

OLACEFS

ORGANIZACIÓN LATINOAMERICANA Y DEL CARIBE
DE ENTIDADES FISCALIZADORAS SUPERIORES



Ventajas de la utilización del Big Data en el proceso auditor

Autores:

Nombres	Correo	Teléfono
Hamlet Abel Morales Carcamo	hamlet.morales@cgr.gob.ni	+(505) 8373-8885
Rodrigo José Berrios Mendoza	rodrigo.berrios@cgr.gob.ni	+(505) 8380-4930

EFS:

- Contraloría General de la República de Nicaragua

Managua, Nicaragua
Agosto 2018

Índice

Constancia emitida por la EFS	2
Resumen	3
Introducción	4
Objetivos de la investigación	5
Revisión de la literatura	6
Conceptualización, definición y dimensiones de Big Data	9
Tendencias respecto a la tecnología para la administración de datos	12
Metodología y procedimientos mínimos que las Entidades de Fiscalización Superior debieran utilizar para la evaluación y análisis de Big Data	15
Beneficios en la utilización del Big Data - Limitantes a los que se enfrentan las EFS	18
Ventajas en la utilización de Big Data en el proceso de auditoría	19
Identificación de posibles riesgos en la utilización de Big Data en el proceso auditor y estrategias para minimizarlas	20
Conclusiones	23
Bibliografía o e-grafía	24

Constancia emitida por la EFS



CONTRALORÍA GENERAL DE LA REPÚBLICA Apartado Postal No. 48 - Managua, Nicaragua

Managua, 17 de agosto del 2018
CGR-URIP-ADS-131-08-2018

Sres. Comisión Técnica de
Prácticas de Buena Gobernanza (CTPG)
Organización Latinoamericana y del Caribe
de Entidades Fiscalizadoras Superiores
(OLACEFS)

Estimados Señores:

La Suscrita Responsable de la Unidad de Relaciones Internacionales y Proyectos de la Contraloría General de la República de Nicaragua, emite la presente constancia a fin de especificar los servidores públicos que estarán participando en el concurso regional sobre buena gobernanza 2018, con el tema: **“Ventajas de la utilización del Big Data en el proceso auditor”** en representación de Contraloría General de la República de Nicaragua.

A continuación se detalla la información requerida de los servidores públicos:

Nombres y apellidos	Cargo desempeñado	Correo electrónico	Sección a la que pertenece
Hamlet Abel Morales Cárcamo	Analista - Desarrollador B	hamlet.morales@cgr.gob.ni	División de Tecnologías de la Información
Rodrigo José Berríos Mendoza	Responsable de Oficina de Desarrollo de Sistemas	rodrigo.berrios@cgr.gob.ni	División de Tecnologías de la Información

Se firma la presente constancia, emitida en la ciudad de Managua, a los 17 días del mes de agosto del año 2018.

Cordialmente,


Msc. Aracelys del Socorro Duarte Sevilla
Responsable Unidad de Relaciones Internacionales y Proyectos
Contraloría General de la República de Nicaragua



Cc: Archivo

Resumen

El término de Big Data se mencionó por primera vez en el año 1941 como “Explosión de la información”, producto del crecimiento del volumen de datos de esa época.

A pesar de que el concepto de los Big Data nunca se ha definido debido a su alta complejidad, a lo largo del tiempo los científicos de datos lo han hecho de manera personal, según sus estudios e investigaciones.

Los datos de los Big Data se clasifican en estructurados, semiestructurados y no estructurados.

Los Big Data al igual que su concepto y definición, los científicos de datos en dependencia de sus estudios, clasifican las dimensiones de los Big Data en 3V, 4V y 5V.

Actualmente producto de la gran demanda en la gestión de datos de los Big Data grandes empresas como Microsoft, Google, Apache, etc, han desarrollado productos que permiten a los usuarios almacenar, clasificar, analizar y gestionar los grandes conjuntos de datos.

Antes de que una organización tome la decisión de implementar los Big Data, se debe realizar un estudio que ayude a determinar la factibilidad y viabilidad del mismo.

Al usar tecnologías que permitan el análisis y gestión de los datos de los Big Data, las organizaciones pueden obtener beneficios que se pueden resumir en competitividad, exactitud y precisión al momento de tomar decisiones.

Las limitaciones en la implementación de los Big Data pueden estar asociadas a la infraestructura tecnológica, los recursos humanos, la definición de los objetivos, las políticas de seguridad y políticas de gobierno existente en las organizaciones.

El uso de los Big Data en el proceso auditor plantea riesgos que según (Coper, 2015) deben ser considerados siendo los más relevantes los siguientes:

- 1.- Falta de habilidades y conocimientos
- 2.- No incorporar integralmente el análisis de datos en el proceso de auditoría.
- 3.- No disponer de los recursos necesarios
- 4.- La sostenibilidad de las soluciones Big Data

Para finalizar, el éxito de la implementación de los Big Data, dependerá de los recursos, infraestructura, metodología y personal capacitado que tenga la EFS para gestionar los datos.

Introducción

Desde la antigüedad, el comportamiento del ser humano ante un determinado evento era impredecible, una sugerencia solo se originaba momento después que este realizaba una determinada acción. Pero en la actualidad todo es diferente, ya que debido a la gran cantidad de datos que genera (por medio de dispositivos móviles inteligentes, sensores, ordenadores, etc) y se almacenan de manera automática día a día, se puede predecir el comportamiento del mismo haciendo uso del análisis y procesamiento de datos de los Big Data

Actualmente las Entidades Fiscalizadoras Superiores (EFS), son organismos autorizados para realizar un sin número de auditorías en el sector público de un país. Su función principal es velar por los bienes y recursos públicos que asigne cada gobierno a cada sector (entidades, instituciones, ministerios, organismos, etc), se use correctamente y de acuerdo a las directrices aprobadas que se encuentran en la normativa que rige cada país.

Teniendo en cuenta las tendencias tecnológicas en la que se encuentra sumergido el mundo actual y la cantidad de información que se genera y se almacena a diario a nivel mundial, ha surgido la tecnología Big Data, debido a la necesidad de clasificar, analizar y procesar estas grandes cantidades de información que permita a las empresas u organizaciones innovar nuevos productos y obtener mayor competitividad ante las demás según los resultados obtenidos..

