UNIVERSIDAD INTERNACIONAL SEK

FACULTAD DE ARQUITECTURA E INGENIERÍAS

Trabajo de fin de carrera titulado:

"Diseño de una metodología de auditoría a la seguridad del proceso de ciencia de datos en una entidad financiera privada del Ecuador"

Realizado por:

Ing. Verónica Alexandra Ramírez Tenecela

Director del proyecto:

Ing. Luis Fabián Hurtado Vargas, Mgs.

Como requisito para la obtención del título de:

MASTER EN TECNOLOGÍAS DE LA INFORMACIÓN CON MENCIÓN EN SEGURIDAD EN REDES Y COMUNICACIÓN

Quito, marzo 2019

DECLARATORIA

El presente trabajo de investigación titulado:

"DISEÑO DE UNA METODOLOGÍA DE AUDITORÍA A LA SEGURIDAD DEL PROCESO DE CIENCIA DE DATOS EN UNA ENTIDAD FINANCIERA PRIVADA DEL ECUADOR"

Realizado por:

VERÓNICA ALEXANDRA RAMÍREZ TENECELA

ha sido dirigido por el docente:

ING. LUIS FABIÁN HURTADO, MGS.

quien considera que constituye un trabajo original de su autor

Ing. Luis Fabián Hurtado Vargas, Mgs.

DIRECTOR

LOS PROFESORES INFORMANTES

Los Profesores Informantes:

Ing. DIEGO FERNANDO RIOFRIO LUZCANO, PhD

Msc. CHRISTIAN DAVID PAZMIÑO FLORES

Después de revisar el trabajo presentado,

lo han calificado como apto para su defensa oral ante el tribunal examinador

Ing. Diego Riofrio, PhD

Msc. Christian Pazmiño

Quito, marzo del 2019

DECLARACIÓN JURAMENTADA

Yo, VERÓNICA ALEXANDRA RAMÍREZ TENECELA, con cédula de identidad 1719894246, declara bajo juramento que el trabajo aquí desarrollado es de mi autoría, que no ha sido previamente presentado para ningún grado a calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración, cedo los derechos de propiedad intelectual correspondientes a este trabajo, a la UNIVERSIDAD INTERNACIONAL SEK, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normativa institucional vigente

_____

Verónica Alexandra Ramírez Tenecela

C.C. 1719894246

AGRADECIMIENTOS

Quiero agradecer en primer lugar a Dios, por darme la fortaleza y permitirme lograr esta meta.

A mi esposo, Fernando por ser mi compañero de vida y estudios, por siempre motivarme a ser una mejor versión de mí y por su apoyo incondicional.

A mi madre, por su fe en mí, por los sacrificios que realiza todos los días y por ser una bendición en nuestras vidas.

A mi papá, hermanos, cuñados y sobrinos por sus ánimos, por el apoyo y consejo en el transcurso de la maestría.

A mi director de tesis, el Ing. Fabian Hurtado por su orientación y ayuda en el presente trabajo.

A mis amigos y compañeros de maestría Anita, Andrés y Jean Pierre por las experiencias compartidas en clase que hicieron más enriquecedora esta etapa de mi vida.

DEDICATORIA

Este proyecto va dedicado a mi madre por ser el ejemplo de constancia y persistencia que me han formado y me han permitido culminar esta etapa.

A mi esposo Fernando y a mi hija Samantha por ser los motores de mi vida y mi inspiración para cada objetivo que me propongo.

