

BIG DATA - ANÁLISIS DE CASO EN LA CONTRALORÍA DE BOGOTÁ  
CON LA ENTREGA DE BONOS

LUIS ALBERTO GIRALDO POLANÍA – REPRESENTANTE DEL GRUPO

JOHN WILSON PARRA ORTIZ

YESID COTRINO GARCÍA

MANUEL FRANCISCO DULCE VANEGAS

JAVIER TAFUR DÍAZ

CONTRALORÍA DE BOGOTÁ D.C.

COLOMBIA

[lgiraldo@contraloriabogota.gov.co](mailto:lgiraldo@contraloriabogota.gov.co)

Tel (057) 3358888 extensión 10803

2018

## ÍNDICE

<b>1. CONSTANCIA</b>	
<b>2. RESUMEN</b>	
<b>3. INTRODUCCIÓN.....</b>	<b>1</b>
<b>4. OBJETIVOS DE LA INVESTIGACIÓN .....</b>	<b>3</b>
<b>5. REVISIÓN DE LITERATURA.....</b>	<b>3</b>
<b>6. CONCEPTUALIZACIÓN, DEFINICIÓN Y DIMENSIONES DE BIG DATA.....</b>	<b>5</b>
<b>7. TENDENCIAS RESPECTO A LA TECNOLOGÍA PARA LA ADMINISTRACIÓN DE DATOS.....</b>	<b>8</b>
<b>8. METODOLOGÍA Y PROCEDIMIENTOS MÍNIMOS QUE LAS ENTIDADES DE FISCALIZACIÓN SUPERIOR DEBERÍAN UTILIZAR PARA LA EVALUACIÓN Y ANÁLISIS DE BIG DATA.....</b>	<b>11</b>
<b>9. BENEFICIOS EN LA UTILIZACIÓN DEL BIG DATA – LIMITANTES A LOS QUE SE ENFRENTAN LAS EFS.....</b>	<b>14</b>
<b>10. VENTAJAS EN LA UTILIZACIÓN DE BIG DATA EN EL PROCESO DE AUDITORÍA.....</b>	<b>19</b>
<b>11. IDENTIFICACIÓN DE POSIBLES RIESGOS EN LA UTILIZACIÓN DE BIG DATA EN EL PROCESO AUDITOR Y ESTRATEGIAS PARA MINIMIZARLOS.....</b>	<b>22</b>
<b>12. CONCLUSIONES</b>	
<b>13. BIBLIOGRAFÍA O E-GRAFÍA</b>	

## 1. CONSTANCIA



*"Una Contraloría aliada con Bogotá"*

### LA DIRECTORA DE APOYO AL DESPACHO DE LA CONTRALORIA DE BOGOTÁ D.C.

#### CERTIFICA QUE:

La Contraloría de Bogotá D.C., con el Liderazgo del Despacho del Contralor Auxiliar, ha organizado un grupo de trabajo para presentarse en el Concurso "Ventajas de la utilización del Big Data" en el proceso auditor - Concurso Regional sobre Buena Gobernanza Edición 2018; el cual está conformado por los siguientes funcionarios:

NOMBRES Y APELLIDOS	CARGO DE DESEMPEÑO	CORREO ELECTRÓNICO	SECCIÓN/DEPENDENCIA
LUIS ALBERTO GIRALDO POLANÍA	SUBDIRECTOR TÉCNICO	lgiraldo@contraloriabogota.gov.co	Dirección de Planeación de la Contraloría de Bogotá D.C.
JAVIER TAFUR DÍAZ	PROFESIONAL ESPECIALIZADO	jtafur@contraloriabogota.gov.co	Dirección de Planeación de la Contraloría de Bogotá D.C.
MANUEL FRANCISCO DULCE VANEGAS	PROFESIONAL UNIVERSITARIO	mdulce@contraloriabogota.gov.co	Dirección de Planeación de la Contraloría de Bogotá D.C.
JOHN WILSON PARRA ORTIZ	CONTRATISTA	jparra@contraloriabogota.gov.co	Despacho del Contralor Auxiliar de la Contraloría de Bogotá D.C.
YESID COTRINO GARCÍA	CONTRATISTA	ycotrino@contraloriabogota.gov.co	Despacho del Contralor Auxiliar de la Contraloría de Bogotá D.C.

Atentamente,

  
**LINA RAQUEL RODRIGUEZ MEZA**

[www.contraloriabogota.gov.co](http://www.contraloriabogota.gov.co)

Cra. 32ª No. 26ª -10  
Código Postal 111321  
PBX: 335 88 88

## **2. RESUMEN**

Esta investigación cualitativa pretende describir un caso del uso de Big Data en el suministro de insumos al proceso de auditoría, el cual arrojó resultados como la consolidación de dos hallazgos fiscales por aproximadamente 72.000 USD, en razón al hecho de otorgar bonos a personas que no debían haber recibido esas ayudas.

La revisión de este caso permite aproximarse al tema, comprenderlo y con base en lo observado, proponer una metodología para la utilización del Big Data en las Entidades Fiscalizadoras. Para ello, se revisó bibliografía relacionada, se describió un caso específico donde se usó el Big Data en la Contraloría de Bogotá y se revisó el informe sobre los hallazgos detectados. Finalmente se analizaron las opciones de mejora en este proceso, para futuros casos que garanticen el éxito en la ejecución de las Auditorías.

### **3. INTRODUCCIÓN**

#### **3.1 Planteamiento del Problema**

En la era del conocimiento, el uso de herramientas tecnológicas para la generación de innovación es indispensable para la subsistencia de las organizaciones, ya sean públicas o privadas. Con el desarrollo de la informática, según Corbí (1992), cambió el concepto tradicional de producción el cual sólo se enfocaba en los bienes, ahora se enfoca hacia la generación de conocimiento, a través de innovaciones y de la ciencia y tecnología.

En este orden de ideas y teniendo en cuenta la velocidad como se crea y se dispone de información vía Web, se puede afirmar que ahora no es tan importante obtener la información como transformarla en algo útil, que genere valor para las organizaciones. Esto es especialmente importante para las entidades que ejercen el control fiscal y que su eficacia y eficiencia podrían favorecer el cumplimiento de los fines esenciales del Estado, con la efectiva provisión de bienes y servicios públicos, a través del mejoramiento de la gestión de las entidades sujetos de control. Como se analizará en capítulos posteriores, se puede decir que el Big Data tiene elementos en común como la gran cantidad de datos que al transformarlos y ordenarlos, se pueden convertir en información de interés para efectos del control fiscal. El Big Data hace referencia a estos altos volúmenes de información, pero también a la forma o metodologías para procesarlos, que sean útiles y generen valor en las organizaciones.

#### **3.2 Justificación**

En Colombia el tema del Big Data adquiere especial relevancia en abril del año 2018, a partir de la adopción de la política de explotación de datos hecha por el Consejo Nacional de Política Económica y Social (CONPES) mediante el documento 3920. Esto quiere decir, que se encuentra enmarcado dentro de una política de Estado. Además, en este documento se expresa que el Big Data aportará al cumplimiento de los Objetivos de Desarrollo Sostenible, concretamente a las metas relacionadas con el desarrollo de la innovación en los países, con un entorno normativo propicio a la adición de valor a los productos básicos; la promoción del

acceso público a la información; y el aumento significativo de datos oportunos, fiables y de alta calidad **(CONPES, 2018)**.

