de una categoría se representan utilizando las barras (Ver Figura 27).



Figura 27. Aumento o disminución del valor inicial en una serie.

Tomado de: (Network Analyst, 2016).

Un diagrama de dispersión (o gráfico X-Y) es un gráfico bidimensional que muestra la variación conjunta de dos elementos de datos. En un diagrama de dispersión, cada marcador (símbolos tales como puntos, cuadrados y signos más) representa una observación. La posición del marcador indica el valor de cada observación. Los diagramas de dispersión también admiten la agrupación. Cuando asigna más de dos medidas, se produce una matriz de diagrama de dispersión. Una matriz de diagrama de dispersión es una serie de gráficos de dispersión que muestra todos los emparejamientos posibles de las medidas que se asignan a la visualización. Los diagramas de dispersión son útiles para examinar la relación o las correlaciones entre las variables X e Y. Se dice que las variables están correlacionadas si tienen una dependencia de, o de alguna manera se influyen entre sí. Por ejemplo, "ganancia" a menudo se relaciona con "ingresos". La relación que existe podría ser que a medida que aumentan los ingresos, la ganancia también aumenta (una correlación positiva). Un diagrama de dispersión es una buena forma de visualizar estas relaciones en los datos (Ver Figura 28).

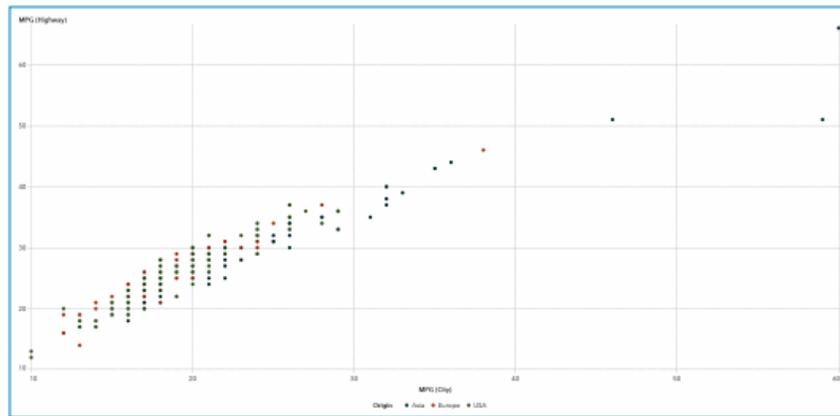


Figura 28. Diagrama de dispersión para visualizar la relación en los datos.

Tomado de: (Network Analyst, 2016).

Diagramas de burbujas: una variación del diagrama de dispersión en el que los marcadores se reemplazan por burbujas. Un diagrama de burbujas muestra las relaciones entre al menos tres medidas. Dos medidas están representadas por los ejes de la gráfica. La tercera medida está representada por el tamaño de las burbujas (Ver Figura 29). Cada burbuja representa una observación.

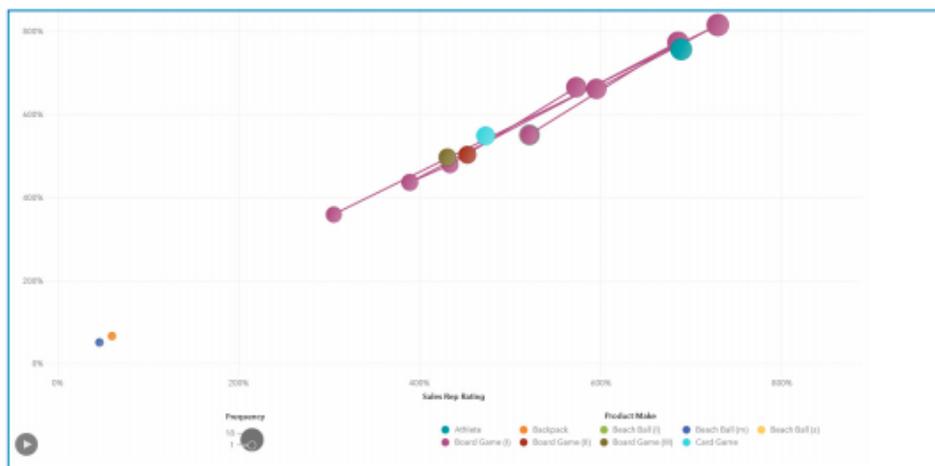


Figura 29. Diagrama de burbujas.

Tomado de: (Network Analyst, 2016).

Los gráficos de pastel y dona se utilizan estos gráficos cuando hay componentes limitados y cuando se incluyen texto y porcentajes para describir el contenido. Al proporcionar información adicional, los consumidores del informe no tienen que adivinar el significado y el valor de cada porción. Si elige usar un gráfico de pastel o dona, las rebanadas / secciones deben ser un porcentaje del total (Ver Figura 30).

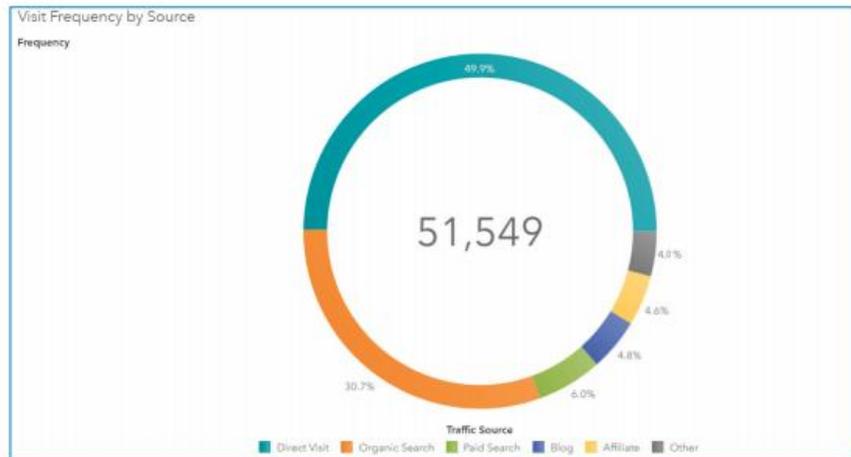


Figura 30. Un gráfico de pastel para comparar los porcentajes.

Tomado de: (Network Analyst, 2016).

Big Data trae nuevos desafíos para la visualización cuando hablamos de velocidad, tamaño y diversidad de los datos, para ello se requiere una nueva manera de presentar los datos que no resulte abrumadora y que también sea posible tenerlos disponibles rápidamente y que los usuarios tengan la posibilidad de explorar los datos en tiempo real.

Por ejemplo, ¿qué sucede si tiene mil millones de filas en un conjunto de datos y desea crear un diagrama de dispersión en dos medidas?, sería imposible ver tantos puntos de datos. Y una aplicación convencional puede no ser capaz de trazar mil millones de puntos de manera oportuna o efectiva. Los diagramas de caja un ejemplo de cómo el volumen de datos puede afectar el aspecto visual que se muestra. Un diagrama de caja es una visualización gráfica de cinco estadísticas (el mínimo, el cuartil inferior, la mediana, el cuartil superior y el máximo) que resumen la distribución de un conjunto de datos. El cuartil inferior (percentil 25) está

representado por el borde inferior de la caja, y el cuartil superior (percentil 75) está representado por el borde superior de la caja. La mediana (percentil 50) está representada por una línea central que divide la caja en secciones. Los valores extremos están representados por bigotes que se extienden desde los bordes de la caja. Usualmente, estos se muestran bien cuando se usan *Big Data* (ver Figura 31).

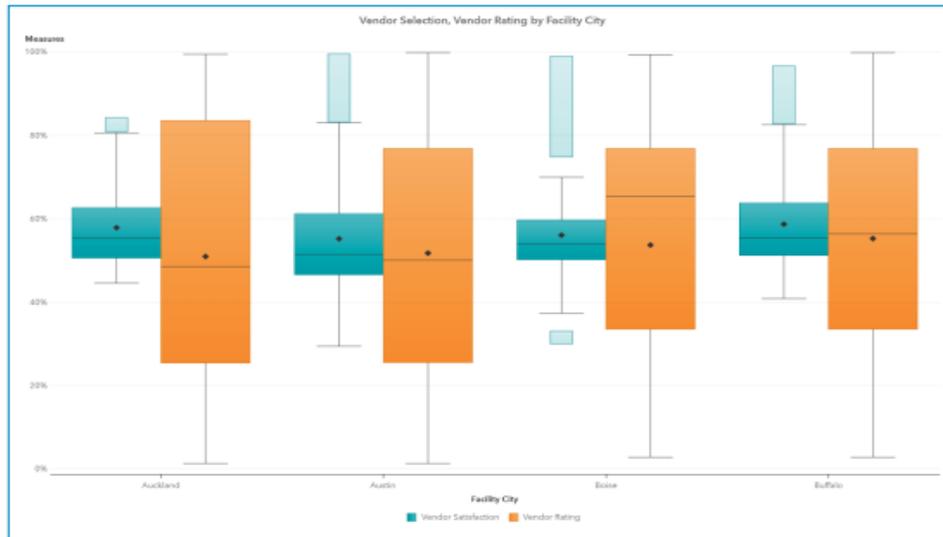


Figura 31. Diagrama de caja que compara la distribución de puntos de datos.

Tomado de: (Network Analyst, 2016).

Los datos semi estructurados y no estructurados traen nuevos desafíos para Big Data ya que este tipo de datos requiere nuevas técnicas de visualización. Una nube de palabras (donde el tamaño de la palabra representa su frecuencia dentro de un cuerpo de texto) puede ser una técnica muy útil para datos no estructurados como una forma de representar la frecuencia alta o baja de palabras como se muestra en la Figura 32.