Tabla de Contenido

CAPÍTULO I. INTRODUCCIÓN 1
1.1. El problema de la Investigación 1
1.1.1. Planteamiento del problema 1
1.1.2. Formulación del problema 2
1.2. Objetivos 3
1.2.1. Objetivo general 3
1.2.2. Objetivos específicos 3
1.3. Justificación 4
CAPÍTULO II. MARCO TEÓRICO 5
2.1. Ciencia de datos 5
2.1.1. Definición de ciencia de datos 5
2.1.2. Procesos de ciencia de datos 7
2.1.3. Áreas de conocimiento de ciencia de datos 8
2.1.4. Diferencia de terminologías 10
2.1.5. Arquitectura 11
2.1.6. Evolución de ciencia de datos 15
2.1.7. La ciencia de datos y las empresas 15
2.1.8. Uso de ciencia de datos en entidades financieras 16
2.1.9. Ventajas de usar ciencia de datos 18
2.1.10. Implementación de un proceso de ciencia de datos 18
2.2. Auditoría 20
2.2.1. Definición de Auditoría 20
2.2.2. Definición de Auditoría de Sistemas 20
2.2.3. Fases de Auditoría 20
2.2.4. Desarrollo de Programas de Auditoría 21
2.3. Metodologías 23
2.3.1. Certificación CISA 23
2.3.2. NIST 23
2.3.3. COBIT 25
2.3.4. Normas para instituciones financieras 29

2.3.5. Estado del arte 29
2.4. Marco conceptual 34
2.4.1. Adopción de una perspectiva teórica 35
CAPÍTULO III. ANÁLISIS SITUACIONAL 37
3.1. Uso de ciencia de datos. 37
3.1. Metodologías existentes 38
3.1.1. Metodologías para ciencia de datos 38
3.1.2. Metodologías para Auditoría 40
3.1.3. Mapeo de metodologías 40
CAPÍTULO IV. PROPUESTA 53
4.1. Objetivo de la revisión 53
4.2. Alcance de la revisión 54
4.3. Creación de plan de auditoría 54
4.3.1. Determinar la materia de la auditoría 55
4.3.1.1. Área de ciencia de datos. 56
4.3.1.2. Flujo de información 57
4.3.2. Realizar la planificación previa a la auditoría 61
4.3.3. Determinar los pasos para la recolección de datos 64
4.3.3.1. Dominio de Gobierno de Datos 64
4.3.3.2. Dominio de Selección y capacitación de personal 68
4.3.3.3. Dominio de Seguridad de infraestructura 72
4.3.3.4. Dominio de Seguridad lógica 76
4.3.3.5. Dominio de Gestión de requerimientos de desarrollo 81
4.3.3.6. Dominio de Gestión de operaciones 85
4.3.3.7. Dominio de Controles de aplicación 89
CAPÍTULO V. CONCLUSIONES Y TRABAJOS FUTUROS 93
5.1. Conclusiones 93
5.2. Trabajos futuros 94
BIBLIOGRAFÍA 96

Índice de tablas

Tabla 1 Dominios para revisiones de Auditoría Interna 41
Tabla 2 Mapeo dominios internos y GTAG 42
Tabla 3 Selección de procesos COBIT 43
Tabla 4 Mapeo procesos internos, GTAG y COBIT. 45
Tabla 5 Selección de procesos Normativa de la Superintendencia No. SB-2018-771 46
Tabla 6 Mapeo procesos internos, GTAG, COBIT y Normativa 48
Tabla 7 Selección de componentes del Framework de Arquitectura de BIG DATA propuesto por la NIST 49
Tabla 8 Mapeo procesos internos, GTAG, COBIT, Normativa y NIST 50
Tabla 9 Áreas involucradas en la revisión - Temas a tratar 55
Tabla 10 Riesgos identificados con GTAG 61
Tabla 11 Riesgos identificados con COBIT5 62
Tabla 12 Riesgos identificados con NIST 63
Tabla 13 Dominio de Gobierno de Datos - Generalidades 64
Tabla 14 Dominio de Gobierno de Datos - Controles a evaluar 66
Tabla 15 Dominio de Selección y capacitación de personal - Generalidades 68
Tabla 16 Dominio de Gobierno de Datos - Controles a evaluar 70
Tabla 17 Dominio de Seguridad de infraestructura - Generalidades 72
Tabla 18 Dominio de Seguridad de infraestructura - Controles a evaluar 73
Tabla 19 Dominio de Seguridad de Seguridad Lógica - Generalidades 76
Tabla 20 Dominio de Seguridad de Seguridad Lógica - Controles a evaluar 79
Tabla 21 Dominio de Gestión de requerimientos de desarrollo - Generalidades 81
Tabla 22 Dominio de Gestión de requerimientos de desarrollo - Controles a evaluar 83
Tabla 23 Dominio de Gestión de operaciones - Generalidades 85
Tabla 24 Dominio de Gestión de operaciones - Controles a evaluar 87
Tabla 25 Dominio de Controles de aplicación - Generalidades 90
Tabla 26 Dominio de Controles de aplicación - Controles a evaluar 91