Este trabajo pretende mostrar un caso real donde se utilizó una gran cantidad de información a través del uso de ciertas herramientas, pues el informe que se desprendió de este ejercicio inicial fue un insumo importante para la configuración de dos hallazgos fiscales por más de 72.000 USD. En este caso, se decidió hacer un cruce de varias fuentes de información, entre ellas las bases de datos de beneficiados de asistencia social como son los bonos de alimentación, pero también se decidió analizar la base de datos de los individuos declarados como fallecidos y los que realmente recibieron estas ayudas. Esto, además que evidenciar fallas administrativas, también logró identificar el hecho que posiblemente se les están otorgando simultáneamente ayudas a las mismas personas, sin cumplir los requisitos y, de otro lado, que algunos individuos estarían recibiendo los bonos, sin encontrarse efectivamente registrados en la base de datos de beneficiados de la Entidad encargada de entregarlos.

Así mismo, se pretende describir cómo se utilizó el Big Data y también identificar los beneficios, limitaciones, ventajas y riesgos para el proceso auditor y para las Entidades Fiscalizadoras superiores. Finalmente, se genera una metodología o paso a paso fácilmente imitable, con el fin de explorar nuevas herramientas para ejercer el control fiscal.

### **3.3 Estructura**

Con el fin de desarrollar la temática planteada, a continuación se propondrá el objetivo general y los específicos; seguidamente, se enunciarán los principales libros y artículos consultados para el desarrollo de la investigación, luego de lo cual se hará una aproximación al marco teórico del tema Big Data a través de su conceptualización, definición y dimensiones, pero también de las tendencias respecto a la tecnología para la administración de datos.

Con base en lo analizado del caso estudiado, se identificarán los beneficios y limitantes en la utilización del Big Data para las Entidades de Fiscalización Superior, así como las ventajas y riesgos en el proceso de auditoría. Para culminar, teniendo en cuenta lo anteriormente observado, se propondrá la metodología y

procedimientos mínimos que las Entidades de Fiscalización Superior deberían utilizar para la evaluación y análisis del Big Data, y las conclusiones finales.

## **4. OBJETIVOS DE LA INVESTIGACIÓN**

### **4.1 Objetivo general**

Fortalecer el proceso auditor de las entidades fiscalizadoras, mediante la propuesta de una metodología en el uso del Big Data.

### **4.2 Objetivos específicos**

1. Describir el uso del Big Data en un caso práctico en la Contraloría de Bogotá.
2. Identificar los beneficios, limitantes, ventajas y riesgos en el proceso.
3. Proponer una metodología, según lo observado.

## **5. REVISIÓN DE LITERATURA**

Dentro de la bibliografía revisada para el desarrollo de este trabajo se tomó en cuenta literatura técnicamente robusta, pero también algunos artículos y lecturas que permitieran asimilar desde distintos puntos de vista el tema del Big Data. En el libro “Big data: Análisis de grandes volúmenes de datos en organizaciones”, el autor Luis Joyanes recopila aspectos técnicos y estratégicos requeridos en la implementación de un modelo organizacional que permita el uso del Big Data. Se describe bajo diferentes perspectivas el concepto de Big Data y sus dimensiones, enmarcado en la relevancia que las organizaciones le dan al tema. De igual forma, describe detalladamente todos los elementos para tener en cuenta en el manejo de grandes volúmenes de datos, ya sea el Big Data, Cloud Computing (computación en la nube), Internet of Things – IoT (Internet de las cosas) y el social media (medios sociales).

En contraste, el autor Bernard Marr, en su libro “Data Strategy. Cómo beneficiarse de un mundo de Big Data, Analytics e Internet de las cosas”, recopila los aspectos con un carácter más estratégico que técnico, referente a la implementación del Big Data en las organizaciones (públicas y privadas) desde la fase de identificación de los tipos de datos que son relevantes para la toma de decisiones, las metodologías para la recopilación, análisis y almacenamiento de estos y la necesidad de

conseguir competencias en materia de datos, para construir dentro de las organizaciones una cultura de los datos que sea la adecuada con el propósito de mejorar la toma de decisiones en las organizaciones. En este libro se pueden apreciar casos concretos de éxito en la adopción de la estrategia Big Data.

Otro estudio consultado fue “Big data Aquí y ahora 2015” de Chema Maroto, el cual hace referencia a las fortalezas y amenazas que las iniciativas Big Data han ocasionado en las organizaciones y en la necesidad de vincular esta tendencia dentro de su modelo de inversión, en respuesta a la creciente velocidad con la que se está adoptando el análisis de datos para conocer con mayor certeza las necesidades de sus clientes y aumentar la capacidad de respuesta a estas demandas. Llama la atención que en este trabajo se hace un análisis del modelo de gestión de datos tradicional y cómo ese modelo ha cambiado desde hace unos años con la incursión del Big Data; igualmente, identifica la contribución de las organizaciones del sector público al ecosistema Big Data, mediante la implementación de políticas y publicación de los datos abiertos, para facilitar no solo la transparencia sino la construcción de un tejido económico y de innovación.

Por otro lado, se encontró el artículo en inglés de Marcos Mazieri y Eduardo Dantas Soares titulado “Conceptualization and Theorization of the Big Data – International Journal of Innovation”, quienes a pesar de encontrar valioso el tema del Big Data, también aseguran que dentro la revisión teórica que realizaron hallaron que varios de los expertos que recomiendan la herramienta son más unos vendedores de software y hardware.

En el marco de la investigación también llamó la atención el hecho de encontrar un artículo de Big Data de investigadores en un país que está en proceso de transición del Comunismo a una sociedad democrática y libre como lo es Cuba, las autoras Yarina Amoroso Fernández y Dévorah Costales Ferrer identifican la importancia del tema para el ejercicio de lo público en su escrito “Big data: una herramienta para la administración pública”. Las autoras explican la relevancia que cobran los datos abiertos recopilados por el Gobierno y que a la luz del principio de transparencia permiten a la ciudadanía en general su uso como herramienta para mejorar la toma

de decisiones y de medidas para el fortalecimiento en el control al mismo Estado y de este último en la generación de políticas públicas, de acuerdo con las necesidades reales de la ciudadanía.

Por otro lado, se consultó el artículo de la Universidad Pedagógica y Tecnológica de Colombia, titulado “Conociendo el Big Data”, donde los autores Juan José Camargo Vega, Jonathan Felipe Camargo Ortega y Luis Joyanes Aguilar, hacen referencia al estado del arte respecto al Big Data y analizan las características, metodologías y herramientas tecnológicas. De igual manera, hacen un análisis de los modelos de datos que existen y de las tecnologías utilizadas para su análisis.

Finalmente, se revisó el libro del autor y periodista latinoamericano Andrés Oppenheimer llamado “Crear o Morir”, donde además de explorar todo lo relacionado con el tema de la innovación y desarrollo tecnológico en América Latina también hace referencia a la importancia del Big Data en el siglo XXI, en ámbitos diversos, desde lo económico hasta la salud. Su punto de vista se enfoca más hacia las perspectivas para la región de no alejarse de estas nuevas tendencias en innovación y tecnología.