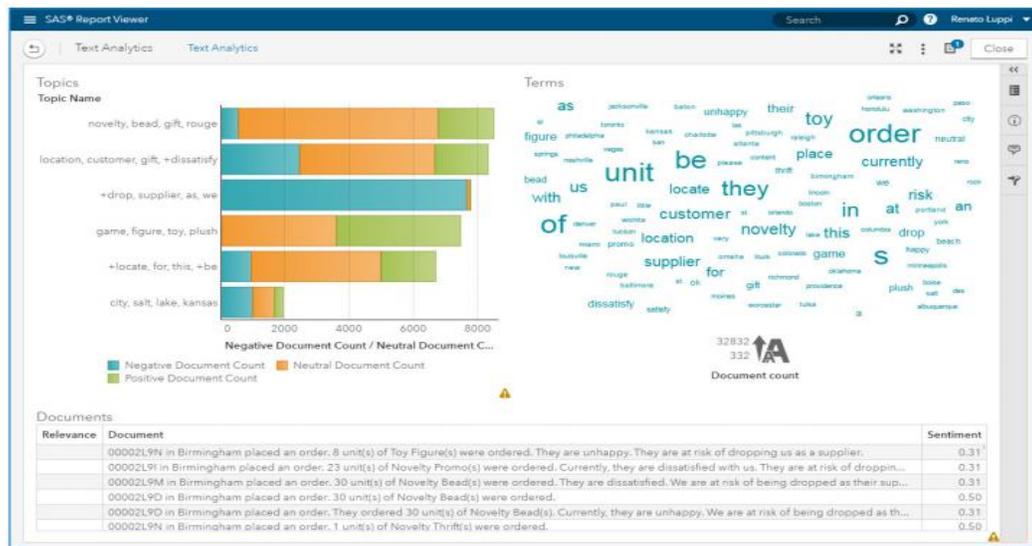


Figura 32. Nube de palabras para representar la frecuencia alta o baja.

Tomado de: (Network Analyst, 2016).

En esta técnica también se puede aprovechar taxonomías y ontologías para hacer asociaciones, quiere decir que se pueden organizar las palabras en temas basados en cómo se usan las palabras. Por ejemplo, podría usar la nube temática para categorizar los comentarios de los clientes en Twitter sobre sus productos o servicios y luego hacer clic en un tema para ver los comentarios reales.

El diagrama de red es otra técnica que permite visualizar datos semiestructurados y no estructurados, en estos diagramas de red se puede visualizar las relaciones en términos de nodos (que representan a los actores individuales dentro de la red) y vínculos (que representan las relaciones entre los individuos, como la amistad, el parentesco, las organizaciones, las relaciones comerciales, etc.). Por ejemplo, las empresas que analizan redes sociales pueden utilizar esta técnica para comprender sus interacciones con los clientes (Ver Figura 33).

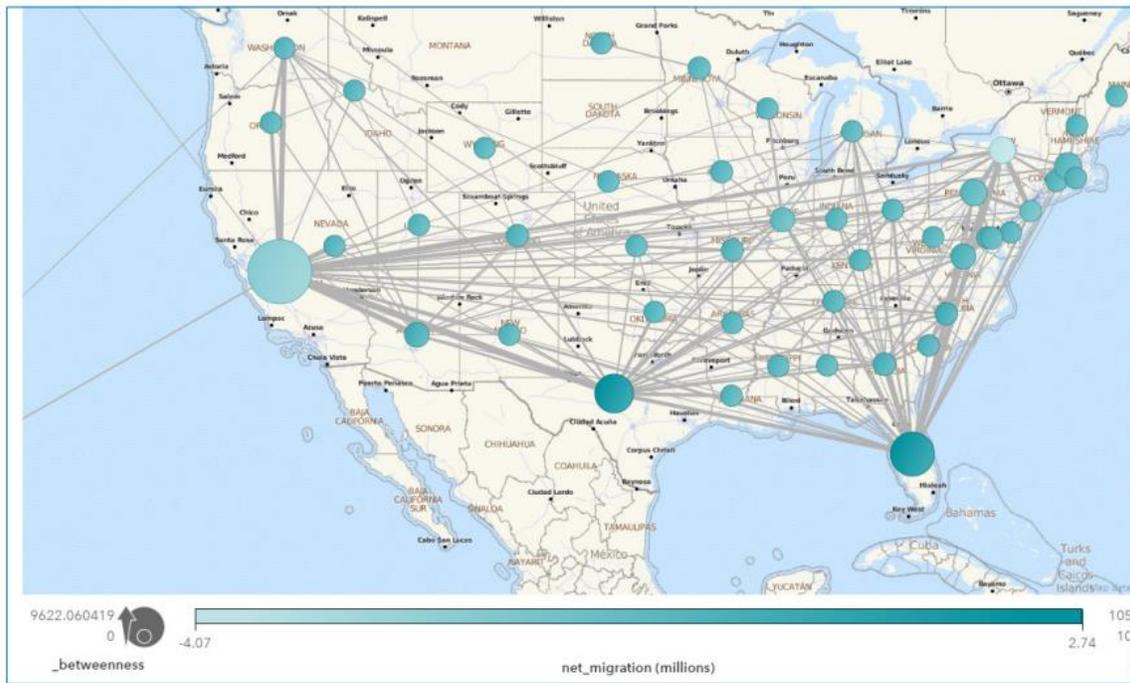


Figura 33. El diagrama de red para visualizar datos semi y no estructurados

Tomado de: (Network Analyst, 2016).

La velocidad de procesamiento en las organizaciones es una capacidad crítica, las matrices de correlación permiten identificar rápidamente todas las variables que están relacionadas y también muestra que tan fuertes son las relaciones entre estas. Los cuadros más oscuros indican una correlación más fuerte; los cuadros más claros indican una correlación más débil (Ver Figura 34).

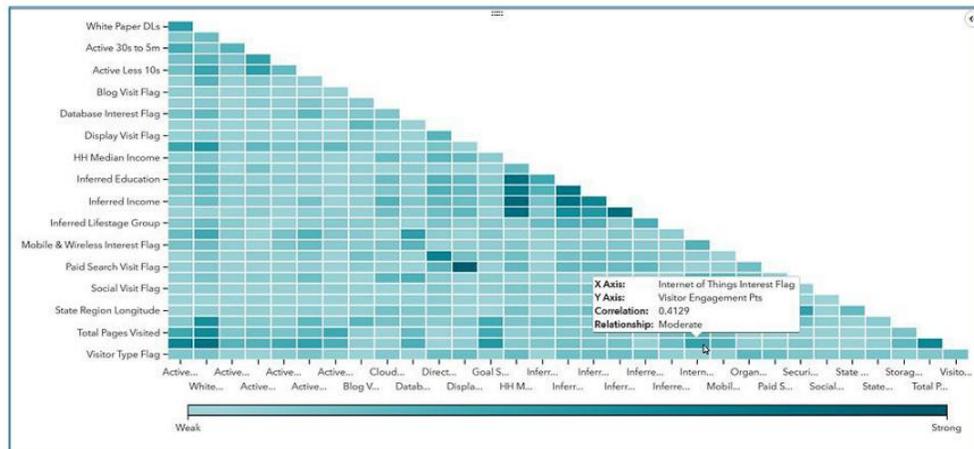


Figura 34. Velocidad de procesamiento en las organizaciones.

Tomado de: (Network Analyst, 2016).

Los gráficos relacionales permiten estudiar la relación entre un cliente, la organización y el nivel de penetración de los servicios (Ver Figura 35).

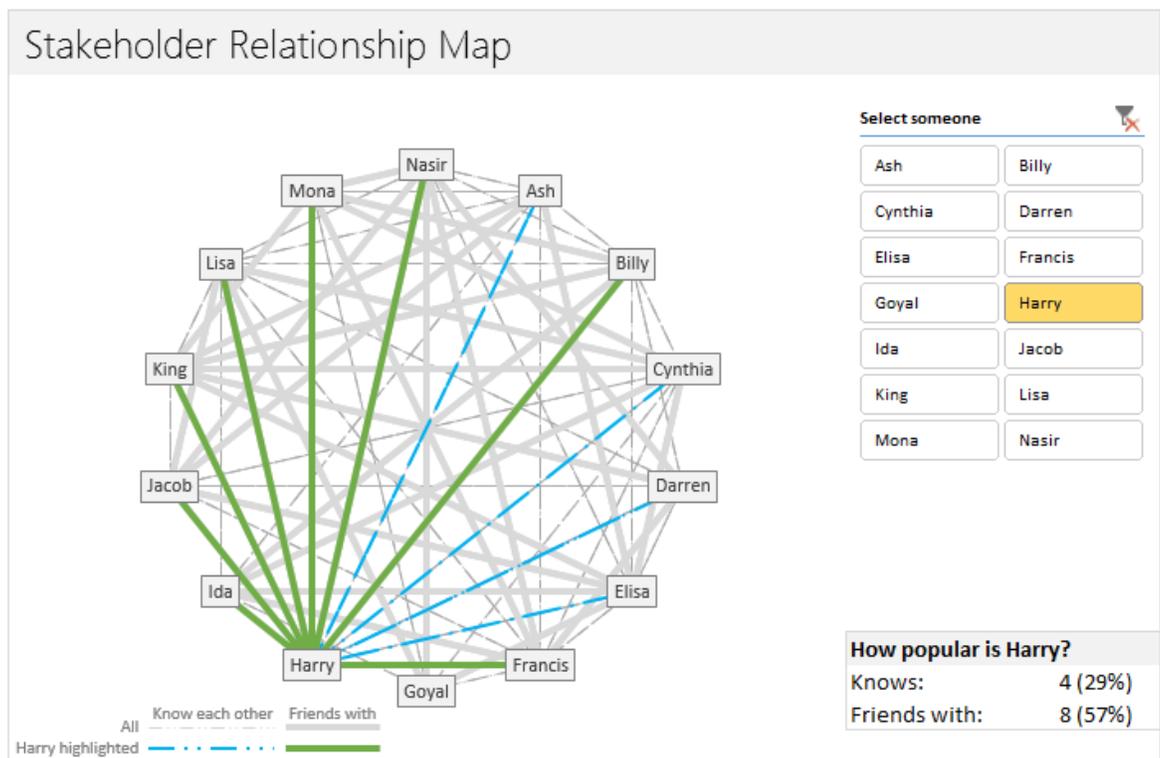


Figura 35. Visualización con gráficos relacionales.