Índice de figuras

Figura 1. Procesos de ciencia de datos 7
Figura 2. Diagrama de Venn "Científico de Datos" 8
Figura 3. Arquitectura Big Data.. 12
Figura 4. Revolución tecnológica 17
Figura 5. Metodología para la ciencia de datos 18
Figura 6. Programas de Auditoría 22
Figura 7. Habilitadores de COBIT5 26
Figura 8. Principios de COBIT5 26
Figura 9. Perfiles en la gestión de riesgo bancario 32
Figura 10. Metodología de Auditoría para evaluar el proceso de ciencia de datos 40
Figura 11. Flujo de información de ciencia de datos 58
Figura 12 Dominio de Gobierno de Datos - Ciclo del dominio 68
Figura 13. Dominio de Selección y capacitación del personal - Ciclo del dominio 71
Figura 14. Dominio de Seguridad de infraestructura - Ciclo del dominio 76
Figura 15. Dominio de Seguridad lógica - Ciclo del dominio. Fuente: Elaborado por autor 80
Figura 16. Dominio de Gestión de requerimientos de desarrollo - Ciclo del dominio 84
Figura 17. Dominio de Gestión de operaciones - Ciclo del dominio 89
Figura 18. Dominio de Controles de aplicación - Ciclo del dominior 92

RESUMEN

Durante años, el análisis de información ha sido una de las actividades más importantes y necesarias dentro de las instituciones financieras, ya que permite conocer datos en torno a los productos y servicios ofertados, el manejo de la cartera de clientes y créditos, entre otra información relevante. Sin embargo, debido a la creciente evolución de las tendencias tecnológicas, la analítica de datos ha evolucionado a una disciplina que mejora el análisis descriptivo que se venía realizando, la cual además hace posible la predicción de situaciones futuras basada en detección de patrones en los datos, esta nueva disciplina se denomina ciencia de datos.

La ciencia de datos integra disciplinas como la estadística, la matemática y la programación, además utiliza técnicas como el aprendizaje automático, para obtener información de valor y garantizar la entrega de resultados rápidos, íntegros y veraces. Por otro lado, con ayuda de la ciencia de datos se puede procesar gran cantidad de información de diferentes tipos de datos de forma eficiente por lo que también está estrechamente relacionado con el concepto de *Big Data*.

Implementar el proceso de ciencia de datos permite crear estrategias empresariales, reconocer nuevas oportunidades de mercado y alcanzar ventajas competitivas.

Si bien, este proceso surge en las instituciones financieras para transformar la información en conocimiento y aportar algunos beneficios, su implementación conlleva riesgos que deben ser identificados, controlados y mitigados. El presente trabajo está enfocado en proponer una metodología de auditoría para identificar los riesgos y evaluar los controles que garanticen la seguridad de un proceso de ciencia de datos en una institución financiera del Ecuador. La metodología está basada en mejores prácticas como COBIT, NIST y GTAG, así como en la normativa de riesgo operativo, resolución No. SB-2018-771 emitida por la Superintendencia de Bancos.