## **6. CONCEPTUALIZACIÓN, DEFINICIÓN Y DIMENSIONES DE BIG DATA**

El creciente desarrollo tecnológico que a nivel global se ha evidenciado en los últimos 10 años, ha ocasionado la disminución en el coste de la tecnología, reducción en el tamaño de los dispositivos, aumento de la aplicabilidad en diferentes sectores y por consiguiente, accesibilidad a gran parte de la población. Este fenómeno va acompañado de un proporcional crecimiento en la generación de datos e información, a partir de gran variedad de fuentes tecnológicas, cambiando el concepto tradicional del manejo y análisis de los datos para la toma de decisiones, como lo menciona Bernard Marr, respecto a la subsistencia y progresos de las organizaciones, quien indica que con el crecimiento masivo del Big Data y el Internet of Things - IoT (Internet de las cosas), además de los métodos para analizar los datos, también aumentará la importancia de estos en cada aspecto de los negocios (Marr, 2018).

Si bien existen múltiples definiciones respecto al término de Big Data, varios autores coinciden en aspectos fundamentales, inherentes al Big Data:

1. Gran cantidad de datos o cantidades masivas de datos.
2. Técnicas, herramientas, procedimientos y metodologías para el procesamiento no convencional de estos datos.

Algunos autores mencionan que las dimensiones o características del Big Data son tres: Volumen, Velocidad y Variedad; sin embargo, se mencionan otras cuatro dimensiones en un concepto más amplio, como son: Veracidad, Valor, Variabilidad y Visualización (Maroto, 2015).

**Volumen:** Hace referencia a la cantidad masiva de datos que constantemente se generan en las diferentes fuentes como Redes Sociales, rastros de clics, Call Center, datos de georreferenciación, Internet de las cosas, dispositivos RFID (Identificación por Radiofrecuencia), consultas de navegación, entre otras. El crecimiento exorbitante de la generación de datos ha traído como consecuencia una serie de cambios en la industria, para recopilar y analizar dicha información. Dentro de las cifras encontradas acerca del crecimiento de los datos por minuto que llaman la atención se encuentran 4.1 millones de búsquedas en Google, 34.7 millones de mensajes instantáneos por Internet, 347.222 Tweets en twitter, 1.572.877 Gigas transferidas, 133.436 USD en ventas (Maroto, 2015).

**Variedad:** Los datos que constantemente se generan son de diferentes tamaños y formas, sin embargo, se han clasificado en tres clases:

**Datos Estructurados:** Éstos son datos que tienen una estructura bien especificada y los cuales a su vez se clasifican en estáticos (array, cadena de caracteres y registros) o dinámicos (listas, pilas, colas, árboles, archivos). Se afirma entonces, que estos formatos facilitan la administración de los datos (por ejemplo cuando se parametriza la fecha con cierto número de caracteres) (Camargo-Vega, Camargo-Ortega, & Joyanes-Aguilar, 2015).

**Datos semiestructurados:** Estos datos están estructurados por medio de etiquetas, que permiten separar los diferentes datos pero que no tienen un formato definido. Adicionalmente, puede presentarse que estos datos son incompletos o que

en sus componentes pueden cambiar de tipo, o contener datos cuya estructura difiere con la existente. Estos datos se generan cuando diferentes fuentes publican los datos a su manera.

**Datos no estructurados:** Son aquellos que no pueden normalizarse, no tienen ningún patrón u orden de manera relacional, es decir, que no tienen un formato determinado. Como ejemplo de este tipo de datos, se encuentran los archivos de audio, video, hojas electrónicas, archivos de texto, correos electrónicos, contenidos Web, entre otros. Respecto a estos, se puede hacer uso de los metadatos, que son datos para describir otros datos, similar a la estructura manejada en las bibliotecas mediante las fichas o sistema de información de los libros.

**Velocidad:** Hace referencia tanto a la rapidez de creación de los datos como de su procesamiento, almacenamiento y análisis. Como un par de ejemplos, en la plataforma Youtube se cargan aproximadamente 100 horas de video por minuto y en Twitter cada segundo se están generando aproximadamente 8.551 Tweets (Maroto, 2015).

**Variabilidad:** Esto se presenta cuando el significado de los datos cambia constantemente. Esta característica aplica cuando la recolección de datos se realiza mediante el procesamiento de lenguaje, ya que al recopilar opiniones, por ejemplo, el lanzamiento de un nuevo producto, cada expresión debe analizarse para revisar el contexto del sentimiento o de la opinión.

*“Este nuevo producto es genial”* es diferente a *“Esperé 30 minutos para probar el nuevo producto. Genial, ya se me arruinó el día”*. En ese contexto, la palabra genial no siempre tiene un sentido positivo.

**Valor:** Reside en el análisis de datos que son precisos, y el conocimiento que de allí se puede obtener. Por ejemplo, en temas de política pública, se puede alertar a los gobiernos cuándo hay más gente consultando por Internet sobre la gripe o sobre medicinas relacionadas con esta enfermedad. Con ello, se podría prevenir la generación de pandemias semanas antes que ocurrieran, focalizando sus esfuerzos para evitar la propagación del virus y así reducir el gasto en materia de salud pública, (Oppenheimer, 2014).

**Veracidad:** Está relacionada con el nivel de certeza y exactitud de los datos, ya que estos se vuelven inútiles cuando su información no es verídica, lo que cobra mayor relevancia cuando existen sistemas automatizados que recopilan los datos para la toma de decisiones o cuando se disponen de algoritmos que alimentan sistemas de aprendizaje automáticos.

**Visualización:** Se ha generado una necesidad apremiante, ya no solo de recopilar datos, sino de representar la información de una manera comprensible y facilitar así la toma de decisiones.

## **7. TENDENCIAS RESPECTO A LA TECNOLOGÍA PARA LA ADMINISTRACIÓN DE DATOS**

La administración de datos involucra en la actualidad grandes pilares sobre los que se sustentan las tecnologías: cloud (la nube), la movilidad (tecnologías, dispositivos y redes), social media (medios sociales), inteligencia artificial, Internet of Things - IoT (Internet de las cosas) y Big Data (grandes volúmenes de datos). La interdependencia de estos pilares está cambiando el paradigma de los negocios a nivel mundial, orientándolos a la toma de decisiones basada en los datos.

En cuanto a tecnología para negocios el **Cloud Computing** (Computación en la nube), es la tendencia más evidente y acogida actualmente por la mayoría de organizaciones, almacena todos los archivos e información en internet. Los servicios que ofrece son infraestructura, software y plataformas, presentando ventajas como costos económicos, toda vez que sólo se paga lo que se consume, reducción de la inversión en equipos, no hay instalaciones, mantenimiento ni soportes.

**La movilidad** define el ámbito cliente-empresa, donde la tendencia Bring Your Own Device BYOD (Lleve su propio dispositivo) permite a los trabajadores llevar a cabo en sus dispositivos portátiles como teléfonos inteligentes, tabletas, aplicaciones, etc., tareas del trabajo, además de estar conectados a la red y a los recursos corporativos. Uno de sus principales beneficios es el incremento de la productividad de los empleados, al sentirse más cómodos y satisfechos utilizando sus dispositivos, lo que para la organización representa ahorro en adquisición de tecnologías.