Considerando que hoy por hoy la información es poder, el propósito de este documento es que las EFS tengan como referencia la información plasmada por medio de este escrito, la cual gira en torno al tema de los beneficios de la implementación del Big Data en el proceso auditor.

Esta pieza comunicativa está estructurada en objetivos de la investigación, revisión de la literatura, conceptualización, definición y dimensiones de los Big Data, tendencias respecto a la tecnologías que permiten la administración de datos, metodología y procedimientos que las EFS deben usar para evaluar y analizar una implementación de Big Data, beneficios y limitantes que las EFS pueden encontrar al momento de la implementación de Big Data, ventajas de la utilización de Big Data en el proceso auditor, del mismo modo a través del presente documento se mencionan los posibles riesgos al usar Big Data en el proceso de auditoría con sus correspondientes estrategias para minimizarlos.

Con la implementación de los Big Data en el proceso de auditoría de las EFS, se obtienen ventajas competitivas que permiten reducir los costos y recursos así como fortalecer los procesos para la obtención de evidencias mediante la implementación de un efectivo análisis y gestión de datos.

Objetivos de la investigación

- Registrar los diferentes conceptos, definiciones y dimensiones de la terminología Big Data, haciendo referencia de manera cronológica desde su origen hasta la actualidad, para una mejor comprensión del lector.
- Mostrar las tendencias tecnológicas que permitan clasificar, analizar, procesar y gestionar los datos de los Big Data a través del presente documento, con el propósito que las Entidades Fiscalizadoras Superiores conozcan dichas herramientas y sus características.
- Describir la metodología y procedimientos que las Entidades de Fiscalización Superior deben realizar para la evaluación y análisis de los Big Data.
- Destacar los beneficios en el uso de los Big Data teniendo en cuenta las limitantes con las que se pueden encontrar las Entidades de Fiscalización Superior.
- Definir las ventajas al hacer uso de los Big Data en los procesos de auditoría.
- Identificar los riesgos que se presentan durante el uso de los Big Data en el proceso de auditor.
- Proponer la estrategia para minimizar los riesgos durante el uso de los Big Data en el proceso de auditor.

Revisión de la literatura

En internet existen una lista amplia de referencias bibliográficas que pueden ser consultada para conocer y aprender sobre la tecnología de análisis de datos Big Data, sin embargo, es importante que hemos tomado en cuenta aquellas referencias que contienen la información más acertada de acuerdo a nuestro objeto de estudio.

Previo al desarrollo de la temática que abarca la estructura del documento fue necesario diseñar un bosquejo que permitiera organizar de una mejor manera las ideas que se plantean durante la elaboración del mismo. Posteriormente a través investigación documental y bibliográfica, se procedió a la selección de fuentes de información confiables usadas a lo largo de nuestro trabajo investigativo, muchas de estas fuentes corresponden a libros online, revistas, artículos y documentos web compartidos por organizaciones y científicos de datos que se han profundizado en la terminología y tecnología Big Data.

A continuación por medio del siguiente cuadro se presenta una síntesis de la literatura consultada

Autores /Año	Título	Tipo de documento	Referencias revisadas	Síntesis del documento
Teradatamagazin e.com / 2011	Teradata Magazine	Revista	Big Data	Revista creada con el objetivo de abordar las principales tendencias de hoy en día en relación al crecimiento de la Big Data y el futuro de los teradata
Silvia Andrián / 2016	El poder de la auditoría a través del análisis de los datos	Documento	Concepto, ¿Qué dicen los analistas del mercado?	Este documento aborda la definición de Big Data en los negocios de hoy, su estructura, oportunidades y amenazas, el papel del análisis de datos en el proceso de auditoría
Inés Álvarez / 2016	¿Qué es el big data?	Artículo web	Completo	En este artículo se aborda el concepto de Big Data y expectativas acerca de la utilización de esta tecnología en el futuro.

Apache / 2014	Apache Hadoop	Artículo web	Completo	En este artículo se describe el proyecto Hadoop, sus módulos y versiones disponibles
BAOSS / 2015	10 herramientas para Big Data Analytics	Artículo Web	Completo	En este artículo se abordan las 10 principales herramientas para manejar Big Data Analytics y sus características
BAOSS / 2017	Big Data: Guía para no perderse	Artículo Web	Completo	En este artículo se aborda el concepto de Big Data, sus dimensiones y los principales desafíos y oportunidades que conlleva la implementación de Big Data
Information Communication and Society / 2012	Critical Questions for Big Data	Documento	Completo	En este documento se abordan 6 aspectos críticos desde la opinión del autor.
Price Waterhouse Cooper / 2015	Análisis de datos en la gestión del riesgo	Documento	Completo	En este documento pricewaterhouse cooper presenta su experiencia y su punto de vista en la implementación de análisis de datos en el proceso de auditoría
Emagister / 2017	Herramientas para Big Data más utilizadas	Artículo	Completo	En este artículo se realiza una descripción de las herramientas para Big Data más usadas.
Google / S.F.	Ventajas de Google Cloud Platform	Artículo	Completo	En este artículo se describen las características principales de la tecnología de Google Cloud y sus principales ventajas
Howard Ellias / 2012	El desafío de Big Data: Cómo desarrollar una estrategia ganadora	Artículo web	Completo	En este artículo se hace un análisis del concepto de big data y los aspectos a considerar para establecer una estrategia para su implementación.