Palabras clave: Ciencia de datos, *Big Data*, auditoría, seguridad, riesgos, controles

ABSTRACT

For years, the analysis of information has been one of the most important and necessary activities within financial institutions, since it provides data which allows us to understand the data generated by the business procedures. However, due to the growing evolution of technological trends, the data analytics evolved to a discipline that improves the descriptive analysis done until today and that makes possible the prediction of future situations based on pattern detection, this new discipline is called data science.

Data science integrates disciplines such as statistics, mathematics and programming, and also uses techniques such as machine learning to obtain valuable information and guarantee the delivery of quick, honest and truthful results. On the other hand, with the help of data science, a large amount of information from different types of data can be processed in an efficient way, which is why it is also closely related to the concept of Big Data.

Implementing the data science process allows us to create business strategies, recognize new market opportunities and achieve competitive advantages.

Although this process was implemented in financial institutions to transform information into knowledge and provide some benefits, its implementation could entail risks that must be identified, controlled and mitigated. The present work proposes an audit methodology to identify the risks and evaluate the controls that guarantee the security of a data science process in a financial institution of Ecuador. The methodology is based on best practices such as COBIT, NIST and GTAG, as well as operational risk regulations, resolution No. SB-2018-771 issued by the Superintendencia de Bancos.

Key words: Data science, Big Data, audit, security, risks, controls

CAPÍTULO I. INTRODUCCIÓN

1.1. El problema de la Investigación

1.1.1. Planteamiento del problema

Las instituciones financieras se enfrentan al manejo de grandes volúmenes de información de diferentes tipos de data y a la necesidad constante de procesarla y obtener resultados, debiendo levantar procesos para realizar estas actividades de forma rápida y facilitando la generación de reportes en base a información confiable y veraz, que permitan posteriormente tomar decisiones estratégicas oportunamente. Al hablar de la información que se maneja en una institución financiera, no sólo se refiere a temas contables y financieros sino también a la información indirecta que se obtiene de los clientes: perfil transaccional, rutinas de pagos, ahorros y gastos, data que puede contribuir a un análisis para madurar los servicios bancarios ofertados, para la consecución de estos objetivos existen soluciones como la ciencia de datos.

La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento desde diferentes tipos de información: estructurada y no estructurada (Liu, 2018).

De acuerdo a entrevista realizado con Jiménez (2018), Gerente de Ciencia de datos de la institución financiera en la que se enfoca el presente trabajo, se ha implementado recientemente este proceso con el objetivo de mejorar oportunidades de negocio, en base al almacenamiento, la extracción, análisis y generación de reportes con información íntegra, veraz y de interés

comercial para el banco. Por ejemplo, es posible obtener datos que sugieran la creación o mejora de un producto o servicio, de acuerdo con el análisis de las preferencias o estilos de vida de los clientes y a su perfil transaccional.

No cabe duda que la implementación de este proceso, está enfocada en obtener numerosos beneficios, sin embargo, se debe tomar en cuenta el surgimiento de algunos riesgos relacionados a la seguridad e integridad de la data, que deben ser identificados y mitigados oportunamente. Por este motivo y tomando en cuenta la criticidad de la información que se maneja en el proceso de ciencia de datos, el área de Auditoría Informática ha incorporado una revisión a este proceso dentro de su planificación, cuyos resultados serán presentados a la Superintendencia de Bancos. El enfoque de la auditoría estará relacionado a identificar riesgos claves y controles establecidos por la entidad que garanticen la integridad de la información y la seguridad de la arquitectura que soporta el proceso de ciencia de datos y sugerir controles que no hayan sido contemplados.

Adicionalmente, en base al análisis de las metodologías utilizadas en la institución para la ejecución de auditorías, se ha identificado que no existen buenas prácticas o normativas enfocadas exclusivamente a este proceso.

1.1.2. Formulación del problema

En la institución financiera privada no se cuenta con una guía de revisión exclusiva para el proceso de ciencia de datos, que especifique los riesgos claves a ser identificados y los controles que deberían aplicarse para mitigarlos, provocando que en las evaluaciones de auditoría no se reconozcan los riesgos relevantes del proceso, como la falta de integridad de los resultados, la falta de certeza de la disponibilidad o fallas en la precisión de los modelos desarrollados.