Tomado de: (Network Analyst, 2016).

Los árboles de decisión permiten analizar problemas, incluso predecir comportamientos y representan relaciones causa efecto. Consiste en encontrar una relación sólida entre los valores de entrada y los datos objetivo en una observación, cuando uno de estos valores de entrada tiene una relación fuerte con un valor objetivo se convierte en una rama del árbol y de inmediato el algoritmo indicará que variable tiene influencia y que valor mayor impacto (Ver Figura 36).

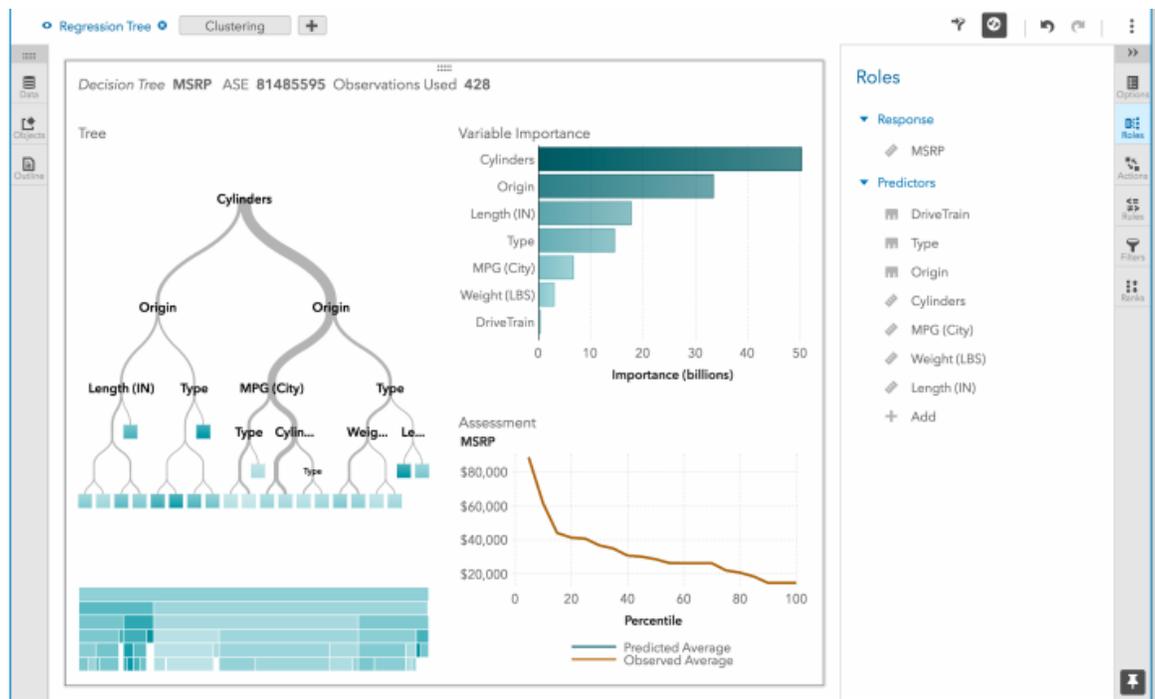


Figura 36. Árbol de decisión, datos segmentados de acuerdo a ramificación.
Tomado de: (Network Analyst, 2016).

El gráfico de matrices es utilizado cuando se tiene varios conjuntos de datos relacionados entre sí a través de vínculos comunes, permite mostrar tanto las posiciones del conjunto como las relaciones entre estos (Ver Figura 37).

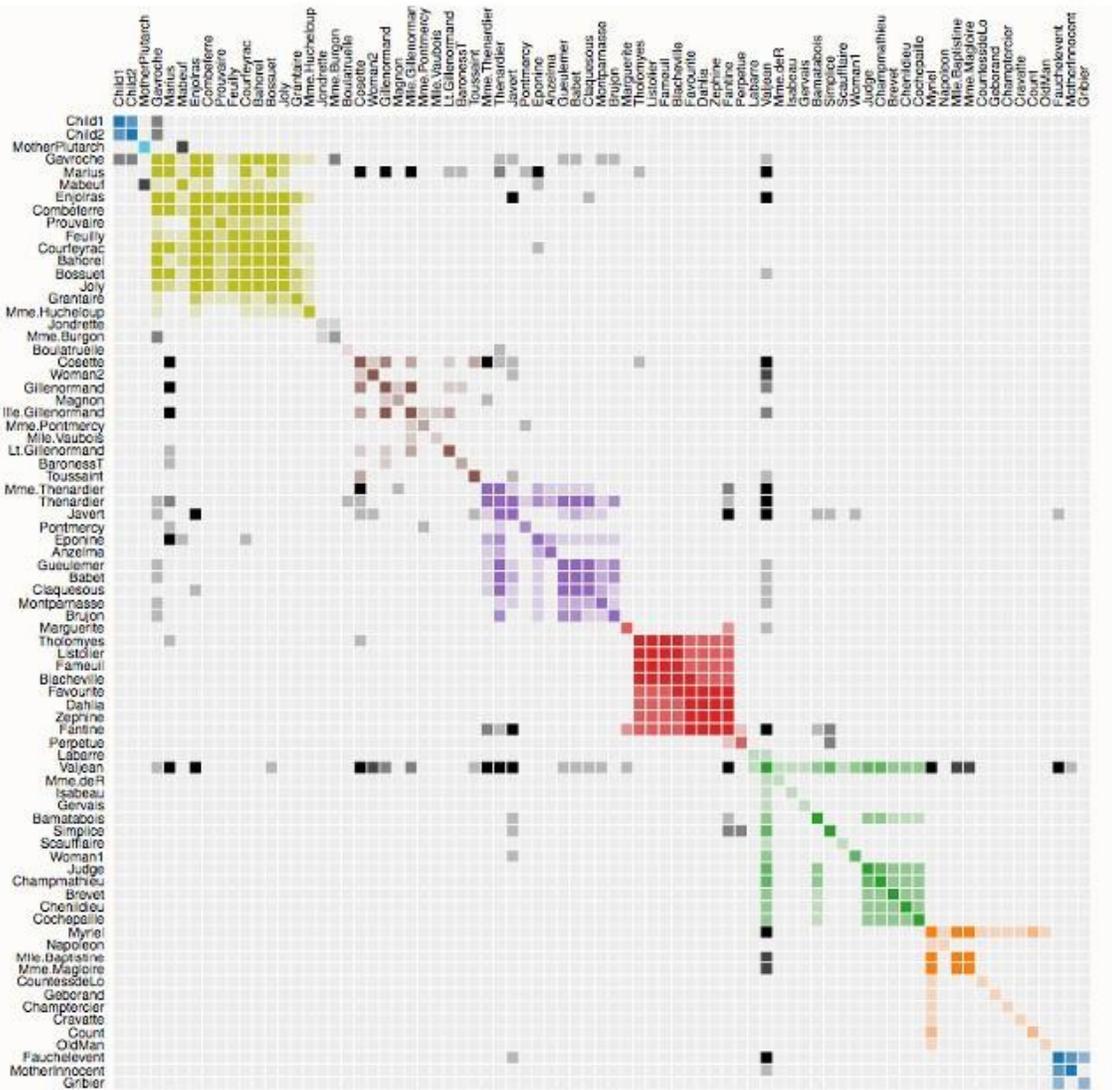


Figura 37. Gráfico de matriz.

Tomado de: (Network Analyst, 2016).

5.6. Repositorio de datos vertical

En esta fase permite el almacenamiento de todos los datos estructurados, semiestructurados y no estructurados en su formato original o procesado por cualquier otra capa del modelo, esta fase también interactúa con todas las demás como un bus.

El repositorio de *Big Data Analytics* (ABDR) es un repositorio de datos CSP, utilizado principalmente con fines analíticos, que permite una reutilización eficiente y directa de los datos para múltiples propósitos, permite manejar datos, por múltiples plataformas de análisis de datos, diversas entidades empresariales, diversos partidos "acoplado", dentro de un mismo grupo empresarial.

Este repositorio de datos proporciona el escalado y la flexibilidad necesaria para manejar el volumen de datos, almacena datos consolidados de diversas fuentes o sistemas de la empresa, estos datos pueden ser locales dentro de un centro de datos interno o acceder a través de una cloud privada o pública.

5.6.1. Datos no estructurados

Los datos no estructurados son datos que no tienen estructura definida. Se almacenan como "documentos" u "objetos" sin estructura, y se tiene poco o ningún control sobre ellos. Los datos no estructurados son típicamente texto, video, audio, fotografía.

5.6.2. Datos semiestructurados

Los datos semiestructurados son estructuras que no se ajustan del todo a la estructura formal de modelos de datos asociados con las bases de datos relacionales o de otras formas de tablas de datos, pero pueden ser tratados como estructurados, tienen un flujo lógico, contiene etiquetas u otros marcadores para separar elementos semánticos y un formato definido pero que aún no es de fácil comprensión para el usuario.

5.6.3. Datos estructurados

Los datos estructurados están organizados en una estructura de acuerdo a un modelo de datos predefinidos, son datos que se encuentran en bases de datos

relacionales y que son fáciles de procesar, y predecible, y que se pueden expresar en filas y columnas con títulos e índices. Este tipo de datos tienen la característica que pueden ser analizados y procesados sencillamente por herramientas de análisis de datos. Estos datos pueden ser internos o externo, donde todo está identificado, etiquetado y tiene fácil acceso.