**Los medios sociales** se han convertido en uno de los principales canales de comunicación, difusión e intercambio de información; se refiere a integración de tecnologías, telecomunicaciones móviles e interacciones sociales como fotografías, imágenes, videos, audios, etc. Estos medios están creando un nuevo comportamiento en las empresas, cambiando la forma en que se relacionan con los clientes, esto se conoce como el “Social Business”. El video en directo, la realidad aumentada, la inmediatez en programas como Instagram y Snapchat, el WhatsApp Business, los influenciadores, entre otros, son la respuesta actualmente a las necesidades sociales en las redes (Joyanes, 2013).

**Inteligencia Artificial**, trata de aprovechar la tecnología disponible en búsqueda de conocimiento, donde agentes tecnológicos o robots desarrollan actividades y conductas propias de la condición humana. En las organizaciones ha evolucionado el Machine Learning (aprendizaje automático), donde el propósito es que las personas y las máquinas trabajen de la mano, al ser estas capaces de aprender como un humano lo haría y desempeñarse con precisión y exactitud. El Chatbots permite una mejor atención al cliente de forma automatizada, en el cual un robot es capaz de simular una conversación con una persona atendiendo sus necesidades.

**Internet de las cosas – IoT**, es la interconexión digital de objetos cotidianos con internet, a los cuales se les puede controlar de forma remota. Actualmente, el uso de vehículos autónomos en entornos controlados y los drones son un ejemplo.

**Big Data** se refiere a captura, transformación, procesamiento, análisis y visualización de grandes volúmenes de datos provenientes de diferentes fuentes y que se generan con rapidez, es una tendencia mundial y se le pronostica un gran crecimiento en el mercado y áreas de investigación asociadas (Mazieri & Dantas Soares, 2016). El manejo de este volumen de datos lleva a las empresas a realizar cambios en sus métodos y a tomar grandes decisiones en tiempo real, pueden crear nuevos productos y servicios, gracias a que conocen lo que sus clientes y usuarios quieren y necesitan, además de prever nuevos negocios, transformándose en más ágiles y competitivas (Joyanes, 2013).

Uno de los elementos más importantes en el Big Data es el uso de tecnologías de bases de datos para su análisis, como las siguientes:

1. Hadoop. La referencia para el tratamiento de los datos, que sirve como motor para gran parte de las aplicaciones basadas en Big Data. Es una biblioteca de software que permite el procesamiento distribuido de grandes volúmenes de datos a través de grupos de computadores que utilizan modelos de programación (Joyanes, 2013). Una característica de esta herramienta es que está diseñado para pasar de los servidores individuales a miles de máquinas.

2. Bases de Datos NoSQL. En 1998 aparece el término NoSQL, que significa no solo SQL. El nombre fue creado por Carlo Strozzi, para denominar su base de datos que no ofrecía SQL. Las NoSQL no presentan el modelo de las bases de datos relacionales; estas no tienen esquemas, no usan SQL, tampoco permiten joins o uniones, no almacenan datos en tablas de filas y columnas de manera uniforme, presentan escalabilidad de forma horizontal, para su labor usan la memoria principal del computador; su objetivo es gestionar grandes volúmenes de información (Camargo-Vega, Camargo-Ortega, & Joyanes-Aguilar, 2015). En estas bases de datos, la información se distribuye en varias máquinas, lo que permite obtener los datos con mayor velocidad en comparación con otras con modelo relacional.

3. MapReduce. Es un sistema basado en hilados para el procesamiento paralelo de grandes conjuntos de datos (Joyanes, 2013). MapReduce es un marco de software creado con el fin de hacer aplicaciones que puedan procesar grandes cantidades de datos de forma paralela, en un mismo hardware (Camargo-Vega, Camargo-Ortega, & Joyanes-Aguilar, 2015).

Otra herramienta útil en los entornos Big Data puede ser ZooKeeper, el cual es un servidor para la coordinación de aplicaciones distribuidas, así mismo HBase y Cassandra como bases de datos NoSQL, también se encuentra Pig y Hive, que facilitan capas de abstracción de nivel superior para datos y consultas; y Mahout como una librería de aprendizaje automático para el descubrimiento de información en los datos (Joyanes, 2013).

Con la expansión del Big Data se ha creado un nuevo concepto llamado Data Science (Ciencia de los datos) el cual hace referencia al tratamiento y manipulación de información que involucra métodos científicos con enfoques estadísticos e

investigativos, con el propósito de obtener un mejor entendimiento y conocimiento de los datos (creación de soluciones). Así, esta tendencia requerirá cada vez más a nivel mundial expertos en estos métodos, que se conocen como científicos de datos.

Para la implementación de Big Data se requiere tener en cuenta algunos pasos: entender el negocio y los datos, determinar los problemas y cómo los datos pueden ayudar, definir metas alcanzables, trabajar en paralelo con el sistema que esté funcionando y ser flexible con la tecnología y las herramientas (Joyanes, 2013).

El almacenamiento, procesamiento y correcto aprovechamiento de los datos se ha convertido en uno de los principales aliados de las organizaciones, estar desconectado de esta realidad es un costo muy alto, además de la pérdida de posicionamiento estratégico.

Finalmente, existen una serie de ventajas del Big Data como la integración de servicios de red, además que se integra con mayor facilidad y rapidez con el resto de aplicaciones empresariales, su implementación es rápida, las actualizaciones se realizan de manera automática, produce un flujo de información más eficiente y confiable, existe una permanente comunicación y análisis de resultados en tiempo real, la velocidad en la toma de decisiones es otra de las ventajas, bajos costos en infraestructura, prepara a la organización para enfrentar el futuro y puede mejorar la Competitividad (Joyanes, 2013).

## **8. METODOLOGÍA Y PROCEDIMIENTOS MÍNIMOS QUE LAS ENTIDADES DE FISCALIZACIÓN SUPERIOR DEBERÍAN UTILIZAR PARA LA EVALUACIÓN Y ANÁLISIS DE BIG DATA**

Con el fin de proponer una metodología viable para que las entidades de fiscalización adopten y utilicen el Big Data, se analiza un caso en la Contraloría de Bogotá, donde se detectan ciertas ideas para el mejoramiento de este proceso, las cuales se listan a continuación:

**8.1 Conformación de equipo.** El primer paso para el correcto uso de esta herramienta es la conformación de un equipo de trabajo para llevar a cabo una prueba piloto. Este además de estar bajo la supervisión y en cabeza de la Alta

Dirección, debe estar capacitado en el manejo de bases de datos y lo pueden conformar como mínimo dos personas. Si en la planta de personal no se encuentran personas capacitadas en el tema, se recomienda darles acceso a video tutoriales o capacitaciones virtuales en bases de datos.

**8.2 Definir los resultados esperados.** Es necesario que este equipo de trabajo tenga presente que el ejercicio que se hará tiene como propósito brindar insumos para la ejecución de auditorías. En este orden de ideas, se debe realizar un trabajo con grupos focales conformados por los auditores, ellos proporcionarán una idea de los temas principales y situaciones de interés por cada uno de los sectores y entidades.