Adicionalmente, no sería posible asegurar que los controles destinados a mitigar los riesgos, están diseñados de forma adecuada y funcionan efectivamente.

1.2. Objetivos

1.2.1. Objetivo general

Diseñar una metodología de auditoría al proceso de ciencia de datos, para la identificación de sus principales riesgos y controles, utilizando COBIT, NIST, GTAG y la normativa de Riesgo Operativo publicada por la Superintendencia de Bancos.

1.2.2. Objetivos específicos

- Analizar la situación actual del área de Ciencia de datos, mediante entrevistas con la gerencia del área, para la identificación de posibles deficiencias en el proceso.
- Identificar riesgos sobre el proceso de ciencia de datos, a través del análisis de sus fases y arquitectura para el levantamiento de controles aplicados a la realidad de la entidad.
- Seleccionar los controles de COBIT, NIST, GTAG y la normativa de Riesgo Operativo, mediante la identificación de temas relacionados a la seguridad del proceso de ciencia de datos, para el levantamiento de controles base de la metodología.
- Desarrollar el programa de auditoría de ciencia de datos de acuerdo a las directrices del proceso auditor CISA, para la identificación de los componentes necesarios que conforman una metodología.

1.3. Justificación

Debido a que actualmente en la institución financiera no se cuenta con un programa de revisión exclusivo para el proceso de ciencia de datos, es necesario diseñar una metodología de auditoría con el objetivo de identificar riesgos y controles claves enfocados a la seguridad del entorno por el que atraviesa la información en este proceso.

Adicionalmente, al no existir una guía dirigida exclusivamente a evaluar la seguridad del proceso de ciencia de datos, se investigaron marcos de referencia que hayan sido desarrollados y aprobados por expertos (COBIT, NIST, GTAG) y que apoyen al desarrollo de una metodología acorde a las necesidades de la institución y sean referentes para identificar controles que garanticen la seguridad de la información.

Levantar una metodología de evaluación de ciencia de datos proporcionará facilidades al área de Auditoría Informática, ya que contará con un esquema de evaluación de este proceso y se podrá emitir resultados que otorguen valor a los procedimientos de gestión de la información, generando confianza en la toma de decisiones y apalancamiento de las estrategias empresariales. Adicionalmente, esta metodología podrá ser utilizada en revisiones a futuro ya que se basará en COBIT, NIST, GTAG y en la normativa de riesgo operativo adaptados a la institución financiera.

CAPÍTULO II. MARCO TEÓRICO

2.1. Ciencia de datos

2.1.1. Definición de ciencia de datos

El concepto de ciencia de datos se ha definido de varias maneras:

- IBM menciona que la ciencia de datos es el proceso de descubrir información oculta en grandes cantidades de datos estructurados y no estructurados, utilizando métodos como la estadística, *machine learning*, la minería de datos y la analítica predictiva. Esta área multidisciplinar está cambiando el modo en que las organizaciones resuelven los problemas. (Liu, 2018).
- Para Waller & Fawcett (2013) el proceso de ciencia de datos es una metodología para proporcionar soluciones de análisis predictivo. El mismo ayuda a mejorar la colaboración en equipo y el aprendizaje.
- Según Van der Aalst (2016) la Ciencia de Datos es una combinación de disciplinas cuyo enfoque está centrado en convertir la información en valor para los individuos, las organizaciones y la sociedad.

Pero, ¿cómo llegó a ser tan necesario este proceso? Sin duda alguna todo nace desde el punto de vista de la importancia del activo más importante en una organización: la información, además de los resultados que se espera de su procesamiento y posterior análisis para la toma de decisiones.