ABDR (Analytics Big Data Repository) puede ser implementado en diversos tipos de plataformas, ABDR no dicta una determinada tecnología de implementación, las implementaciones actuales de repositorio de datos analíticos generalmente tendrán un componente *Hadoop* y HDFS y en muchos casos se combinarán otras tecnologías, por ejemplo, bases de datos columnares, bases de datos no sql, sistemas de archivos e incluso un *Data Warehouse*. Con respecto a la infraestructura o plataformas, los CSP optarán por implementar un ABDR en función de sus propias limitaciones de los sistemas existentes y los objetivos de rendimiento.

Los grandes repositorios de datos generalmente utilizan bases de datos NoSQL con procesamiento paralelo masivo como *MapReduce* para almacenar rápidamente la gran cantidad de datos. Estos datos pueden ser exportados a un RDBMS para procesar y presentar los resultados almacenados en el repositorio y, con el uso de herramientas gráficas, produce informes, paneles, cuadros de mandos, infografía, potencialmente interactivo.

5.7. Gobernanza de datos

La implementación de iniciativas de *Big Data* puede conducir al descubrimiento de nueva información confidencial o desconocida, esto un reto para las organizaciones al tratar con sus clientes. El *Big Data* representa una gran oportunidad, pero también implica importantes riesgos. Las organizaciones que buscan implementar soluciones en *Big Data* y que no cuentan con un marco de gobernanza se arriesgan a enfrentar muchos problemas ante la ausencia de procedimientos y directrices

concretas que seguir; así como nuevas herramientas y metodologías deberán ser adoptados para hacer frente a esta creciente amenaza. Con el enorme crecimiento del volumen de datos ha provocado un incremento en las medidas de seguridad, por lo tanto, disponer de procesos, entrenamiento, personal y medidas adecuadas es imperativo para gobernar todo lo relacionado con el programa de *Big Data*.

Con el fin de ganarse la confianza del cliente, los proveedores de servicios de comunicaciones deben:

- Adoptar y constantemente se adhieren a las mejores prácticas de la industria y los códigos de conducta; y
- Cumplir con leyes y reglamentos jurisdiccionales.

Este crecimiento en el volumen de datos ha obligado a la industria de telecomunicaciones a considerar un enfoque para grandes volúmenes de datos que abarca la privacidad y la seguridad como partes constitutivas del modelo de buenas prácticas y forma parte de la guía de análisis de grandes volúmenes de datos.

En el mismo ámbito global son las preocupaciones de privacidad, así como un desafío a ser tratado por los CSPs dispuestos a beneficiarse de grandes volúmenes de datos.

La gobernanza de datos es un proceso empresarial, liderazgo por un equipo de gobernanza con los líderes de la organización, este proceso cubre varias áreas dentro de la organización y define las políticas que se deben aplicar para cada tipo de cliente o datos dentro de la red y el cumplimiento a cada una de estas políticas.

El cumplimiento y seguimiento de estas políticas y reglamentos junto con las obligaciones contractuales que tiene que manejar el equipo de gobernanza y los resultados en términos de ejecución de las reglas es visible en la 'governabilidad'

considerado como uno de los componentes de la aplicación del marco de referencia *Big Data Analytics*.

Procesos para la implementación de la gobernanza de datos se han definido como una buena práctica sobre cómo aprovechar la gobernanza de datos para lograr los objetivos empresariales.

Los 6 procesos principales son (Ver Figura 38):

Data Governance (DG) Roadmap Level 1:

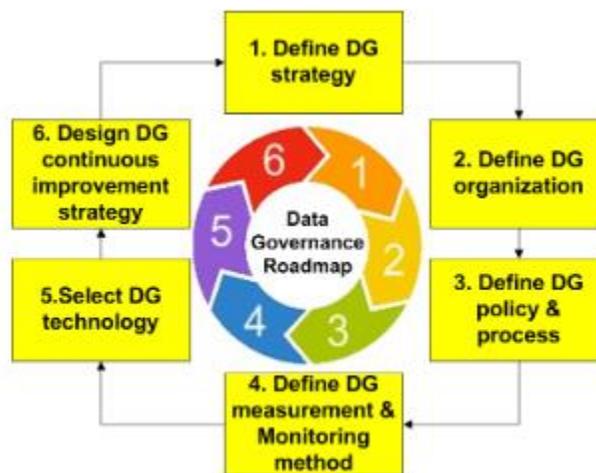


Figura 38. Datos de gobernanza.

Tomado de: (Network Analyst, 2016).

Estos 6 procesos se coordinan mediante el uso de la DG "hoja de ruta", Como una mejor práctica sobre cómo aprovechar la gobernanza de datos para lograr los objetivos de negocio, como se observa a continuación en la tabla 2:

Tabla 2.

Procesos se coordinan mediante el uso de la DG "hoja de ruta".

| Paso | Descripción |
|--------------------------|--|
| 1. Definir estrategia DG | El primer paso es definir la estrategia DG y alinear la estrategia DG con la estrategia comercial. La estrategia DG definida guiará la implementación de los siguientes pasos. |

| | |
|--|--|
| 2. Definir la organización DG | El segundo paso es definir la organización de DG, de acuerdo con la estrategia DG definida. En este paso, diferentes tipos de datos requieren un modelo operativo diferente, y propietarios de datos, roles, responsabilidades. |
| 3. Definir la política y el proceso de DG | El tercer paso es definir la política y el proceso de la DG de acuerdo con la organización definida de la DG. |
| 4. Definir el método de medición y monitoreo de DG | El cuarto paso es definir las métricas DG, el método de monitoreo y el método de medición. |
| 5. Seleccione la tecnología DG | El quinto paso es seleccionar la tecnología DG según el paso anterior define el método de medición y monitoreo. |
| 6. Diseño de la estrategia de mejora continua de la DG | El sexto paso es diseñar un circuito cerrado de DG, y verificar el desempeño de la DG para apoyar los objetivos del negocio. |

Tomado de: (TMForum, 2017).

La DG en una empresa puede ser dividida en una junta de gobernanza y en cada una unidad de negocio, administradores de datos. La junta de gobernanza necesita ser impulsada por la dirección ejecutiva, para convertirse en una organización impulsada por datos enfocada en cambiar la forma de trabajar de las personas y tiene que hacerse de arriba hacia abajo, pero también ganar la aceptación de las personas dentro de la organización.

Funciones del modelo de referencia:

Las capas de gobernanza de datos encapsulan otras capas en la plataforma de *Big Data Analytics* para abordar las mejores prácticas introducidas anteriormente y provee las siguientes funciones:

- Privacidad: gestión, protección y conservación.
- Seguridad: el cifrado, autenticación y control de acceso.
- Cumplimiento: legal y regulatorio.

Estas funciones trabajan a través de 3 dimensiones contextuales:

- Colección - la adquisición de los datos personales del cliente.
- Uso: el almacenamiento, la manipulación y la aplicación a los datos personales del cliente aplicable al negocio del CSP.
- Divulgación - la liberación de los datos personales del cliente o cualquier agregado que puede vincularse a un cliente individual o el usuario final.

5.7.1. Gestión de la privacidad

La gestión de la privacidad se aborda con la necesidad de proteger los datos de los clientes, para ganar la confianza de la sociedad, la única alternativa es la transparencia, elección y control al proporcionar visibilidad y configuración de prácticas y preferencias de privacidad.

Los datos personales pueden ser:

- Compartidos por el titular de manera voluntaria como la información de las redes sociales
- Datos Observados: capturados por una organización a raíz de las acciones de las personas (por ejemplo, datos de localización geográfica de teléfonos móviles).
- Datos inferidos / agregados: obtenidos del análisis de otros datos voluntarios, observados o también inferidos (aquí radica la verdadera potencialidad de los Big Data).

Las técnicas de anonimización permiten identificar y ocultar la información sensible, permitiendo su divulgación sin que ello implique vulnerar los derechos a la protección de datos de las personas y organizaciones que se puedan referenciar en los mismos. Estas técnicas de anonimización se utilizan por ejemplo en soluciones de movilidad urbana que utiliza datos de los clientes y sus móviles, para estimar el movimiento de grupo de personas dentro de las ciudades.

5.7.1.1. Administración Política de privacidad de CSP

La política de privacidad de los proveedores de servicios de comunicaciones (CSP) en el contexto de *Big Data* requiere un enfoque especial para proveer a sus clientes privacidad a través de una administración de las políticas.

La política de privacidad debe contar con:

- Visibilidad de la recopilación, uso y divulgación de datos personales
- Velar por la calidad del dato
- Controlar el acceso de terceros a dato
- Limitación de acceso
- Evaluación de riesgos

Los CSP deben garantizar el cumplimiento de las políticas en sus análisis *Big Data* y ofrecer las garantías.

“Garantizar el cumplimiento de los principios de privacidad significa que las políticas de privacidad de una institución son consistentes con los principios de privacidad establecidos, como tener un organismo externo que establezca un conjunto de reglas, pautas o prohibiciones” (ISO 22307,2008).

5.7.1.2. Protección de privacidad

La Protección de privacidad le otorga y asegura al cliente privacidad de sus datos adherido a:

- Política de privacidad gestionada por CSP y
- Preferencias individuales administradas por el cliente.

La protección se proporciona en la función de seguridad para controlar la recopilación, el almacenamiento, el uso y la divulgación de los datos del Cliente en función de la política anterior.

Un enfoque alternativo es que los CSP brinden a sus clientes un 'panel de administración de privacidad' que permita a los clientes / usuarios finales controlar con precisión cómo se usarán sus datos personales. Sería mejor aún si dicho tablero no estuviera vinculado a ningún CSP o proveedor específico; es decir, funcionaría con todos.

Actuar con responsabilidad permite a los CSP transmitir que el uso de datos para obtener mejores productos y servicios, el cliente continuará cediendo los datos y facilitando la adopción de conocimientos, Eva Plera, directora de Relaciones Institucionales de BBVA indica que “los clientes tienen que percibir confianza con la cesión de sus datos, bajo su consentimiento se adoptará conocimiento y se añadirá valor a los productos y servicios” (BBVA,2017) (Ver Figura 39).