IBM / S.F.	The four Vs of Big Data	Infografía	Completo	Un recurso por medio el cual los científicos de datos de IBM describen las características principales de las 4 dimensiones del Big Data.
IDC / 2012	Nuevas tendencias tecnológicas han entrado a AL	Artículo web	Completo	En este artículo la consultora IDC menciona conceptos, definiciones, estudio de los Big Data
Laney / 2001	Application Delivery Strategies	Documento Web	Completo	Un documento web de la consultora Gartner que hace mención por primera vez de las dimensiones de los Big data
McKinsey / 2011	Big Data: The next frontier for innovation, competition and productivity	Documento Web	Completo	La consultora McKinsey menciona conceptos, dimensiones, ventajas y predicciones referente al aumento del volumen de datos de los Big Data.
Microsoft / 2014	Microsoft presenta la más completa para la gestión de Big Data	Artículo web	Completo	El artículo aborda las características de la plataforma de Microsoft Azure para la gestión de Big Data
Prometeus / 2017	Volumen, Variedad, Velocidad, Veracidad y Valor, las 5 dimensiones del Big Data	Artículo Web	Completo	Prometeus quien es una empresa española dedicada al análisis de datos describe las dimensiones de los Big Data en 5V
Soft / S. F.	Seis retos en la implementación del Big Data	Artículo Web	Completo	En este artículo se abordan las principales dimensiones del big data y los retos desde la perspectiva del autor que se deben superar al momento de considerar una implementación de Big Data.
Gil Press / 2013	A Very Short History of Big Data	Artículo Web	Completo	Este artículo aborda la historia del Big Data desde sus inicios hasta

				el día de hoy y como se ha venido adaptando su definición.
Goberna / 2017	Manual sobre utilidades del Big Data para bienes públicos	Libro	23, 59, 81	Este es un manual elaborado para compartir las experiencias de expertos en la materia y funcionarios públicos vinculados con el ámbito tecnológico.

Conceptualización, definición y dimensiones de Big Data

En la actualidad es muy frecuente escuchar la terminología “Big Data”, las personas de diferentes organizaciones hablan e investigan sobre este interesante tema, durante la investigación es común que estas se hagan preguntas como: ¿Cuándo surgió esta terminología?, ¿Cuáles fueron sus orígenes?, ¿Qué son los Big Data?, ¿Cuáles son sus características o dimensiones?, etc.

Es importante destacar que los Big Data se mencionaron por primera vez según (Press, 2013) en el año 1941, bajo el término “Explosión de la información”, producto del crecimiento en el volumen de los datos de esa época.

Tres años después en (Press, 2013) afirma que en 1944 el escritor Fremont Rider publica en su libro “The Scholar and the Future of the Research Library”, el mismo fenómeno de crecimiento en los volúmenes de datos, al indicar que las bibliotecas estadounidenses aumentaban su tamaño aproximadamente cada dieciséis años y que según sus cálculos para el 2040, en la biblioteca en la cual trabajaba, iban a existir alrededor de 200,000,000 de volúmenes, que necesitan alrededor de 6,000 millas de estantes y un personal mayor a seis mil personas para controlarlo.

Posteriormente en 1961 (Press, 2013) menciona el autor de “Science Since Babylon” Derek Price, expresa que el crecimiento científico se debe al aumento en el número de publicaciones y artículos. Del mismo modo hizo referencia al fenómeno de la “Ley de aumento exponencial”, basado en el incremento de la población mundial, el autor aducía que cada avance o estudio científico genera en las nuevas generaciones mayor conocimiento y nuevos estudios que producen una mayor documentación.

No obstante (Press, 2013) expresa que en 1967 Marron y de Maine publicaron un artículo llamado “Automatic data compression”, donde se menciona que debido a la explosión de la información las unidades de almacenamiento de esa época no eran suficiente para guardar tantos datos, por lo que proponían aumentar la tasa de transmisión de la información de manera rápida y automática a través de ordenadores y medios externos.

(Press, 2013) indica que durante el mes de abril de 1980 IA Tjomsland brinda una charla en el cuarto simposio de la IEEE sobre sistemas de almacenamiento masivo, donde expresa a los presentes una problemática que tenía que ver directamente con los usuarios, la cual era no saber clasificar la información dando lugar al almacenamiento de datos obsoletos e inútiles, en lugar de guardar datos potencialmente útil a futuro.

Consecutivamente (Press, 2013) describe que en 1990 Peter J. Denning publica "Saving All the Bits", donde plantea la construcción de ordenadores que ayuden a clasificar y procesar las grandes cantidades de información como una solución al problema que es ocasionado por guardar todos los bits que producen la velocidad y volumen de los datos.

Seguidamente (Press, 2013) expresa que en 1997 Michael Cox y David Ellsworth mencionan por primera vez el término "Big Data" en un artículo de la 8va conferencia de la IEEE, donde se indican que el problema de los Big Data se resuelve adquiriendo más recursos.

(Press, 2013) menciona que en febrero del 2001 Doug Laney a través de una publicación define 3 dimensiones de los Big Data los cuales son: Volumen, Variedad y Velocidad, estos fueron aceptadas por la mayoría de los científicos de datos. IDC en marzo del 2007 publica a través del libro "The Expanding Digital Universe: A Forecast of Worldwide Information Growth through 2010", el primer estudio que estima y pronostica la cantidad de datos digitales creados y replicados cada año, donde IDC estimó que para el 2006 se creó 161 exabytes de datos y pronostica que se dupliquen cada 18 meses.

En lo sucesivo a esto, en el año 2010 y 2012 se realizó una continuidad al mismo estudio donde se encontró que la cantidad de datos digitales creados anualmente superó el pronóstico realizado por el IDC en el estudio del 2007.

A pesar de que el concepto de los Big Data nunca se ha definido debido a su alta complejidad, a lo largo del tiempo los científicos de datos lo han hecho de manera personal, según sus estudios e investigaciones (McKinsey, 2011) Define los Big Data como "Conjuntos de datos cuyo tamaño está más allá de las capacidades de las herramientas típicas de software de bases de datos para capturar, almacenar, gestionar y analizar", de igual forma (Adrian, 2011) Menciona que los Big Data "Exceden el alcance usado comúnmente en los entornos de hardware y herramientas de software para capturar, gestionar y procesar los datos dentro de un lapso de tiempo razonable para los usuarios".

Otras definiciones que surgieron posterior a este año, fueron la de (Boyd & Crawford, 2012) cuando consideran que los Big Data son "un fenómeno cultural, tecnológico y académico que está basado en la interacción de la tecnología, análisis y mitología", mientras que (Howard, 2012) describe a los Big Data como

“Conjuntos de datos tan grandes como diversos, que rompen las infraestructuras de TI tradicionales”.

Así mismo (IDC, 2012) se expresa de los Big Data como “nueva generación de tecnologías y arquitecturas, diseñadas para extraer un Valor económico de un Volumen muy grande de amplia Variedad de datos, con una alta-Velocidad de captura, descubrimiento, y / o análisis de la información”.