Sin embargo, con el paso de los años y a medida que tecnologías nuevas iban apareciendo, se presentaron algunos problemas con respecto al procesamiento de la información, principalmente con respecto a la obtención de información y al tipo de datos que se deben procesar:

- La obtención de la información relevante está relacionada a extraer conocimiento, a partir de una cantidad diversa de datos (Provost & Fawcett, 2013).
- Por otro lado, Pérez Marqués(2015) menciona que existe una gran variedad de datos que pueden ser representados de diversas maneras en todo el mundo, por ejemplo de dispositivos móviles, audio, video, sistemas GPS, sensores, automóviles, medidores eléctricos, veletas, anemómetros, etc., los cuales pueden medir y comunicar el posicionamiento, movimiento, vibración, temperatura, humedad y hasta los cambios químicos que sufre el aire, de tal forma que las aplicaciones que analizan estos datos requieren que la velocidad de respuesta sea lo demasiado rápida para lograr obtener la información correcta en el momento preciso. Según Rayo (2016), para el año 2020 se espera que haya entre 25000 y 35000 millones de dispositivos, con una tasa de crecimiento anual del 40%, capaces de generar datos solo en el mundo del Internet de las cosas.

Es aquí en donde toma importancia el concepto de ciencia de datos, ya que su razón de ser está enfocada en utilizar la ciencia y el análisis de datos para generar información útil para el futuro. El manejo adecuado del proceso de ciencia de datos, permite generar información de valor para las empresas, para la toma de decisiones y afinamiento de sus estrategias de negocio.

2.1.2. Procesos de ciencia de datos

Según (O'Neil & Schutt, 2013), la ciencia de datos requiere atravesar el siguiente conjunto de fases, que permitirán obtener conocimiento para la toma de decisiones a partir de la data. Las fases pueden ser iterativas, ya que puede existir retroalimentación de otras fases:

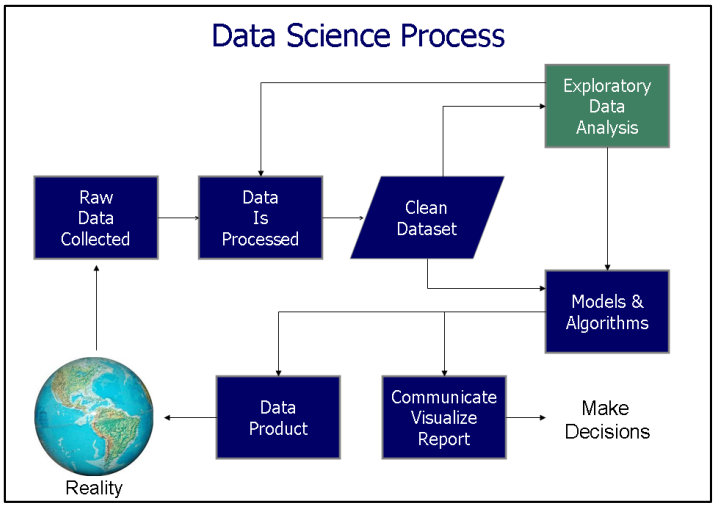

Figura 1. Procesos de ciencia de datos. **Fuente:** O'Neil & Schutt (2013)

- **Requerimiento de datos:** Nacen de la necesidad planteada, la misma que proporcionará la información que será la entrada para el análisis.
- **Recopilación de datos:** Hace referencia a la obtención de la data desde diferentes dispositivos y fuentes. Adicionalmente, la extracción de la información puede ser de diferentes tipos de datos.
- **Procesamiento de datos:** Es la organización de la información previo al análisis que se realizará sobre la data.
- **Limpieza de datos:** Al ingresar y almacenar datos pueden presentarse errores o duplicación de la información, por lo que existe un paso de limpieza como un control de prevención.