Figura 39. Big Data privacidad la banca ante nuevos desafíos éticos.

Tomado de: (BBVA, 2017).

5.7.1.3. Técnicas de anonimización

La preservación de la privacidad se centra en técnicas de anonimización de datos que son requeridas antes de la publicación de los datos de los clientes, como:

- K-anonimato (*k-anonymity*) es una técnica de anonimización para datos estructurados con campos específicos personales, donde algunos de los valores o atributos se reemplazan por un asterisco "*" o en su caso los atributos son reemplazados por una categoría genérica, por ejemplo, la edad si el atributo es 10 se reemplaza por $\geq 1 \leq 20$.
- Seudonimización es una técnica que reemplaza los datos de identificación dentro de un conjunto de datos por uno o más seudónimos, puede existir un solo seudónimo o varios por cada dato reemplazado. Estos datos se pueden desidentificar y desensibilizar de tal manera que la información personal se mantenga anónima en el contexto de procesamiento, análisis y visualización.
- Removing Personally Identifiable Information (PII) esta técnica funciona como un redactor en documentos y archivos eliminando intencionalmente la información sensible.

Las técnicas de anonimización de datos buscan ocultar la identidad, estos atributos personales aplica para cualquier tipo de persona, estos atributos pueden ser:

- Número de celular
- Número de identificación personal
- Dirección de facturación
- Teléfonos de contacto
- Número de serie del terminal
- Número de serie de Simcard
- Ubicación del terminal
- Historial de navegación

5.7.1.4. Técnicas de privacidad diferencial

Investigaciones recientes, en las áreas de técnicas de privacidad diferencial, ofrecen técnicas de preservación de la privacidad matemáticamente probadas. La privacidad diferencial busca preservar la privacidad (minimizar el riesgo) mientras

se mantiene la utilidad de los datos. La privacidad diferencial no garantiza privacidad perfecta o utilidad perfecta. Si bien la privacidad diferencial no es perfecta, en general es más aceptable que la redacción PII o la seudonimización cuando la reidentificación de un solo individuo es motivo de preocupación.

Conjunto de datos sintéticos para proporcionar datos CSP anónimos a terceros:

Datos abiertos - uso compartido de datos seguro - para proporcionar registros de datos de llamadas (CDR) en formato de conjunto de datos sintéticos para permitir el suministro de datos anónimos.

Existe una técnica llamada 'formato de conjunto de datos sintéticos' para proporcionar CDR (Registros de detalles de llamadas) como datos anónimos a terceros. Esta técnica utiliza técnicas de anonimato y de privacidad diferencial.

Con el fin de desconectar los datos y el usuario real del teléfono móvil, se propuso la creación de un conjunto de datos sintéticos a partir de los CDR originales, es decir, un conjunto ficticio de datos que tienen las mismas características que un conjunto de datos verdaderos, de modo que no sea posible volver a identificar a los usuarios verdaderos iniciales. Tal método todavía está en la etapa experimental, pero los resultados iniciales parecen realmente prometedores.

5.7.2. Cumplimiento

El cumplimiento es un mecanismo de integridad para garantizar el desarrollo y cumplimiento de leyes y regulaciones jurisdiccionales de acuerdo con el contexto, lo que significa principalmente la ubicación, la fecha y la hora en que se recopilan, utilizan y divulgan los datos. El cumplimiento puede ser de dos tipos: Opcional (por ejemplo, leyes o regulaciones) y Requerido (por ejemplo, mejores prácticas, códigos de conducta).

5.7.2.1. Cumplimiento Legal

El Cumplimiento legal se refiere al cumplimiento de las leyes jurídicas del país, estado o federación donde se utilizan los datos del cliente, por ejemplo, la legislación de privacidad en la Unión Europea, Marco de Regulación de Privacidad Global de Datos (GPDR) llamada directiva de privacidad electrónica es un marco legal de varios países y que menciona que los datos de los clientes deben mantenerse dentro del país o deben ser rastreables dentro del mismo.

5.7.2.2. Cumplimiento Normativo

El cumplimiento normativo es controlado por entidades basadas en países, estados, federaciones o múltiples países como, por ejemplo, la Comisión Nacional Francesa de Informática y Libertad (CNIL) es responsable de proteger los derechos humanos y la identidad y privacidad de las personas. Antes de cualquier recopilación de datos, esta entidad debe asignar una autorización al CSP o DSP. La duración de la recopilación de datos está limitada a 1 año o 14 meses como para los Registros de datos de llamadas.

En Ecuador, en el artículo 66 numeral 19 de la norma suprema “garantiza el derecho a la protección de datos de carácter personal, que incluye el acceso y la decisión sobre la información y datos de ese carácter, así como la correspondiente protección” (Asamblea Nacional del Ecuador, 2016). Esta ley tiene como objetivo garantizar el derecho de privacidad de las personas y tratamiento de datos personales en bases de datos, ficheros, archivos físicos o digitales en ámbitos públicos o probados.

5.7.2.3. Mejores prácticas y códigos de conducta

Adicional a las leyes y regulaciones, los proveedores de servicios de telecomunicaciones pueden implementar regulaciones y mejores prácticas recomendadas por líderes de la industria, entidades gubernamentales no

legislativas, tales como, la el Gobierno de Estados Unidos propuso la protección de datos personales a través de la declaración de derechos de privacidad del consumidor, que incluye el análisis y tratamiento de datos personales del consumidor, que generan información privada de sus usuarios y las prácticas de su implementación. La MMA (Mobile Marketing Association) proporciona un código de conducta global que contiene todos los estándares y buenas prácticas de privacidad para los usuarios de dispositivos móviles.

5.7.3. Gestión de la seguridad

5.7.3.1. Cifrado

Las capacidades de cifrado se abarcan para datos en reposo y para datos en tránsito, es un proceso por el cual la información legible se transforma mediante un algoritmo en información ilegible, llamada criptograma o secreto. Esta información ilegible se puede enviar a un destinatario con muchos menos riesgos de ser leída por terceras partes. Estos datos pueden residir dentro del Repositorio de *Big Data* o en ubicaciones de almacenamiento temporales dentro de la Ingestión de Datos y /o Servicios de Intercambio de Datos. Todos los datos en tránsito deben estar cifrados en todos los puntos de acceso. Un sistema de cifrado eficiente requiere al menos un doble juego de claves, de lo contrario no es resistente.

5.7.3.2. Autenticación

Los servicios de autenticación se proveen en todas las capas de Big Data Analytics de extremo a extremo, en un contexto que incluye la capa a la que se accede, la función de la entidad autenticada y el objetivo de la aplicación que accede.

5.7.3.3. Control de acceso

Las administraciones de control de acceso se proporcionan tanto como para el cliente como para los proveedores de servicios de telecomunicaciones, estos últimos definen un control de acceso enfocado en roles y políticas de privacidad y

mejores prácticas de seguridad. El control de acceso proporciona un medio razonable para que un cliente acceda a sus datos personales y a la configuración de Gestión de Privacidad.

6. CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

Big Data es un avance tecnológico que ha cambiado el enfoque de entendimiento y toma de decisiones, ha permitido sacar provecho de datos de diferente tipo y estructura en enormes cantidades, es decir *Big Data* aplica para toda información que no puede ser procesada o analizada por sistemas tradicionales. Cuando hablamos de *Big Data* no solo se refiere a grandes volúmenes de información, existe también otros atributos como variedad y velocidad de datos que pueden provenir de diversas fuentes y en tiempo real, todo esto se logra con la combinación de técnicas de recolección, minería de datos, analítica avanzada, visualización, seguridad, privacidad y gobernanza de datos.

La industria de telecomunicaciones actualmente es una industria privilegiada que está viviendo una auténtica revolución, gracias a masificación de dispositivos inteligentes y creciente incremento del volumen de datos, se ha convertido en una oportunidad clave para un ecosistema tan competitivo donde los jugadores OTT (Over-The-Top), como Whastapp, Facebook, Google, entre otros están devorando las ganancias.

Los operadores de telecomunicaciones móviles actualmente son una mina de oro de información ya que cuentan con grandes bases de clientes, información de llamadas, mensajes y navegación, sin contar con la información que se genera a través de las interacciones con los clientes como canales de venta y call center. Esta información ofrece un conocimiento importante sobre comportamientos, preferencias y movimientos de sus clientes.

Líderes en la industria de telecomunicaciones, consultoras, empresas de desarrollo de software, proveedores de servicios, entre otras han evidenciado a través de casos de éxitos reales los beneficios de Incorporar tecnología *Big Data*. Las empresas que han logrado adoptar este concepto de manera efectiva han revolucionado la forma en que construyen, ejecutan y comercializan sus servicios, han incrementado la lealtad de marca, la experiencia con el cliente realizando predicciones con alto grado de precisión, poniendo de manifiesto la fiabilidad del producto y/o servicio y mejorando los procesos internos.

El nuevo mundo de la experiencia del cliente, tiene aplicabilidad en muchos ámbitos, este nuevo concepto se basa en entregar un emocionante conjunto de nuevos servicios, personalizados y contextualizados para cada cliente en cada punto de interacción, estos servicios no solo necesitan saber qué quiere el cliente y cuándo lo quiere, sino que también necesita saber cómo se lo va a enviar. La experiencia con el cliente debe ser dinámica constante, independientemente de cómo estén accediendo a ella. Por último, integrar todos estos elementos en una solución Big Data se convierte en la única ventaja competitiva ya que los competidores pueden replicar casi todo lo demás que se realiza (Ver Anexo 2).

En tiempo real es un término que sobresale sobre los demás en el sector de las telecomunicaciones que, a diferencia de otras industrias, el tiempo real es relevante a medida que los servicios basados en localización y la prevención de fraude requieren un enfoque en tiempo real. Este enfoque apoya a mejorar la experiencia del cliente, desarrollar nuevos productos y servicios, sin embargo, abre nuevos negocios con nuevas fuentes de ingresos añadiendo socios como retailers, agencias de publicidad (Ver Anexo 2).