(Álvarez, 2016) Define el concepto de los Big Data como “el análisis de información en cantidades industriales, todo un universo de oportunidades para las empresas aún por explorar”.

Actualmente los datos de los Big Data se clasifican en 3 partes:

- Datos estructurados:
Aquellos datos que se generan comúnmente en las transacciones de los sistemas de información con base de datos relacionales.
- Datos semiestructurados:
Son datos que están separados y ordenados a través de etiquetas o marcadores, por ejemplo archivos en formato JSON, XML y HTML.
- Datos no estructurados:
Son los datos más difícil de gestionar y administrar, pero producto al incremento de los mismos han surgido herramientas para facilitar su administración como bases de datos no SQL, un ejemplo de estos tipos de datos pueden ser la fotografías, audios, videos, mensajes de textos, correos electrónicos, documentos pdf, Excel, Word, etc.

Las dimensiones de los Big Data son las características que deben cumplir los conjuntos de datos para ser considerados como Big Data, pero estas al igual que las definiciones varían entre los científicos de datos; Inicialmente para (Laney, 2001) las dimensiones eran las famosas 3 V “Volumen, Variedad, Velocidad”, tiempo después científicos de datos de (IBM) clasifican a través de una infografía las dimensiones de los Big Data en 4V “Volumen, Variedad, Veracidad, Velocidad”, posteriormente a estos (Prometeus, 2017) clasifica las dimensiones en 5v “Volumen, Variedad, Veracidad, Velocidad y Valor”.

Cuando se habla de **volumen** es un término que alude al continuo incremento en los datos día a día, según los estudios realizados por (IBM) “la información crecerá tanto que para el 2020 se crearán 43 billones de gigabytes de datos, un aumento 300 veces mayor que la cantidad de datos en 2005”.

Otra de las dimensiones de los Big Data es la **variedad**, esta tiene que ver con las distintas formas en que se encuentran los datos, (IBM) expresa que los “Big Data abarcan información de múltiples fuentes tanto internas como externas, como transacciones, redes sociales, contenido empresarial, sensores y dispositivos móviles, pero que esto va a depender según el tipo industria u organización”.

La **velocidad** es una dimensión más de los Big Data que tiene que ver mucho con el desempeño en el procesamiento de los datos, ya que estos según (Prometeus, 2017) “deben hacerse en el menor tiempo posible e incluso en tiempo real”, esta misma fuente menciona que “la velocidad se requiere para acceder y visualizar los datos, con el objetivo de facilitar el análisis y extraer conclusiones”. Un ejemplo claro de la velocidad es el que nos brinda los estudios de (IBM) donde indica que en “La bolsa de valores de Nueva York se procesa 1 terabyte de información comercial durante cada sesión de negociación”.

La **veracidad** es una de las dimensiones más importantes de los Big Data, ya que de esta según (Prometeus, 2017) “depende el acierto de las tomas de decisiones”, encuestas realizadas a los líderes empresariales por (IBM) menciona que “1 de cada 3 no confían en la información que utilizan para tomar decisiones”, es por esto que (Prometeus, 2017) expresa lo fundamental que es “controlar la integridad de los datos durante la manipulación de los mismos”.

En última instancia y no menos importante se encuentra la dimensión del **valor**, que consiste según (Prometeus, 2017) en “utilizar los datos de manera rentable y a la vez eficiente”, una de las ventajas de esta característica, es que producto de la analítica de los datos pueden surgir nuevos productos además de nuevas competencias y tendencias que permitan a las organizaciones generar mayores ingresos y ser competitivas ante las demás empresas.

En definitiva (Prometeus, 2017) define los Big Data como un “conjunto de tecnologías y procesos que están permitiendo capturar y almacenar cantidades inteligentes de datos de diversos orígenes y tipologías”.

Debido al continuo crecimiento en los datos de los Big Data y la necesidad de almacenar y gestionar los mismos, surgen las tendencias tecnológicas que permiten la administración de datos.

Tendencias respecto a la tecnología para la administración de datos

Teniendo en cuenta los constantes cambios tecnológicos y la demanda actual que existe en la administración y gestión de datos de los Big Data, por parte de muchas empresas u organizaciones, Google, Microsoft, Apache, entre otros han desarrollado productos que permiten a los usuarios visualizar, gestionar y explotar a través de la analítica, datos de los Big Data.

En primera instancia tenemos **Hadoop** versión 3.1, una herramienta de código abierto desarrollada y ofrecida a los usuarios por (Apache, 2018) gratuitamente que permite “el procesamiento distribuido de grandes conjuntos de datos a través de clústeres de computadores que usan modelos de programación simples”.

Hadoop está compuesto por 4 módulos:

- Hadoop Common que tiene utilidades o librerías comunes que son compatibles con los otros módulos de Hadoop.
- Sistema de archivos distribuidos de Hadoop el cual se encarga de proporcionar acceso de alto rendimiento a los datos de las aplicaciones.
- Hadoop YARN, una librería que permite administrar la programación de trabajo y los recursos del cluster.
- Hadoop Mapreduce, este es usado para el procesamiento paralelo de grandes conjuntos de datos.

Sin embargo, (BAOSS, 2015) comparte en su sitio web que “empresas como Flipboard, Groupon, Twitter, Spotify, Yahoo (Japón), WebMD, Baidu y muchos más” usan **Storm** como una herramienta alternativa para la gestión de datos similar a Hadoop. Esto se debe a que “Es fácil de usar y puede ser integrada con cualquier lenguaje de programación” además que “Ofrece una interfaz fácil y fiable que permite el procesamiento de los datos en tiempo real”.

Así mismo (BAOSS, 2015) propone reemplazar el módulo Mapreduce de Hadoop por **Spark** ya que este puede ser 100 veces más rápido que Mapreduce y “Permite combinar SQL, streaming y librerías analíticas más avanzadas”.

Del mismo modo (González, 2017) propone hacer uso de un sistema de almacén de datos llamado **Hive** el cual “permite realizar resúmenes de datos, consultas y análisis de datos mediante HiveQL (un lenguaje de consultas similar a SQL)”, una de las principales características que brinda Hive es que “permite proyectar la estructura del proyecto en datos que en gran medida no están estructurados”.