Es importante el conocimiento del tema con el que cuentan los auditores, con el fin de llegar a conclusiones interesantes se deben establecer varios parámetros, sentencias o cruces de información, según la normatividad y la ley, por ejemplo, que se encuentren niños entre ciertas edades que no deberían estar recibiendo algún tipo de bono en particular o personas que estén recibiendo varios bonos a la vez.

**8.3 Recabar datos.** Una vez se tenga conocimiento del tema, se puede comenzar a recolectar datos. Para esto hay que explorar como primera medida las bases de datos abiertos de entidades del sector público y separar los archivos en carpetas por temas. Paralelamente, el equipo debe preparar los oficios dirigidos a las entidades priorizadas y se debe contar con una parte legal donde se recuerde la obligatoriedad de suministrar los datos, los plazos y las competencias, pero también una parte técnica donde se realicen recomendaciones y observaciones respecto a los principales errores detectados en el suministro de información, con el fin de minimizar el riesgo de la oportunidad en la obtención de la información y mejorar su calidad. Se recomienda diseñar una matriz de seguimiento a los requerimientos con las fechas y condiciones solicitadas.

**8.4 Priorizar y ordenar datos.** Cuando se cuente con los datos completos se deben ordenar los archivos en carpetas por temas o sectoriales. Según lo recolectado se debe clasificar los datos útiles y no útiles según los objetivos y resultados esperados ya definidos. Al ordenar los datos se debe detectar posibles inconsistencias y errores, verificar que se haya cumplido por parte de las entidades

con la forma y condiciones solicitadas, corregir algunos campos y demás procedimientos relacionados con la depuración de datos.

En este sentido se podría utilizar Excel o Access, sin embargo, al encontrar bases de datos no estructuradas, semiestructuradas o de grandes volúmenes, se recomienda utilizar el programa gratuito Hadoop, toda vez que cuenta con código abierto comunitario para el procesamiento de conjuntos de macrodatos, para su creación en forma local, se requiere el uso de servidores, configuración, adaptación y mantenimiento, es por ello que algunas compañías incluida Microsoft ofrecen Hadoop como servicio en la nube, permitiendo poner en marcha un cluster sin costo inicial alguno, sólo se paga por el almacenamiento y los servicios de proceso que se utilicen, cuando se utilicen (Joyanes, 2013). Igualmente, con este programa se puede poner en marcha un proceso, analizar los datos y apagarlo para detener el contador, además su integración con bases de datos como Sql, Access y Excel, permite flexibilización y rapidez.

#### **8.5 Transformar los datos en información y preparar los insumos finales.**

Después de la depuración y los cruces de datos, se pueden observar los temas priorizados en el análisis inicial con los expertos. En este punto se debe realizar un corto informe con lo detectado, según los cruces que se propusieron. Con las carpetas organizadas y los informes ejecutivos por sector o entidad, se remitirán estos insumos y se socializarán las conclusiones a los auditores.

Esto puede ser complementado con la visualización de los resultados a través de software de inteligencia de negocios como Power BI, que es una herramienta de Microsoft gratuita con una licencia básica, pero también existen otras herramientas sin costo como Qlik, que también ofrece elementos visuales interesantes.

**8.6 Custodiar la información.** Es importante una asesoría legal en materia de protección de datos personales, con el objeto de anonimizar bases de datos con información personal. Las bases además deben mantenerse en custodia informática, y en lo posible los informes y bases convertirlos en archivos no modificables o imagen.

Es importante guardar en la nube estos cruces y bases originales como posibles elementos probatorios para la configuración de los hallazgos fiscales.

**8.7 Otras herramientas tecnológicas.** Adicional a las ideas anteriores, existen puntualmente algunas otras herramientas o software para el proceso de adopción del Big Data, las cuales tienen diferentes facilidades y se pueden adaptar a cada una de las realidades de las Entidades Fiscalizadoras. Las aplicaciones open source (código abierto), a pesar de ser gratuitas, tienen la desventaja que su infraestructura, administración, seguridad e implementación van por cuenta de la organización. Algunas buenas aplicaciones que tienen constante trabajo por parte de la comunidad que las desarrolla son Apache Hadoop, Apache Storm, ElasticSearch, HPCC, Cassandra, MongoDB, RapidMiner y Talend (el open studio). Por otro lado, las herramientas de pago ofrecen escalabilidad, seguridad, administración y gestión de infraestructura en el caso de que sean cloud (en la nube). Las posibles opciones son Cloudera, HortonWorks, TeraData, Oracle BigData suite, IBM Watson, SAP Vora (se debe tener en cuenta que este utiliza HANA rdbms in-memory) y Sisense.

## **9. BENEFICIOS EN LA UTILIZACIÓN DEL BIG DATA – LIMITANTES A LOS QUE SE ENFRENTAN LAS EFS**

En este capítulo se describe el caso en la Contraloría de Bogotá, y se identifican los beneficios y las limitantes observadas en el desarrollo del ejercicio.

### **9.1 Caso en la Contraloría de Bogotá.**

La Contraloría de Bogotá tiene como misión vigilar la gestión fiscal de la Administración Distrital y de los particulares que manejan fondos o bienes públicos, con el fin de conseguir el mejoramiento de la calidad de vida de los ciudadanos del Distrito Capital. Este objetivo se desarrolla por medio de sus procesos misionales, Proceso de Vigilancia y Control a la Gestión Fiscal donde audita y controla a sus sujetos de control fiscal como es la Secretaría Distrital de Integración Social - SDIS. El presupuesto de esta Secretaría supera los 340 millones USD y es la encargada de ejecutar proyectos sociales en los aspectos de nutrición y calidad de vida en vejez, por medio de la entrega de bonos canjeables por alimentos y apoyos

económicos. Para el desarrollo del ejercicio auditor se utilizó como insumo el informe final, donde a través del Big Data se llegó a ciertas conclusiones sobre estos programas llamados “Bogotá te nutre” y “Envejecimiento digno, activo y feliz”, cuyos recursos superan los 126 millones USD.

La Contraloría de Bogotá realiza periódicamente auditorías a los diferentes contratos que realiza el Distrito para cumplir sus objetivos de desarrollo planteados en su Plan de desarrollo, audita la gestión fiscal, plan estratégico corporativo, gestión financiera y de contratación de la vigencia anterior, evaluación a estados contables, plan de mejoramiento, control interno por normatividad vigente y procedimientos asociados y sirven como insumo para la elaboración de los informes obligatorios y sectoriales.

En el ejercicio de la auditoría, se solicitaron las siguientes bases de datos:

- De beneficiarios de apoyos económicos y bonos canjeables por alimentos vigencia 2016-2017, de la Secretaría de Integración.
- De personas fallecidas en los últimos 12 meses en la ciudad de Bogotá durante el periodo comprendido entre enero de 2016 y septiembre de 2017, de la Secretaría Distrital de Salud.
- De personas fallecidas en los últimos doce meses a nivel nacional, que debe ser gestionada ante la Registraduría Nacional del Estado Civil.
- De personas únicas atendidas por vigencia, registrada en el sistema de información misional SIRBE, de la Secretaría de Integración.
- De beneficiarios de apoyos económicos y bonos canjeables por alimentos, con indicadores adicionales entre enero de 2016 y septiembre de 2017 de la Dirección de Análisis y Diseño Estratégico de la Secretaría de Integración.
  - De beneficiarios de apoyos económicos al adulto mayor.
  - De beneficiarios de bonos canjeables por alimentos personas en condición de discapacidad.
  - De beneficiarios de bonos canjeables por alimentos programa Creciendo en Familia.
  - De beneficiarios de bonos canjeables por alimentos personas en situación de emergencia social.