Una estrategia de *Big Data* debe ser encontrar la mejor manera de aprovechar los datos para resultados de negocio más predecibles en un proceso de descubrimiento y se debe tener en cuenta qué datos se dispone, dónde, a quién pertenece y lo controla, y cómo se usa actualmente. Este proceso de descubrimiento proporcionará conocimiento sobre todas las fuentes ya sean internas o externas y dependencia de terceros.

El aumento exponencial en el volumen de datos ha generado grandes cambios no sólo la manera en cómo las empresas se relacionan con sus clientes, sino que también en la forma en que las empresas desarrollan sus modelos de negocios. El *Big Data* ha impactado en cada uno de los procesos de las organizaciones convirtiéndose en un recurso vital en la creación de ventajas competitivas sostenibles. Ahora bien, dado este nuevo contexto, impulsado por la innovación *Big Data* permite generar varias oportunidades para desarrollar nuevos modelos de negocio y maximizar el rendimiento de la empresa. Los operadores de telecomunicaciones móviles requieren desarrollar capacidades para el uso de estos nuevos modelos de negocio basados en *Big Data*, esto requiere no solo una infraestructura de TI avanzada, sino también una cultura organizacional y de gobierno que apoyen diferentes iniciativas. Este nuevo modelo de negocios basados en el cliente requiere mayores responsabilidades para las organizaciones que hacen uso de los datos de sus clientes, estas propuestas tienen su atención en el uso de los datos más que en la recolección. Todos estos cambios para explotar los beneficios de *Big Data* no se requieren de forma inmediata, las capacidades se pueden ir desarrollando de forma incremental involucrando personal clave en los procesos que apoyan su efectividad, inteligencia analítica y una estrategia basada en *Big Data*. En este contexto la industria de telecomunicaciones se convierte en un ecosistema atractivo para el surgimiento de nuevos negocios así, como mayor investigación en nuevos modelos de negocio.

El marco de referencia Big Data Analytics proporciona no solo una arquitectura estándar para *Big Data*, si no también introduce un conjunto de casos de uso a nivel empresarial y de negocio como la monetización de datos, publicidad móvil, exposición de datos del cliente que pueden ser adaptados de acuerdo a las necesidades de cada organización permitiendo obtener el máximo valor de los datos.

Finalmente, el marco de referencia Big Data Analytics proporciona un conjunto de directrices para la adopción de Big Data en operadores de telecomunicaciones móviles en torno a un problema que se desea resolver proporciona un lenguaje común para acelerar, simplificar y eliminar los riesgos de los proyectos de análisis de datos. Cada uno de los elementos del marco de referencia:

- Evaluación del caso comercial
- Ingestión de datos
- Procesamiento de datos
- Análisis de datos
- Visualización de datos
- Repositorio de datos vertical
- Gobernanza

Incorpora las mejores prácticas y estándares para los procesos relacionados con empresas de telecomunicaciones como es TMForum de tal manera que permite incursionar en nuevos modelos de negocio, nuevos productos, servicios y así mejorar y recuperar competitividad aprovechando la interacción con los clientes.

6.2. RECOMENDACIONES

Antes de iniciar con un proyecto de *Big Data* y tener una comprensión inicial de lo que puede ser posible se debe establecer objetivos a corto y largo plazo con una colaboración entre las unidades de TI y de negocios para llegar a objetivos bien definidos. ¿Qué se espera lograr con *Big Data*? ¿Podrían algunas partes del negocio ser más rentables con la infusión de más datos para predecir el comportamiento del cliente o patrones de compra?

Después de comprender los objetivos de cómo aprovechar *Big Data* se necesita involucrar a todas las partes interesadas del negocio. Conformar un equipo de trabajo es una excelente forma de reunir a los representantes del negocio para que puedan ver cómo se relacionan sus problemas de administración de datos. Este equipo puede convertirse en un equipo que puede ayudar a varias unidades de negocios con las mejores prácticas. El equipo de trabajo debe tener representantes de los líderes de la alta gerencia que están estableciendo la estrategia y la dirección del negocio.

Dedicar tiempo y los esfuerzos necesarios para realizar el proceso de descubrimiento, que será la base para la planificación y ejecución de la estrategia de *Big Data*.

La complejidad que aborda *Big Data* es que es costoso de administrar y difícil de extraer valor de él, incorporar tecnología *Big Data* en operadores de telecomunicaciones representa un desafío y radica en la capacidad del operador para recopilar y analizar los datos de manera rentable, esto es lo que hace realmente poderoso a *Big Data*.

El marco de referencia propuesto proporciona una visión amplia, un esquema de elementos relacionados que apoya un enfoque particular de un determinado objetivo, sirve como guía y se recomienda que sea modificado de acuerdo a las necesidades de cada organización.

Si bien las empresas siempre mencionan la seguridad de los datos como uno de los problemas más importantes que deben gestionar, se recomienda que, en las primeras etapas del análisis de *Big Data*, el analista no protegerá los datos, ya que solo una pequeña parte de esos datos se guardará para un análisis posterior. Sin embargo, cuando un analista selecciona una cantidad de datos que se incorporarán a la empresa, los datos deben estar protegidos contra riesgos internos y externos. Algunos de estos datos tendrán información privada que debe enmascarse para que nadie sin autorización tenga acceso. Para que la seguridad sea efectiva en el contexto del *Big Data*, necesita tener políticas robustas y asegurar el cumplimiento de las mismas.

Una estrategia de gobernanza es una responsabilidad conjunta de TI y el negocio, proporciona reglas concretas que dicta cómo se regirá Big Data. Existen reglas que definen cómo deben protegerse los datos según las circunstancias y los requisitos gubernamentales, Define cómo los datos personales deben permanecer privados, aboga para que los datos sean confiables y otorga responsabilidad a los equipos de trabajo. Una estrategia de gobernanza implica un cambio de mentalidad en una organización donde los grandes volúmenes de datos que se maneja es cada vez mayor y, por tanto, es mucho más complicado trabajarla y estructurarla de manera idónea. Una estrategia de gobernanza permite empoderar a los datos y, sobretodo, a los equipos de trabajo.

Contar con los proveedores adecuados que cuenten con la experiencia necesaria que permita fortalecer las capacidades y brindar apoyo para seleccionar las herramientas y los procesos tecnológicos apropiados para alcanzar los objetivos, es vital. Amdocs, Cloudera, IBM, NOUS, Nokia, Ericsson se han convertido en los

mejores aliados de operadores de telecomunicaciones móviles como Vodafone, Telefonica, AT&T, China Telecom, entre otros.

Todos los modelos de negocio en los CSPs para ser digitales deben modernizarse, para esto se recomienda no solo la adopción Big Data, si no también se debe realizar analítica avanzada ya que añade un valor real a los datos, debe existir sinergia y convergencia de sistemas y un cambio de mentalidad que va mas allá de la satisfacción del cliente.

Se recomienda que los casos de uso sean definidos en un lenguaje de negocio y, sobre todo, explicar el valor que aportan para que sean comprendidos por los ejecutivos de alto nivel. Suele suceder que los casos de uso se definen en términos demasiado técnicos y comunican que hacen, pero no el valor.

En los próximos años el mercado de la analítica Big Data para el sector de telecomunicaciones habrá crecido de forma exponencial y continuará fomentando la capacidad de innovar, por ello se recomienda establecer un proceso de mejoramiento continuo sobre la solución Big Data para enfrentar nuevos retos dentro de la industria.

REFERENCIAS

- Accenture. (2017). Big Data puede ser un gran potencial para las empresas de telecomunicación. Recuperado el 3 de marzo de 2017 de <https://www.accenture.com/us-en/insight-highlights-cmt-digital-analytics-big-data-telcos>.
- Accenture. (2017). Iniciativa de transformación digital para la industria de telecomunicaciones. Recuperado el 28 de febrero de 2017 de https://www.accenture.com/t20170116T084457__w__/us-en/_acnmedia/Accenture/Conversion-Assets/WEF/PDF/Accenture-Telecommunications-Industry.pdf.
- Amazon.com, I. Amazon Web Services. <https://aws.amazon.com>, 2006.
- Amin, Feizi (2014). Big Data Strategy for Telco: Network Transformation. Recuperado el 6 de noviembre de 2017 de <https://waset.org/publications/9997406/big-data-strategy-for-telco-network-transformation>.
- Apache, F. (2017). Apache Hadoop. Recuperado el 12 de diciembre de 2017 de <http://hadoop.apache.org/hadoop/>.
- Apache, F. (2017). HDFS Architecture. Recuperado el 13 de diciembre de 2017 de <http://hadoop.apache.org/docs/r2.6.0/>.
- Asamblea Nacional del Ecuador. (2016). Ley orgánica de los derechos a la intimidad y privacidad de datos personales. Recuperado 11 de noviembre de 2017 de <http://www.asambleanacional.gob.ec/sites/default/files/private/asambleanacional/filesasambleanacionalnameuid-29/Leyes%202013-2017/250%20protec-intimidad-grivadeneira-12-07-2016/PP-protec-intimidad-grivadeneira-12-07-2016.pdf>