Sin embargo (Microsoft, 2014) ofrece componentes que ayudan a la gestión de los Big Data a través de su plataforma en la nube **Microsoft Azure**, estos componentes son:

- SQL Server 2014 un software que presenta un alto desempeño al capturar, procesar y entregar resultados de las consultas realizadas a las diferentes bases de datos.
- Intelligent Systems Service (ISS), su función consiste en capturar los datos de forma ordenada sin importar la fuente (puede ser machine to machine, sensores, dispositivos, transacciones, etc) y el tipo (Estructurados, Semiestructurados y no estructurados).
- Analytics Platform System (APS), sirve para almacenar y gestionar los datos a través de un Data Warehouse, este software según (Microsoft,

2014) “combina las mejores funcionalidades de Microsoft SQL y de la tecnología Hadoop en un producto de bajo coste”.

- Power BI para office 365, es un componente que permite a los usuarios combinar datos locales con la nube, con el propósito de brindar a los mismos una mejor visualización y entendimiento de los datos.

Por otro lado (Google, s.f.) propone la plataforma **Google Cloud** que permite la administración de datos a través de los siguientes componentes:

- BigQuery, software por medio del cual se puede realizar consultas a grandes volúmenes de datos de manera eficiente y rápida mejor dicho casi en tiempo real.
- Cloud Storage, un servicio en el que los usuarios pueden almacenar datos ya sea de tipo SQL o NO SQL de forma unificada para su posterior análisis.
- Cloud Dataflow, herramienta mediante el cual se puede desarrollar y ejecutar una gran variedad de patrones de procesamientos de datos.
- Cloud Pub / Sub, permite capturar de forma automática, datos que se generan durante las transmisiones de un determinado evento por ejemplo “las interacciones de los usuarios con los diferentes botones de Me gusta, Me encanta, Me divierte, etc de Facebook durante una transmisión en vivo”, posteriormente las procesa a través de Cloud Dataflow y BigQuery para su debido análisis.
- Cloud DataProc, integra las herramientas Spark y Hadoop en la nube.
- Cloud DataLab, un componente por medio del cual se puede explorar, analizar, transformar y visualizar los datos además de generar modelos de aprendizaje de forma automática.
- Cloud Dataprep, herramienta que permite a los usuarios que puedan explorar, limpiar y preparar los datos estructurados y sin estructurar para su debido análisis. Una de sus principales características es que además de la visualización de datos predice y sugiere la transformación de datos ideal.
- Google Data Studio, es un servicio en cual los usuarios pueden realizar informes interactivos de manera gráfica de cualquier fuente de datos una de sus principales ventajas es que permite que múltiples usuarios puedan colaborar con el desarrollo del informe desde cualquier lugar.

En nuestra opinión consideramos que para la implementación de una plataforma que permita gestionar los datos de los Big Data en las EFS van en dependencia de los recursos o presupuesto asignado que tenga la misma y el tamaño de los Big Data.

Por un lado podrían tener su propia infraestructura en la cual puedan tener instalado las herramientas antes mencionada (**Hadoop, Spark, Storm, Hive, etc**) que permitan a la EFS clasificar, analizar, procesar y gestionar los datos las veces que deseen sin costo alguno y sin restricciones hablando monetariamente, ya que empresas como (Google, s.f.) Y (Microsoft, 2014) además de cobrar por cada componente o servicio adicional en su plataformas, cobran por la transmisión, almacenamiento y procesamiento de los datos.

Por otro lado al adquirir el licenciamiento del uso de los diferentes componentes o servicios en las plataformas **Microsoft Azure** o **Google Cloud**, evita a la EFS el pago de personal que de mantenimiento a la infraestructura de los Big Data, además de gastar en instalación y configuración adicional cada vez que se necesite más recursos para el almacenamiento y procesamiento de los datos. Una de las ventajas es que al contratar cualquiera de estos servicios, es que los datos y la información estarán disponibles 24 / 7 desde cualquier dispositivo.

Considerando todas estas tendencias tecnológicas que permiten la administración y gestión de los datos, es necesario conocer, métodos de evaluación y análisis que las Entidades de Fiscalización superior puedan usar al momento de gestionar e implementar los Big Data.

Metodología y procedimientos mínimos que las Entidades de Fiscalización Superior debieran utilizar para la evaluación y análisis de Big Data

Previo a la implementación de los Big Data, recomendamos realizar un estudio que ayude a determinar la **factibilidad, rentabilidad y viabilidad** del mismo:

- El estudio de factibilidad brindará a la EFS un análisis de su realidad actual, ya que mostrará si esta cuenta con los recursos (económicos y talento humano capacitado) necesarios con los que se pueda llevar a cabo el proyecto de la implementación de los Big Data.
- El estudio de rentabilidad permitirá a la EFS visualizar y medir el costo – beneficio de esta implementación, ya que evalúa el monto a invertir y el beneficio a obtener a lo largo del tiempo.
- El estudio de viabilidad ayudará a la EFS a determinar a través de un estudio más profundo, si el proyecto de implementación de los Big data se puede realizar, además se encargará de estudiar y valorar detalladamente la sostenibilidad y rentabilidad del mismo a lo largo de los años.

Sin embargo, (Baoss, 2017) recomienda a todas las empresas y organizaciones que plantearse cuatro preguntas básicas antes de cualquier decisión de implementar los Big Data, descritas a continuación:

1. ¿Los datos de la organización son lo suficientemente grandes?

Si bien es cierto, el tamaño de los Big Data aún no está establecido, pero algunas empresas toman como medida los Petabytes y otras los Terabytes, sin embargo a lo que (Baoss, 2017) se refiere es a que si los datos que posee la organización cumplen al menos con las dimensiones de los Big Data.

2. ¿Los resultados de los datos son relevantes para la organización?

Los Big Data surgen a raíz de las preguntas que se realizan los usuarios y las respuestas que esperan del procesamiento y análisis de datos, a como (Baoss, 2017) lo indica “es muy costoso y complejo implementar una solución Big Data”, por lo que recomienda considerar anticipadamente los resultados que se obtendrán a través del procesamiento de datos y su nivel de impacto dentro de la organización.

3. ¿La organización se encuentra preparada para la implementación de los Big Data?