-De beneficiarios de bonos canjeables por alimentos proyecto Bogotá te Nutre.

## **9.2 Beneficios identificados en la auditoría.**

Los principales beneficios encontrados en la realización de la auditoría:

- Manejo de la información en base de datos, esto da alguna certeza sobre el inventario de los beneficiados de los programas.
- Verificación de reportes de personas que no pueden estar como beneficiarios, por ejemplo personas fallecidas, validado fechas de entrada al beneficio, fechas de atención y fecha de fallecimiento.
- Validación de beneficiarios que pueden encontrarse en varios programas, esto no debería presentarse en razón a que los individuos sólo pueden contar con un beneficio.
- Facilita la observancia de errores en los procedimientos y en la operación en la aplicación de las ayudas a los beneficiarios.
- Como resultado de las validaciones de esta información se pudieron configurar HALLAZGOS por beneficiarios que no pertenecen al programa, por beneficiarios que reciben doble beneficio y por beneficiarios fallecidos.
- Esto puede permitir el mejoramiento de la gestión del sujeto auditado y una asignación óptima del recurso público, fin último de las Entidades Fiscalizadoras.
- Optimización del tiempo, en razón a que se contaba con conclusiones importantes desde el escritorio.
- Un mejor proceso de planeación del trabajo en campo, dado que se contaba con bases sólidas para la ejecución de la auditoría.
- Mayor cobertura en la muestra, en razón a que se puede contar con el universo de información y de este extraer una muestra más amplia.

## **9.3 Limitantes.**

Según el ejercicio analizado se pueden identificar varias limitantes como el tiempo de respuesta del sujeto de control, la unificación de información en cuanto a su estructura, y la calidad de la misma, ya que en algunos casos se presentaba información incompleta o inexacta y el limitado porcentaje de evaluación de

información, debido a la baja calidad de la información encontrando la necesidad de desechar la información.

### **9.3.1 Información no homogénea.**

Con el fin de contar con información completa se requirió a las entidades columnas similares, sin embargo, al momento de analizar la estructura de las bases, se evidenciaban falencias como campos incompletos, en diferente orden o sin diligenciar. Esto demora el proceso de consolidación, depuración y análisis de la información.

### **9.3.2 Tiempo de respuesta del sujeto de control.**

La Secretaría de Integración Social entregó satisfactoriamente la información en un tiempo aproximado de dos meses, este tiempo es considerable, según el plazo establecido para la etapa de planeación de una Auditoría. Estos días se requirieron en razón a que la información no se encontraba correctamente consolidada en los aplicativos de la entidad, entonces fue necesario construir la información a través de varias subdirecciones, cada una con una base de datos y estructura diferente.

### **9.3.3 Volumen de la información analizada.**

Debido a la calidad de la información y con el fin de no alterar las conclusiones a las que se podrían llegar se optó por prescindir de aproximadamente el 30% de los registros, esto quiere decir, que las conclusiones se realizaron sobre el 70% de los datos suministrados.

### **9.3.4 Capacidad de análisis de la información.**

Una limitante que se identificó es que al tratarse de bonos que se otorgan con una frecuencia alta, las bases de datos sufren varios cambios en razón al aumento constante de los beneficiados, además por las situaciones que se puedan presentar y novedades, lo cual dificulta la capacidad de análisis de la información.

### **9.3.5 Limitantes Tecnológicas Software.**

Una de las principales desventajas encontradas hace referencia a la capacidad tecnológica de la entidad, en lo que tiene que ver con el software, esta Entidad

Fiscalizadora cuenta con licencia sobre los productos de Microsoft Office, con esta licencia se puede trabajar en archivo en formato Excel, pero eventualmente en Access, sin embargo, para este último programa fue necesario un permiso especial de administrador, debido a las escasas licencias para instalar dicho programa en la entidad.

A pesar de que se dispone de estas herramientas, estos programas cuentan con limitantes que a continuación se explicarán, razón por la cual se descartó la posibilidad de realizar un cruce de información con las bases de datos del SISBEN (Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales), cuyas conclusiones en este análisis hubiesen podido ser mucho más profundas.

Igualmente, debido a estas limitaciones el proceso de limpieza o depuración de información también tomó unos catorce días, tiempo en el cual se hubiese podido realizar algunas otras verificaciones para enriquecer el producto final.

#### **Desventajas de Excel.**

Microsoft Excel se utiliza como una hoja de cálculo que al ser utilizada como base de datos, presenta algunas falencias como generar archivos pesados con poca información, consume importantes recursos de máquina o hardware, es poco seguro, ya que puede abrirse y modificarse fácilmente, se vuelve inestable presentando errores cuando se cierra inesperadamente y provocando pérdida de información, corre solo en sistemas operativos Windows y Mac, adicionalmente, limita el número de registros, ya que no se puede tener más de 1.048.576 registros, así mismo, limita en los datos tipos numéricos y en el alfabético, toda vez que acepta hasta ciertos números de caracteres.

#### **Desventajas de Access.**

Dentro de las desventajas de la Access se encuentra que limita un tamaño máximo de 2 Gb. Access es todavía una base de datos insegura, no permite algunas sentencias de SQL, una vez abierta cualquier persona puede modificar la estructura y la información, cuando alcanza la capacidad máxima de almacenamiento presenta errores inesperados, ocasionando pérdida de información, al ser utilizada vía web presenta limitaciones, ya que permite hasta cierto número de usuarios concurrentes,

limitaciones en los sistemas operativos, ya que sólo funciona en Windows; por último, es débil en cuanto al diseño relacional (Entidad - Relación).

#### **Desventajas de SQL server.**

SQL Server, al igual que los dos productos anteriores, fue desarrollado por Microsoft, dentro de las limitaciones de Sql Server, que ocupa importantes recursos de máquina especialmente memoria RAM, procesadores y espacio en disco; en cuanto a la paginación se puede decir que es pequeño comparado con otro motor de base de datos, no funciona igual en los diferentes sistemas operativos.

#### **9.3.6 Limitantes en el Talento Humano.**

En la composición de la planta de personal de la Contraloría de Bogotá se identifica que la mayoría de los profesionales vinculados son expertos en temas de derecho administrativo, contables y financieros y provienen de áreas del conocimiento relacionadas a las ciencias sociales, sin embargo, hay escasez de personal con amplios conocimientos en sistemas, análisis de datos y programación. Debido a este aspecto, los encargados de realizar todo el cruce y análisis de información recolectada por la entidad en el caso estudiado, fueron dos ingenieros industriales.

## **10. VENTAJAS EN LA UTILIZACIÓN DE BIG DATA EN EL PROCESO DE AUDITORÍA**

A continuación se identifican las ventajas en la utilización de Big Data en el proceso de auditoría, a la luz del caso estudiado en esta investigación. Dichas ventajas encierran temas como la posibilidad de consolidar hallazgos sólidos, facilitar la labor de los auditores en campo, mejorar la gestión de las entidades sujetos de control y la ejecución de políticas públicas; transparencia y oportunidad en la ejecución de las auditorías y finalmente, se podrían explorar nuevas líneas de investigación hasta ahora aisladas por parte de las Entidades Fiscalizadoras.