- BBVA. (2017). ¿Qué son las smart cities? Recuperado 4 de abril de 2018 de <https://www.bbva.com/es/las-smart-cities/>.
- BVEX. (2015). Big Data abre la puerta a monetizar los datos de los clientes. Recuperado el 16 de diciembre de 2017 de <https://businessvalueexchange.com/es/2015/11/12/big-data-abre-la-puerta-a-monetizar-los-datos-de-los-clientes/>.
- Cardosa, M., Wang, C., Nangia, A., Chandra, A., and Weissman, J. Exploring. (2011). Exploring mapreduce efficiency with highly distributed data. Recuperado el 4 de noviembre de 2017 de <https://dl.acm.org/citation.cfm?id=1996100>.
- Deloitte. (2015). Nuevas oportunidades del sector de las telecomunicaciones. Recuperado el 5 de mayo de 2017 de <https://www2.deloitte.com/content/dam/Deloitte/in/Documents/technology-media-telecommunications/in-tmt-opportunities-in-telecom-sector-noexp.pdf>.
- Demchenko Y., Ngo C., Membrey P. (2013). Big Data Ecosystem. White Paper. Recuperado el 2 de enero de 2018 de <https://ieeexplore.ieee.org/abstract/document/6867550/>
- Forbes. (2013). ¿Qué es Big Data? Recuperado el 24 de abril de 2017 de <https://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/#180e722642f6>.
- Forbes. (2015). Big Data As a Service. Recuperado el 10 de noviembre de 2017 de <https://www.forbes.com/sites/bernardmarr/2015/04/27/big-data-as-a-service-is-next-big-thing/#61b4dd6333d5>
- Fundación Apache. (2013). Apache Chukwa. Recuperado el 15 de julio de 2017 de <https://chukwa.apache.org/docs/r0.8.0/design.html>
- Fundación Apache. (2013). Apache Cwiki. Recuperado el 15 de julio de 2017 de <https://cwiki.apache.org/confluence/display/Hive/Design>

- Fundación Apache. (2013). Apache Iwith. Recuperado el 20 de julio de 2017 de http://www.iwith.org/pdf/Libro_BI_Compertir_con_Información.pdf.
- Fundación Apache. (2013). Apache Zookeeper. Recuperado el 15 de julio de 2017 de <https://zookeeper.apache.org/>.
- Gartner. (2014). Gartner Says Beware of the Data Lake Fallacy. Recuperado el 2 de Agosto de 2017 de http://www.gartner.com/newsroom/id/2809117?utm_content=buffer5fc11&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer.
- GSMA. (2016) The-Internet-Value-Chain. Recuperado el 23 de julio de 2017 de http://www.gsma.com/publicpolicy/wp-content/uploads/2016/05/GSMA_The-internet-Value-Chain_WEB.pdf.
- Gualtieri, M., & Yuhanna, N. (2014). The forrestter wave: Big data Hadoop solutions. Recuperado el 6 de noviembre de 2017 de hadoop-project-dist/hadoop-hdfs/HdfsDesign.html.
- Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3.a ed.). Elsevier. Recuperado el 3 de octubre de 2017 de http://liacs.leidenuniv.nl/~bakkerem2/dbdm2007/05_dbdm2007_Data%20Mining.pdf.
- Huawei. (2013). Big Data & Analítica Avanzada en telecomunicaciones. Recuperado el 25 de marzo 2017, Recuperado el 6 de junio de 2017 de www.huawei.com/ilink/en/download/HW_323807.
- IBM (2012). Análisis de Big Data con Apache Pig. Recuperado el 11 de julio de 2017 de <https://www.ibm.com/developerworks/ssa/data/library/bigdata-apachepig/bigdata-apachepig-pdf.pdf>
- IBM. (2013). Exploiting Big Data in telecommunications to increase revenue, reduce customer churn and operating costs. Recuperado el 1 de marzo de 2017 de <http://www-01.ibm.com/software/data/bigdata/industry-telco.html>.

- IBM. (2013). Introducción a la clasificación de Big Data y arquitectura. Recuperado el 10 de junio de 2017 de <https://www.ibm.com/developerworks/library/bd-archpatterns1/>.
- Juárez, Barrios, I. (2013). Predicción del precio de la energía eléctrica utilizando modelos de minería de datos: arboles de clasificación y regresión, random forest y bagging. (2.ª ed.). España, Madrid: Universidad Politecnica de Madrid.
- Krazytech. (2017). Introduction Apache Hadoop. Recuperado el 10 de julio de 2017 de <https://krazytech.com/technical-papers/apache-hadoop-introduction>.
- Weiss, M.A. (2013). Estructuras de Datos en Java. (4.ª ed.). Nueva York, Estados Unidos: PEARSON
- Marz N., y Warren J. (2012). Big Data Principles and best practices of scalable real data systems. (2.ª ed.). Nueva York, Estados Unidos: Manning Publications
- McCarthy C., y Bali S. (2013). Big Data analytics and the telco: How telcos can monetize customer data. Ovum. White Paper. Recuperado el 4 de octubre de 2017 de <http://bigdata-madesimple.com/ovum-warns-telcos-about-missing-out-on-big-data/>.
- McKinsey & Company. (2017) Overwhelming OTT: Telcos' growth strategy in a digital world. Recuperado el 4 de agosto de 2017 de <http://www.mckinsey.com/industries/telecommunications/our-insights/overwhelming-ott-telcos-growth-strategy-in-a-digital-world>.
- Microsoft. (2017). ¿Qué es? IaaS, Infraestructura como servicio. Recuperado el 1 de noviembre de 2017 de <https://azure.microsoft.com/es-es/overview/what-is-iaas/>
- Mike Barlow, O'Reilly. (2013). Real-Time Big Data Analytics: Emerging Architecture. Recuperado el 4 de noviembre de 2017 de <http://pubs.sciepub.com/dt/1/1/7/>

- Network Anayst & SAS. (2017). Tecnicas de visualización de datos. Recuperado el 15 de enero de 2018 de https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/data-visualization-techniques-106006.pdf.
- Olmedo, J. (2011). Comunicaciones móviles. Recuperado el 20 de Agosto de 2017 de <http://comunicaciónesmviles.blogspot.com/2010/05/handoff-yo-handover.html>.
- Oracle. (2016). An Enterprise Architect's Guide to Big Data. Recuperado el 3 de julio de 2017 de <http://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf>.
- Rivadera, G. (2010). La metodología de Kimball para el diseño de almacenes de datos (Data warehouses). Recuperado el 3 de septiembre de 2017 de <http://www1.ucasal.edu.ar/hm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>.
- Rodriguez, Pablo. (s.f.). Director of research and innovation and head of the Barcelona R&D lab at Telefonica Digital. Recuperado el 9 de noviembre de 2017 de www.rodriuezrodriuez.com.
- Ronald Bouman, J. v. (2009). Pentaho Solutions: Business Intelligence and DataWarehousing with Pentaho and MySQL. (2.ª ed.) Indianapolis, Estados Unidos: Wiley Publishing.
- Salesforce. (2017). Salesforce IoT Cloud. Recuperado el 15 de diciembre de 2017 de <https://www.salesforce.com/mx/products/salesforce-iot/overview/>
- T. Matias, F. Souza, R. Araujo, and C.H. Antunes. (2014). "Learning of a single-hidden layer feedforward neutral network using an optimized extreme learning machine". (4.ª ed.). Detroit, Estados Unidos: Neurocomputing.
- Telefonica (2017). Economía Digital. Recuperado el 5 de julio de 2017 de <http://www.expansion.com/economia-digital/protagonistas/2017/06/22/59411b13e5fdea985f8b45af.html>

- Terán, D. (2014). *Administración Estratégica De La Función Informática*. (1. ed.). Madrid, España: ALFAOMEGA
- TMForum. (2012). *Insights Research: Big Data: Big volumes, big payback and big challenge*. Recuperado el 15 de diciembre de 2017 de <https://www.tmforum.org/>.
- TMForum. (2018). *Big Data Analytics*. Recuperado el 1 de febrero de 2018 de <https://www.tmforum.org/resources/toolkit/analytics-toolkit/>.
- Villada, F. Muñoz, N. & García, E. (2014). *Application of artificial neural networks to price forecasting in the stock Exchange market*” *Neurocomputing*. (1.ª ed.). Medellín, Colombia: Universidad de Antioquia.
- Vitria Corporate. (2012). *CEP - Procesamiento de Eventos Complejos*. Recuperado el 1 de febrero de 2018 de <http://blog.vitria.com/es/bid/81509/Parte-II-CEP-Procesamiento-de-Eventos-Complejos>.
- Williams, S. Williams, N. (2016). *The Profit Impact of Business Intelligence*, Morgan Kaufmann. . (1.ª ed.). Nueva York, Estados Unidos: Elsevier.
- Yuri Demchenko, UNIVERSITY OF AMSTERDAM. (2013). *Defining the Big Data Architecture Framework* Amsterdam. Recuperado el 6 de marzo de 2017 de http://bigdatawg.nist.gov/_uploadfiles/M0055_v1_7606723276.pdf

ANEXOS

Anexo 1. Glosario

AOL (América Online): Empresa de servicios publicitarios.

AMPS (Advanced Mobile Phone System): Sistema de telefonía móvil de primera generación que se implementó en Estados Unidos en los años 80.

ARPU (Average Revenue Per User, ingresos medios por usuario): Término que representa el ingreso promedio por usuario que mide el consumo de clientes por servicio.

BDL (Business Data Lake, Lago de datos de negocio): es un repositorio de datos que almacena una gran cantidad de datos en su formato original que utiliza una arquitectura plana para almacenar datos sin ser tabulada o clasificada.

B2B (Business-to-business): Modelo de negocio en la cual la prestación de servicios o comercialización de productos se hacen entre dos empresas.

B2C (Business-to-Consumer): Modelo de negocio de comercio directo entre una empresa y cliente o consumidor.

CAGR (Compound annual growth rate): Tasa anual de crecimiento, indicador que se utiliza para medir el retorno de una inversión dentro de un período dado.

CDR (Call Detail Record): Registro que contiene información de una llamada.

CEO (Chief Executive Officer): Concepto que representa a la persona que tiene el cargo de mayor responsabilidad directiva en una empresa.

CSP (Communications Services Providers, proveedor de servicios de comunicaciones): Son empresas que proveen servicios de comunicaciones y que poseen su propia infraestructura.