(Baoss, 2017) menciona que no solo se necesita contar con los recursos de Hardware y Software necesarios para la gestión de datos de los Big Data. A nuestro criterio él tiene razón ya que se requiere mucho más que eso, se necesita personal capacitado o que al menos tengan conocimiento y experiencias en análisis y estadísticas; además se considera necesario que existan o se definan procesos para la recopilación, clasificación, análisis y gestión de los datos asegurando la integridad de los mismos.

4. ¿Las expectativas de la organización son realistas?

Los costos y expectativas definidas para un proyecto de Big Data deben ser realistas conforme al tiempo planificado para su implementación. (Baoss, 2017) propone que debe existir una comunicación estrecha entre Tecnologías de la Información y las áreas de negocio cuyo resultado sea la definición de objetivos claros y selección de las herramientas tecnológicas necesarias.

De esta forma la organización obtiene una visión de la magnitud, complejidad, beneficios, etc. de la implementación de los Big Data.

En resumen, según las recomendaciones que ha mencionado (Baoss, 2017) y los conocimientos que se adquirieron durante el proceso de investigación y redacción de este escrito, se proponen los siguientes procedimientos mínimos que se deberían tomar en cuenta para el análisis y evaluación de la implementación de los Big Data en las EFS:

1. Realizar un estudio que determine la factibilidad, rentabilidad y viabilidad de implementar los Big Data en la EFS deseada.
2. Evaluar si la EFS cuenta al menos con un conjunto de datos lo suficientemente grande, que por lo menos se pueda medir en Terabytes.
3. Verificar si el conjunto de datos con el que se desea trabajar posee datos estructurados, semiestructurados y no estructurados.
4. Elaborar preguntas que sirvan como pauta al momento de analizar y procesar los datos.
5. Evaluar si las preguntas realizadas anteriormente requieren de una implementación de Big Data para ser contestadas o si pueden ser contestadas vía SQL a través de datos estructurados.
6. Establecer una planificación especificando las tareas que se llevarán a cabo para implementar los Big Data en las EFS.
7. Realizar un plan de pruebas que garantice que tanto la infraestructura como los componentes de software adquiridos funcionan correctamente.
8. Definir el alcance del proyecto y las metas a lograr a nivel institucional con la implementación de los Big Data.
9. Establecer estrategias que permitan tanto al área de informática como a las distintas áreas de la EFS cumplir con las metas propuestas anteriormente como plan de contingencia ante eventos inesperados.
10. Investigar, analizar, evaluar y seleccionar las herramientas de software que son necesarias según las metas establecidas durante la implementación de los Big Data.
11. Seleccionar la infraestructura en la que se implementará los Big Data ya sea de manera local en la EFS o a través de las plataformas en la nube que ofrecen empresas como (Google, s.f.) y (Microsoft, 2014).
12. Evaluar que los recursos económicos que posee la EFS garanticen la sostenibilidad del proyecto a lo largo del tiempo.
13. Adquirir la infraestructura y software seleccionados en los procedimientos anteriores concluyendo con debida configuración e instalación.
14. Ejecutar el plan de pruebas realizado anteriormente, para garantizar que la implementación de Big Data está correcta.

15. Capacitar al personal en los diferentes componentes de software seleccionados, por medio del cual se analizarán y procesarán los datos de los Big Data.
16. Establecer y aplicar procedimientos de uso, políticas de integridad, confidencialidad y normas de seguridad que garanticen la correcta gestión de los datos.
17. Notificar al personal de auditorías o personal indicado por la EFS con el propósito de involucrarlos en el proceso de la implementación de Big Data, con el fin promover su uso y aprovechamiento de la herramienta.

Beneficios en la utilización del Big Data - Limitantes a los que se enfrentan las EFS

Beneficios

Transparencia (Andrián, 2016): En la medida que las entidades fiscalizadoras sean capaces de proporcionar información a los ciudadanos, mediante la gestión de la información y documentación generada a través de Big Data se fortalecerá el sistema de transparencia y rendición de cuentas de la administración pública.

Exactitud (Andrián, 2016): Las herramientas de Big Data permiten establecer un universo mayor de la información analizada, lo que genera una mayor confianza en la exactitud de las pruebas y resultados obtenidos, disminuyendo los riesgos existentes en la ejecución de procesos de auditoría. .

Precisión (Andrián, 2016): Se debe considerar la calidad de la información analizada de forma masiva a través de Big Data, dado que no siempre significa que la información analizada tenga el valor requerido, por ello, el volumen de los datos además de ser grande, debe ser profundo, de calidad y procesable, a fin de analizarlos con cierta velocidad y obtener información precisa y confiable.

Toma de decisiones (Andrián, 2016): Un efectivo modelo de análisis de la información garantizará que las fuentes de información usadas para el análisis de los datos mediante herramientas de Big Data generen información oportuna y en tiempo real que permitan tomar las decisiones acertadas.

Crecimiento (Andrián, 2016): El uso efectivo de los datos a través de sistemas analíticos permitirá transformarlos en información útil permitirá ser más eficientes en la administración y manejo de los datos.

Limitantes

A continuación se detallarán las limitantes que pueden presentarse en las EFS al momento de implementar Big Data:

1.- Infraestructura tecnológica (soft, s.f.): Las entidades deberán superar las restricciones presupuestarias y la dificultad para estimar la rentabilidad de desarrollar y adoptar la tecnología necesaria y requerida para almacenar, procesar y analizar adecuadamente altos volúmenes de información a la velocidad requerida.

2.- Recursos humanos(soft, s.f.): Esto representa una de las mayores limitantes que se puede presentar en las entidades fiscalizadoras, disponer del capital humano necesario y capacitado para explotar el potencial de dicho desarrollo tecnológico, basado en que la complejidad de los sistemas de Big Data, necesitará de equipos técnicos totalmente dedicados a dar soporte a la plataforma, a nivel de infraestructura y para el tratamiento de la información mediante el conocimiento de técnicas, procesos y metodologías para interpretar y sacar valor de los datos.

3.- Identificar los datos útiles (soft, s.f.): Con el volumen y variedad de la información que existe en las entidades fiscalizadoras, otra limitante es la identificación de los datos útiles y su aprovechamiento mediante la implementación del Big Data.