### **10.1 Configuración de hallazgos fiscales.**

La principal ventaja que se presentó en la realización del proceso auditor y validación de la información encontrada en las bases de datos, fue la de tener un insumo sólido técnicamente y comprobado, el cual permitió la configuración de algunos hallazgos que se mencionarán a continuación:

**Hallazgo fiscal relacionado con el Big Data No. 1 por aproximadamente 58.000 USD.** Con la información entregada por la SDIS, se procedió a realizar un cruce de las personas que cambiaron los bonos por alimentos en dos supermercados llamados Colsubsidio y Cencosud, versus los beneficiarios de esta secretaría, encontrándose inconsistencias importantes como la identificación de 1.324 personas que recibieron y canjearon bonos, pero No registran como beneficiarios de la entidad.

Es de aclarar que para poder realizar la comparación de las dos bases de datos, se tuvo que unificar la información de los dos almacenes y se extrajo la lista de los beneficiarios que mes a mes cambiaron los bonos.

**Hallazgo fiscal relacionado con el Big Data No. 2 por aproximadamente 14.000 USD.** Se verificó mediante el cruce de información en la ejecución, en lo relacionado a la entrega de bonos, donde estos fueron entregados a un mismo beneficiario en un mismo periodo, lo cual se conoce como simultaneidad de bonos. Una vez analizada esta información, se evidenciaron 333 casos en los cuales el beneficiario recibió y canjeó bonos en el mismo periodo simultáneamente, inobservando la resolución donde se incluyen las condiciones para recibir ciertos tipos de bonos. El hallazgo surge de la comparación de la lista de los beneficiarios, las fechas y los bonos canjeados simultáneamente.

## **10.2 Mejoramiento de la gestión del Sujeto de Control.**

En este caso, se realizaron observaciones de mejora al sujeto de control con relación a su procedimiento en la entrega de bonos. Esto significa una mejor utilización de los recursos que se invierten en la realización de los proyectos a través del ejercicio del proceso auditor, lo cual a su vez influye en la mejora de las políticas sociales. Esto podría aunarse con la participación ciudadana y la lógica del Gobierno Abierto, con el fin de que la misma ciudadanía identifique tempranamente riesgos en la ejecución de políticas públicas (Amoroso Fernández & Costales Ferrer, 2016). A continuación se explican los dos hallazgos administrativos que pueden permitir el mejoramiento de la gestión de la entidad en la inversión del recurso público:

**Hallazgo No. 1 Administrativo.** En el análisis realizado a la ejecución del contrato, se verificó que las personas que recibían y canjearon los bonos se encontraron con

vida, razón por la cual el equipo auditor comprobó que una persona hizo canje de un bono un año después de fallecida. La persona fallece en enero de 2016 y la auditoría constató que cobró el bono en el mes de marzo de 2017.

**Hallazgo No. 2 Administrativo.** La Secretaría Distrital de Integración Social entregó la información de los beneficiarios, tales como: número del documento de identificación, tipo de bono, fecha en que fue canjeado y los productos que adquirieron al precio vigente a la fecha de cambio. Mediante el cruce de esta información se permitió verificar la variación en los precios de los productos preferiblemente no perecederos durante el mismo mes, encontrando productos que presentan una diferencia considerable en el precio entre el 15% y el 60%; se escogieron productos en cada uno de los almacenes.

### **10.3 Proceso auditor transparente, rápido y oportuno.**

El manejo de información en base de datos permite la realización del proceso auditor de una forma más transparente y rápida. Esto generaría confianza en el principal cliente del Estado, es decir el Ciudadano. Además, por la capacidad de analizar un mayor volumen de datos se podría alcanzar una mayor cobertura a las muestras a auditar.

En este orden de ideas, al reunir un importante volumen de datos y a través del ejercicio de depuración, estos datos se transforman en información que al analizarla puede generar conocimiento, lo cual, además de permitir comprender algunos fenómenos, también abre la posibilidad de predecirlos o detectar oportunamente fallas, esto facilita el hecho de anticiparse a posibles situaciones (Mazieri & Dantas Soares, 2016), como las que puedan afectar la ejecución de recursos por medio de fraudes al sistema de seguridad social o ayudas otorgadas a individuos fallecidos, entre otras.

### **10.4 Enfoques en las auditorías.**

El uso del Big Data permite la comprensión de fenómenos inexplorados, por ejemplo grandes cruces de datos con beneficiados en los programas sociales, no sólo en el nivel subnacional sino también nacional. Esto a su vez generaría nuevas líneas de negocio o de investigación para el proceso auditor, como la efectividad de estos

programas de asistencia social, el impacto que generan y su efectividad en la lucha contra la pobreza, por ejemplo.

Al contar con un ejercicio del Big Data enfocado a temas o sectores, se podrían generar insumos robustos donde se identifiquen situaciones que valdría la pena revisar en el marco de la ejecución de las auditorías en campo. Es decir, el Big Data significa una herramienta indispensable para el proceso de planeación de las auditorías de las Entidades Fiscalizadoras.

### **10.5 Beneficios del control fiscal.**

Una de las formas de impacto de la gestión en el proceso auditor es la generación de beneficios del control fiscal, que pueden ser cuantificables, cuando existe una efectiva recuperación del recurso público o ahorro, o cualificables, que tienen que ver con los beneficios en mejoras en la gestión de las entidades auditadas.

## **11. IDENTIFICACIÓN DE POSIBLES RIESGOS EN LA UTILIZACIÓN DE BIG DATA EN EL PROCESO AUDITOR Y ESTRATEGIAS PARA MINIMIZARLOS**

Los riesgos identificados en el proceso investigado se encuentran asociados a sesgos en la consolidación del resultado en la información, cuando se cuente con baja calidad en la información recibida, también, se encuentra un riesgo relacionado con los costos de oportunidad vinculados a demoras en el proceso de entrega de la información y finalmente, un riesgo con la protección de la información suministrada y procesada.

### **11.1 Riesgo de sesgos en las conclusiones, en razón a la calidad de la información.**

En ocasiones se identificaba información incompleta, por ejemplo, nombre sin apellidos, número de tarjetas de identidad sin nombres asociados, o campos vacíos. Así mismo, casos donde por error humano se duplicaban los registros, es decir, se identificaron errores en la incorporación de la información por parte de los generadores de la misma, ya que se encontraron faltantes, información incompleta o duplicada. También se observaron falencias en la estructura de la información, la cual no era homogénea en los diferentes campos, esto demoró el análisis de la misma.

En razón a estas situaciones la conclusión final del informe se realizó sobre el 70% de los registros. A pesar que inicialmente se requirió a la entidad para el arreglo de ciertos errores encontrados, también fue necesario un proceso de depuración de información, de lo contrario las observaciones y conclusiones finales tendrían algún tipo de sesgo y esto podría afectar la solidez del hallazgo final.