CRM (Client Relationship Management, Gestión sobre la Relación con los Consumidores): se define como una estrategia orientada a la administración del cliente y permite la administración de la relación que se tiene con el cliente.

CDMA (Code Division Multiple Access): Tecnología inalámbrica propiedad del fabricante de chips Qualcomm para teléfonos móviles.

Churn: Es el porcentaje de clientes que dejan de utilizar los servicios que ofrece una empresa.

DM (Data mart): Son subconjunto de datos que permiten la toma de decisiones y se considera como una versión especial de un almacén de datos.

DPI (Deep Packet inspection): Elemento de red que permite inspeccionar paquetes de datos y filtrar tipo de tráfico de datos de aplicaciones, como por ejemplo YouTube, Skype, Whastapp, Facebook y cualquier tipo de aplicación.

DWH (Data Warehouse, Almacén de datos) Repositorio donde se unifica toda la información de la empresa que se recogen de diversos sistemas.

ECM (Enterprise Content Management): Conjunto de sistemas que almacenan información de los procesos dentro de una organización.

ERP (Enterprise Resource Planning, Planificación de recursos empresariales): es un conjunto de sistemas de información que integra de ciertas operaciones de una empresa, como recursos humanos, finanzas, logística, inventarios, entre otros.

ETL (Extract, Transform and Load): Proceso que permite extraer, transformar y cargar datos en una base de datos.

e-Commerce (Comercio Electrónico): Son negocios por internet o negocios en línea que permite la compra y venta de productos o servicios a través de medios electrónico.

e-Advertising (Publicidad en línea): Contenido publicitario de las empresas que encontramos y se publicidad en internet

FCC (Federal Communications Commission): Agencia reguladora de telecomunicaciones de Estados Unidos.

FDMA (Frequency Division Multiple Access): Técnica de multiplexación utilizado en telecomunicaciones divide el espectro en canales para comunicación según las necesidades de la red.

Gbps (Gigabit por segundo): Unidad de medida de velocidad de transmisión 1Gb/s equivale a 1000000000 b/s.

GSM (Global System for Mobile communications, Sistema Global para Comunicaciones Móviles): Es una tecnología inalámbrica de segunda generación que actualmente es el sistema básico para todas las comunicaciones móviles. Al ser un sistema digital no se requiere un terminal fijo para acceder a cualquier información existente en la red y permite su uso en cualquier lugar con cobertura, incluso en ámbitos internacionales.

HDFS (Hadoop Distributed File System, Sistema distribuido de archivos Hadoop): es un sistema de archivos distribuido, escalable y portátil escrito en Java para el framework Hadoop.

HTML (HyperText Markup Language): Lenguaje para elaboración de páginas web.

IoT (Internet of Things): Sistemas de dispositivos, sensores, terminales móviles interconectados con internet que intercambian información.

JSON (JavaScript Object Notation - Notación de Objetos de JavaScript): Formato ligero de intercambio de información.

KPI (Key Performance Indicator): Medidor de desempeño, medida del nivel de desempeño de un proceso.

LTE (Long Term Evolution): Estándar de comunicaciones móvil de alta velocidad (173 Mbps de bajada y 86 Mbps de subida), baja latencia y fácil de desplegar.

LUCA (Last Universal Common Ancestor): Unidad de servicios Big Data parte del grupo Telefónica para clientes corporativos, se comercializará desde las unidades de empresa de los países donde opera Telefónica.

MHz (Megahercio): Unidad de medida de frecuencia.

M2M (Machine to Machine): Comunicación entre máquinas a través de internet.

Mbps (megabit por segundo): Término que cuantifica la velocidad de datos.

MMS (Multimedia Messaging Service): Estándar de mensajería que permite enviar contenidos multimedia, como fotos y videos.

NMT (Nordic Mobile Telephones): Sistema de telefonía móvil definido por el estado escandinavo.

NoSQL: Soluciones para almacenar datos que no cumplen con un esquema de entidad relación, no requieren estructura de datos en tablas, no utilizan SQL como lenguaje principal de consulta y permiten almacenar grandes cantidades de información.

OTT (Over-The-Top, Capa superior): Los servicios OTT son servicios de valor agregado los que se brindan y se comercializan a través de internet.

Proveedor de contenidos OTT: Son empresas que prestan servicios de comunicación, contenido multimedia que utilizan la red de datos de un tercero para comercializar sus servicios.

RFID (Radio Frequency Identification): Método de recuperación y almacenamiento de datos que permite transmitir la identificación de un elemento a distancia, mediante tarjetas o transportadores RFID.

SIM (Subscriber identity module): Tarjeta inteligente utilizada en terminales móviles que almacenan la información del suscriptor.

SIP (Session Initiation Protocol): Es un protocolo de señalización que permite utilizar sesiones con varios participantes en telefonía.

SS7: es un protocolo que permite iniciar y finalizar llamadas en telefonía.

SMS (Short Message Service, Servicio de Mensaje Corto): Servicio que proveen los operadores de telecomunicaciones móviles a través de dispositivos móviles permite el envío de mensajes cortos, conocidos como mensajes de texto.

SQL (Structured Query Language, lenguaje de consulta estructurada): Lenguaje de programación estándar para bases de datos relacionales, que permite especificar diversas operaciones en ellas.

SIM (Subscriber Identification Module): Es una tarjeta desmontable usada en teléfonos móviles que contiene la información del suscriptor.

TACS (Total Access Communication System): Sistema de Telefonía celular adoptado en el Reino Unido adoptado en los años de 1980.

TDMA (Time Division Multiple Access) es una tecnología inalámbrica de segunda generación para telecomunicaciones.

XML (eXtensible Markup Language): Estándar que permite definir lenguajes de marca que permite almacenar información de forma legible.

3GPP (3rd Generation Partnership Project): Estándar de telecomunicaciones de tercera generación para teléfonos móviles basado en GSM.

ANEXO 2. CASOS DE ÉXITO

Caso 1: Ufone (Pakistán), uso de analítica avanzada para estudiar y capitalizar el comportamiento del cliente.

Empresa: Ufone

Negocio: Proveedor de servicios móviles

Reto: reducir las tasas de deserción y mejorar la satisfacción del cliente

Solución: El uso de la analítica avanzada para analizar grandes volúmenes de datos y mejorar la comercialización esfuerzos de la campaña de hardware: un IBM Power Systems 795 del programa: IBM InfoSphere Streams, IBM Solución de gestión de campañas Única y varias herramientas de middleware de IBM WebSphere

Descripción:

Pakistán ha experimentado un crecimiento en la industria de telecomunicaciones de 1 millón de clientes hace diez años a 120 millones hoy. Ufone es la marca de Pak Telecom Mobile Limited (PTML), utiliza Big Data y Sistemas empresariales para fidelizar clientes y reducir la tasa de churn.

"Tenemos completa visibilidad y transparencia en cuanto a cuántos clientes se apuntan, cuántos fueron alcanzados, cuántos de ellos optaron por una campaña y, si optaron, cuál era su comportamiento", señala Faisal Khaliq, CIO de Ufone. "Ahora estamos en condiciones de calcular nuestro ROI de marketing muy rápidamente."

Según IBM (2013), Ufone implemento una solución que ha sido crucial para poder almacenar, procesar este crecimiento de información, analizando a detalle todas las llamadas que cursan por la red, la empresa puede agrupar servicios u ofrecer planes de llamadas diferentes que se adapten a las necesidades y a las preferencias individuales, de esta manera los clientes pueden aprovechar al máximo los servicios.

Antes de implementar InfoSphere Streams de IBM, Ufone había estado utilizando una solución de data-warehousing de un tercero que ejecuta sus modelos y transforma la data en listas. Esa información basada en CDR, podría tener tres horas de antigüedad. Esta información es consumida casi en línea por los ejecutivos de Marketing que ejecuta campañas internamente.

Ufone decidió reevaluar la forma en que rastreó los datos de clientes y creó ofertas de marketing individualizadas y específicas. Para ello, decidió implementar InfoSphere Streams y la solución Única Campaign Management en un entorno integrado de un solo silo, la mayoría de los cuales se ejecutan en el Power 795 en un entorno seguro.

Los empleados de marketing desarrollarán ofertas de marketing basadas en CDR, que se empujan a los servidores de Streams. Ese sistema analizará las CDR-en 30 segundos en lugar de las tres horas anteriores, gracias al procesamiento de gran velocidad de datos de Steams, y, sobre la base de reglas predeterminadas, recomendará una recompensa o servicio a ser provisionado. A continuación, el servidor examinará la base de datos para los suscriptores que coincidan con los esquemas de aprovisionamiento orientados.

El middleware de WebSphere realiza búsquedas de datos simples en una variedad de sistemas de Ufone, incluyendo CRM y facturación. Una vez completada la validación, el middleware proporcionará las ofertas de marketing basadas en reglas de negocio almacenadas en middleware o en archivos y bases de datos externos. Las respuestas resultantes se devuelven a la base de datos primaria para generar informes de inteligencia empresarial". Así indica IBM.

"Hay una gran cantidad de datos fuera de datos no-grandes," Khaliq añade: "Pero para sacar provecho de ello, hay que juntar las piezas de tecnología que le permitirán miras y ver cómo se puede añadir valor a diferentes segmentos de clientes. Esto es válido para otras industrias, como la banca y el comercio minorista,

que tienen o deberían tener información sobre los hábitos de compra de clientes y la tecnología en su lugar para darle una idea rápida. La información es una herramienta muy poderosa, pero sólo si se convierte en acción ".

Este es el caso de Ufone. Y los números lo soportan. Según Khaliq, cuando ciertas campañas están funcionando, la tasa de churn disminuye. *"Puede que no sea un número enorme", dice, "y las tasas de éxito van a variar de microsegment a microsegment y campaña a campaña, pero la diferencia es allí y notable. De hecho, ya hemos duplicado nuestras tasas de respuesta a la campaña, y todavía tenemos mucho más que hacer ".*