4.- Definir claramente los objetivos(soft, s.f.): Se debe estar claro en cuál es el objetivo que se quiere alcanzar mediante la utilización de grandes volúmenes de datos a fin de mejorar la toma de decisiones y una adecuada implementación de Big Data para su correcta aplicación.

5.- Políticas de seguridad (soft, s.f.): Con la implementación del Big Data la seguridad informática constituye un factor importante a considerar en las entidades fiscalizadoras, en vista, que se deberán fortalecer o rediseñar los esquemas de seguridad de la información actuales a fin de tener la capacidad de respuesta frente a incidentes que se presenten en el futuro.

Ventajas en la utilización de Big Data en el proceso de auditoría

Las ventajas que puede obtener una EFS al implementar los Big Data en el proceso de auditoría son muchas entre estas tenemos:

1. Ofrece una mayor confianza en el resultado de las pruebas al aumentar el universo de una muestra.
2. Reduce costos y tiempos de ejecución al obtener información de una forma más ágil y en menor tiempo.

3. Mejora la valoración de los riesgos dentro de una organización identificando anomalías y patrones de alto riesgo para lograr una mejor cobertura.(Coper, 2015).
4. Permite una comunicación eficaz de aspectos y resultados importantes a través de tableros de control dinámicos.(Coper, 2015)
5. Se obtiene un conocimiento más profundo de una organización para ayudar en la inteligencia de la auditoría.(Coper, 2015)
6. Incorpora funciones de auditoría más relevantes para la organización y la toma de decisiones.(Coper, 2015)
7. Según los datos del conjunto de datos que se analiza y procesa, se puede predecir el comportamiento de un servidor público ante un posible acto de corrupción.
8. Los usuarios pueden crear, consultar y visualizar informes usando diferentes fuentes de datos de manera fácil, ágil y rápida simultáneamente.
9. Se puede optimizar recursos del presupuesto general de la República, enfocándose a través de un plan generado por la implementación de los Big Data en aquellas instituciones públicas que pueden tener altos índices de corrupción.
10. La información está disponible 24/7 según la plataforma desde cualquier dispositivo u ordenador que tenga conexión a internet en caso que no se encuentre en la EFS.

Identificación de posibles riesgos en la utilización de Big Data en el proceso auditor y estrategias para minimizarlas

El uso de Big Data plantea una serie de riesgos en el proceso auditor, logrando identificar los siguientes:

- 1.- Falta de habilidades y conocimientos. (Coper, 2015)

Los auditores que trabajan con Big Data requieren conocimientos de análisis de datos, modelado estadístico y seguridad de TI, ya que sin las habilidades apropiadas resulta muy arduo para los auditores sintetizar diversos datos puntuales, comprender tendencias y decidir dónde enfocarse.

Para minimizar este riesgo es importante que dentro de la organización se creen grupos de usuarios avanzados con sólidas habilidades técnicas y analíticas que se encarguen de asistir a los usuarios menos hábiles quienes a su vez deberán

recibir capacitaciones para comprender cómo reconocer el uso de datos y sintetizar los resultados.

2.- No incorporar integralmente el análisis de datos en el proceso de auditoría. (Coper, 2015)

Los equipos de auditoría hoy en día intentan incorporar técnicas de análisis de datos en forma separada y no utilizan el análisis de datos en el ciclo completo de una auditoría, lo que limita la capacidad de identificar y evaluar áreas de mayor riesgo.

Para dar mayor valor al análisis de datos, los auditores deben considerar hacer uso de los datos en todos los aspectos del proceso de auditoría, desde la evaluación del riesgo hasta la presentación de sus informes y el seguimiento y de esta forma podrán aportar conclusiones de mayor impacto en forma constante, demostrando mayor eficiencia en la asignación de recursos.

3.- No aplicar adecuadamente la tecnología disponible (Coper, 2015)

La tecnología en el proceso de auditoría debe ser vista como un instrumento posibilitador por tanto, una capacitación deficiente de los equipos de auditoría puede afectar a corto y largo plazo el cumplimiento de los objetivos organizacionales.

Elaborar planes de capacitación orientados a fortalecer habilidades y conocimientos de los auditores en el uso de las herramientas TIC disponibles en su organización, constituye un factor fundamental para avanzar en el proceso de integración del análisis de datos y que debe ser considerado e incorporado en el proceso de formulación del presupuesto anual de la organización.

4.- No recabar datos relevantes (Coper, 2015)

Los equipos de auditoría presentan dificultades para obtener información oportuna, completa, exacta y relevante como resultado de la relación existente con la disponibilidad de los datos provenientes de múltiples sistemas, por problemas con la calidad de los datos, a la tecnología limitada para acceder a los datos y a datos disponibles fuera de sistemas transaccionales (no estructurados) representan los obstáculos que los auditores deben sortear para incorporar el análisis de datos en sus procesos de auditoría.

Para superar esto, será necesario que los auditores evalúen qué datos usados por otros departamentos sirven para su análisis y se asocien con el departamento de TI de su organización, a fin de identificar en conjunto los datos necesarios, fiables y de fácil acceso que estén disponibles en múltiples fuentes y formatos por medio del aprovechamiento de la tecnología de Big Data.

Según la experiencia de Price Waterhouse Cooper, los departamentos de auditoría que obtienen mayor valor del análisis de los datos son los que construyen habilidades y tecnologías requeridas para hacer uso de los datos en todas las etapas de sus procesos de auditoría.

5.- No disponer de los recursos necesarios (Coper, 2015)

Un riesgo que se puede presentar durante el proceso de implementación del Big Data es que debido a una mala planificación, no se hayan contemplado los procesos, capacidades profesionales, tecnología y el presupuesto requerido para culminar el plan de implementación con éxito.

Por ello es importante realizar una inversión en capacidad tecnológica y habilidades de análisis datos los cuales deben estar claramente definidos en objetivos realistas y plazos que se ajusten al presupuesto y capacidades de la organización.

6.- La sostenibilidad de las soluciones Big Data (Coper, 2015)

Si no se establece una sólida gestión del proyecto y monitoreo, así como una administración del cambio, con el pasar del tiempo la implementación del Big Data puede pasar desapercibida e incluso desaparecer.