**Forma de mitigación:** la forma más oportuna para mitigar el riesgo en el proceso de análisis de información es el conocimiento de los procesos que generan la información solicitada. Igualmente, se debe tener en cuenta cuál puede ser el resultado esperado y tener claro que el proceso de análisis aporta insumos para la toma de decisiones. Respecto a las tecnologías para mitigar riesgos sobre la calidad de información, se recomienda realizar un análisis de cuál se dispone, verificar su capacidad de procesamiento, compatibilidad, protección, etc., antes de comenzar a realizar inversiones significativas.

Desde el punto de vista del ejercicio auditor, es importante que los equipos que adelantan las auditorías consignen observaciones en los informes de las entidades, con el fin de manifestar la necesidad de incorporar mejoras al manejo de la información, haciendo énfasis en los temas antes relacionados y que afectan el análisis o crean riesgos.

### **11.2 Costos de oportunidad asociados a demoras en el suministro de Información.**

Por otro lado, uno de los principales riesgos identificados es la oportunidad en el suministro de información, esto se evidenció cuando se intentó consolidar la información que debería haber estado agrupada en el aplicativo SIRBE de la Secretaría Distrital de Integración Social, sin embargo, la información se encontraba incompleta y se observaron falencias en la actualización. Adicionalmente, al no contar con un administrador de bases de datos en esa entidad, fue necesario construir la información a través de varias subdirecciones, cada una con una base de datos y estructura diferente. El tiempo de espera para recibir la información fue de dos meses, este periodo es considerable, si se tiene en cuenta que para la planificación de una auditoría hay ciertos días estipulados.

Adicionalmente, para el suministro de las bases de datos de la Registraduría y Medicina Legal, las entidades se demoraron cinco días hábiles. Para el proceso de depuración de la información al interior de la Contraloría de Bogotá se tomaron dos semanas, y se usó el programa Access de Office.

**Forma de mitigación:** Con el fin de incentivar la entrega oportuna de información y la calidad de la misma, las Entidades Fiscalizadoras deben realizar primero una investigación de los posibles insumos que deben solicitar y en qué forma se pueden suministrar, se recomienda enviar un oficio donde se recuerde la obligatoriedad en el suministro de información, también el tiempo de respuesta o entrega de la información, asimismo las condiciones, el tipo de archivo y su forma física de transmisión al equipo auditor.

### **11.3 Riesgos en la protección de datos.**

Un riesgo común en el uso de Big Data es la protección de datos, se hace énfasis que los datos personales son responsabilidad de la organización que los posee, y esta debe garantizar su protección, vinculando técnicas que le permitan anonimizar la información tales como el encriptado, detección de intrusiones y procesos de capacitación a empleados para que no revelen esta información (Marr, 2018).

La pérdida de información puede llevar a costos legales y compensaciones financieras, pasando por los daños en la reputación organizacional. El gran volumen de datos que se genera, se almacena y se procesa trae como consecuencia un interés malverso de los delincuentes informáticos, por filtrarse y robar dichos datos. Ejemplo de ello los robos masivos de información en Tumblr en 2013 y Myspace en 2016, pues para que un hacker pueda tener acceso a esta información, debe tener conocimientos superiores de los programadores y contar con la capacidad de infraestructura tecnológica para almacenar y compartir estos datos de manera anónima (Marr, 2018).

En general, es necesario no atentar contra la intimidad y la privacidad de las personas, toda vez que un importante volumen de información contiene datos personales (Amoroso Fernández & Costales Ferrer, 2016) entonces las Entidades Fiscalizadoras deben evitar un mal uso de dichos datos.

**Forma de mitigación:** La responsabilidad del manejo de la información por parte de las Entidades Fiscalizadoras es la custodia y utilización de la información para su objetivo principal que es el control fiscal, por contener información vulnerable como es la de género, salud, religiosa o financiera. Para lo cual se deben tener las mejores herramientas o actualizadas para la protección de la información, las personas que manipulan, verifican, o utilizan la información sólo lo harán en ejercicio del control fiscal.

Al terminar el proceso auditor, se aconseja que las bases de datos originales se mantengan en custodia informática, y los resultados convertirlos en archivos no modificables o de imagen, para así no ser utilizadas en actividades diferentes a las de control fiscal. También es importante en lo posible anonimizar bases de datos con información personal.

## **12. CONCLUSIONES.**

El Big Data puede convertirse en una herramienta indispensable en el marco de la sociedad del conocimiento, toda vez que ofrece la posibilidad de analizar y depurar un gran volumen de datos, los cuales una vez organizados se convierten en información valiosa, que permite conocer fenómenos y eventualmente predecirlos para tomar decisiones en el ámbito de las competencias y prevenir o detectar situaciones que puedan afectar la correcta ejecución de los recursos públicos, además que también brinda insumos importantes para la ejecución de las auditorías.

Se recomienda comenzar con pruebas piloto en cabeza de la Alta Dirección, para explorar la utilidad de la herramienta. En este sentido, la revisión de casos de éxito es importante en la identificación de ventajas, limitantes, beneficios y riesgos en el uso del Big Data para las Entidades Fiscalizadoras.

La metodología que se propone una vez revisado el caso de éxito en la Contraloría de Bogotá, tiene como primer paso la conformación de un equipo de trabajo para una prueba piloto. También es necesario el conocimiento del tema por parte de los auditores, quienes harán recomendaciones y definirán ciertos parámetros al momento del análisis de la información, esto con el fin de definir los resultados esperados. Luego, es necesario recabar los datos a través de oficios a las Entidades auditadas y con datos abiertos de las Entidades públicas. Una vez se cuente con los datos, es necesario depurarlos y priorizarlos según los temas definidos para después transformarlos en información, que será el insumo remitido a los auditores con los soportes y el informe ejecutivo. Finalmente, es necesario custodiar correctamente la información sensible, sobretodo la que contenga datos personales.

### 13. BIBLIOGRAFÍA O E-GRAFÍA

- Amoroso Fernández, Y., & Costales Ferrer, D. (2016). Big Data: una herramienta para la administración pública. *Ciencias de la Información*, 3-8.
- Camargo-Vega, J. J., Camargo-Ortega, J. F., & Joyanes-Aguilar, L. (2015). Conociendo Big Data. *Facultad de Ingeniería - Universidad Pedagógica y Tecnológica de Colombia*, 63-77.
- CONPES, C. N. (17 de Abril de 2018). *Departamento Nacional de Planeación*. Obtenido de <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf>
- Corbí, M. (1992). *Proyectar la sociedad-reconvertir la religión. Los nuevos ciudadanos*. Barcelona: Herder.
- Joyanes, L. (2013). *Big Data. Análisis de grandes volúmenes de datos en organizaciones*. México D.F.: Alfaomega grupo Editor, S.A. de C.V.
- Maroto, C. (2015). *Big Data Aquí y ahora 2015*. Obtenido de Chema Maroto: <http://chemamaroto.com/home/wp-content/uploads/2018/02/Investigaci%C3%B3n-chema-maroto-Big-Data-2015.pdf>
- Marr, B. (2018). *Data Strategy. Cómo beneficiarse de un mundo de Big Data, Analytics e Internet de las Cosa*. Bogotá: Ecoe Ediciones.
- Mazieri, M., & Dantas Soares, E. (2016). Conceptualization and theorization of the Big Data. *International Journal of Innovation*, 23-41.
- Oppenheimer, A. (2014). *!Crear o morir!* México DF: Penguin Random House Grupo Editorial.