Se debe fomentar una cultura apropiada en la organización y hacer que los auditores se entusiasmen con el proyecto, a fin de fomentar una nueva mentalidad que les permita identificar nuevas técnicas de análisis que vengán a reemplazar o complementar enfoques de auditorías anteriores.

Conclusiones

De acuerdo a la información recopilada durante el desarrollo y elaboración del presente documento, se puede observar que los Big Data no tienen un concepto o definición específico, sin embargo cronológicamente se aprecia que aunque la terminología Big Data para muchas personas es nueva, en realidad no lo es ya que tiene sus orígenes alrededor del año 1941.

De la misma manera aunque existan variaciones en las dimensiones y características de los Big Data, se observa que siempre están presente las 3V “Volumen, Variedad, Velocidad” propuestas por (Laney, 2001).

Es importante que las EFS que deseen implementar los Big Data en su organización, tomen en cuenta los procedimientos propuestos a través de este escrito, puesto que facilitan una guía que apoya a la toma de decisiones al momento de evaluar y analizar la posibilidad de la implementación de los Big Data en la misma.

Durante el desarrollo y cumplimiento de los objetivos para elaborar esta investigación, se observa que es importante elegir una tecnología que permita la administración y gestión de los datos en dependencia de los recursos y necesidades de la EFS. Para ello se recomienda realizar un cuadro comparativo donde se muestren las ventajas y desventajas de tener su propia infraestructura versus alquilarla a través de una plataforma en la nube, tomando en cuenta las variables como recursos, costos, mantenimiento, tiempo de procesamiento, entre otros.

Para finalizar, se considera que la implementación de Big Data en el proceso auditor, brinda a la EFS un sin número de ventajas que pueden ayudar a mejorar el proceso como tal en base a la toma de decisiones originadas por el procesamiento y gestión de datos de los Big Data, pero que para ello es necesario que cuenten no solo con los recursos de Hardware y Software necesarios, sino que también dispongan de un personal calificado que haga uso de políticas de seguridad, integridad y privacidad durante la gestión de los mismos.

Bibliografía o e-grafía

- Adrian, M. (2011). Big Data. *Teradata Magazine*, 40-41. Obtenido de nextbook media: http://www.nextbook.com/nxtbooks/mspcomm/teradata_2011q1/index.php?startid=8#/42
- Andrián, S. (24 de 05 de 2016). *Federación Latinoamericana de Bancos*. Obtenido de Felaban: https://www.felaban.net/archivos_memorias/archivo20160524213147.PM.pdf
- Álvarez, I. (10 de 02 de 2016). *Que es el big data*. Obtenido de Forbes: <http://forbes.es/business/3238/que-es-el-big-data>
- Apache. (08 de 08 de 2018). *Hadoop*. Obtenido de Apache: <http://hadoop.apache.org/>
- BAOSS. (18 de 11 de 2015). *10 HERRAMIENTAS PARA MANEJAR BIG DATA ANALYTICS*. Obtenido de BAOSS Analytics Everywhere: <https://www.baoss.es/10-herramientas-para-manejar-big-data-analytics/>
- BAOSS. (2017,08 17). Big Data: Guía para no perderse, Obtenido de *BAOSS Analytics Everywhere*: <https://www.baoss.es/guia-definitiva-big-data/>
- Boyd, D., & Crawford, K. (05 de 2012). *Taylor & Francis online*. Obtenido de Taylor & Francis online: <https://www.tandfonline.com/doi/pdf/10.1080/1369118X.2012.678878>
- Coper, P. W. (12 de 08 de 2015). *PWC*. Obtenido de PWC: <https://www.pwc.com/mx/es/risk-assurance-services/archivo/20150812-am-foll-data-analytics.pdf>
- González, C. (2017, 01 10). Herramientas para Big Data más utilizadas, Obtenido de *Emagister*. From Emagister: <https://www.emagister.com/blog/herramientas-big-data-mas-utilizadas/>
- Google. (s.f.). Cloud computing, Obtenido de Google Cloud: https://cloud.google.com/gcp/?hl=es&utm_source=google&utm_medium=cpc&utm_campaign=latam-LATAMsp-all-es-dr-skws-all-all-trial-b-latam-1003997-LUAC0000387&utm_content=text-ad-none-any-DEV_c-CRE_257142332802-

[ADGP_SKWS%20%7C%20BMM%20~%20Hadoop-KWID_4370003008](#)

Howard, E. (03 de 07 de 2012). *El desafío de Big Data: Cómo desarrollar una estrategia ganadora*. Obtenido de CIO PERÚ: <https://cioperu.pe/articulo/10442/el-desafio-de-big-data-como-desarrollar-una-estrategia-ganadora/>

IBM. (s.f.). *4-vs-of-big-data*. Obtenido de IBM: https://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg

IDC. (20 de 11 de 2012). *IDC Release*. Obtenido de IDC LatinoAmerica: <http://mx.idclatin.com/releases/news.aspx?id=1433>

Laney, D. (06 de 02 de 2001). *Application Delivery Strategies*. Obtenido de Meta Group: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

McKinsey. (2011, 05). *MGI_big_data_exec_summary.ashx*. Obtenido de McKinsey Global Institute: https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx

Microsoft. (23 de 04 de 2014). *Microsoft presenta la plataforma más completa para la gestión de Big Data*. Obtenido de Microsoft: <https://news.microsoft.com/es-es/2014/04/23/microsoft-estrategia-bigdata/>

Press G. (09 de 05 de 2013) *A very short history of Big data*, Obtenido de Revista Forbes: <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#69223fe65a18>

Prometeus. (20 de 09 de 2017). Obtenido de Prometeus Global Solutions: <https://prometeusgs.com/volumen-variedad-velocidad-veracidad-y-valor-las-5-dimensiones-del-big-data-la/>

soft, P. (s.f.). *SEIS RETOS EN LA IMPLEMENTACIÓN DEL BIG DATA*. Obtenido de personal soft:

<http://www.personalsoft.com/index.php/es/noticias/60-seis-retos-en-la-implementacion-del-big-data-2>

Otras referencias:

Manual sobre utilidades del big data para bienes públicos.

https://goberna.org/wp-content/uploads/sites/9/2017/09/Big_data.pdf