- **Análisis exploratorio de datos:** Es la aplicación de técnicas para la comprensión de la información.
- **Modelado y algoritmos:** Es la construcción de modelos para evaluar una variable y su comportamiento con otras variables y determinar relaciones.
- **Producto de datos:** Es el resultado final, puede ser una aplicación que toma los datos de entrada y devuelve una salida basada en análisis realizado en fases anteriores.
- **Comunicación:** Es la decisión de qué forma se comunicarán, presentarán y visualizarán los resultados. Esta es una fase en la que participan quienes levantaron el requerimiento, por lo que pueden existir comentarios y solicitudes de cambio, lo que convierte a este ciclo en un proceso iterativo.

2.1.3. Áreas de conocimiento de ciencia de datos

El científico de datos estadounidense Conway (2013), realizó un diagrama de Venn para representar las disciplinas que requieren ser explotadas en el proceso de la ciencia de datos:

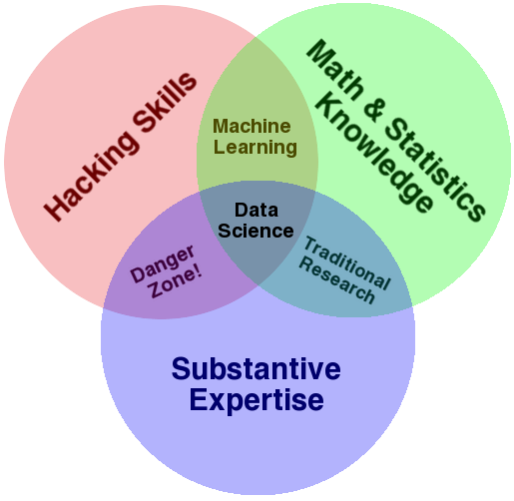

Figura 2. Diagrama de Venn "Científico de Datos". **Fuente:** Conway (2013)

- **Hacking Skills (Habilidades informáticas):** Habilidad para organizar, analizar y manipular los datos que provienen de fuentes de diferentes.
- **Math & Statistics Knowledge (Estadística y matemáticas):** Modelamiento matemático y estadístico para interpretar y procesar la información con las herramientas adecuadas.
- **Substantive expertise (Conocimiento del entorno):** Hace referencia al conocimiento del contexto y dominio del área en la que se está aplicando ciencia de datos.

Adicionalmente, Conway (2013) menciona que existen subconjuntos que se forman de la relación entre las áreas mencionadas, las cuáles hacen referencia en lo que pasa si se adquieren 2 habilidades y no la fusión con una tercera:

- **Machine Learning:** Fusión en la que no se tendría conocimiento del entorno del trabajo, por lo que no tendría razón de ser la ciencia de datos, ya que se podrán tener las habilidades de programación, matemáticas y estadística, pero lo fundamental es dominar el entorno en donde se van a aplicar esos conocimientos.
- **Investigación tradicional:** En este subconjunto no se podría manejar la información rápida y ágilmente, ya que los conocimientos de lenguajes de programación serían una gran limitante.
- **Zona peligrosa:** Si no se mantiene conocimientos en matemáticas y estadística se corre el riesgo de que se procese o interprete la información incorrectamente, comprometiendo incluso futuros trabajos que se basen en estos resultados.

2.1.4. Diferencia de terminologías

Ante la inminente evolución de las tecnologías, se genera confusión entre algunos de los términos *Business Intelligence*, *Business Analytics*, ciencia de datos y *Big Data*, que, si bien están interrelacionados, mantienen diferentes alcances y objetivos

Business Intelligence hace referencia al análisis descriptivo de información, los datos analizados son estructurados y ayuda a conocer el rendimiento de las operaciones de una institución. Brinda información relacionada al estado actual de la institución y lo presenta en un esquema de visualización de alto nivel, de tal forma que usuarios no técnicos puedan apreciarlo (Chen, Chiang, & Storey).

Business Analytics, se basa en estadística para realizar análisis predictivos y prescriptivos de la información, ayuda a pronosticar situaciones futuras (Chen, Chiang, & Storey).

El *Big Data*, se refiere al almacenamiento y procesamiento de un conjunto de datos que pueden ser estructurados y/o no estructurados (Provost & Fawcett, 2013):

- **Estructurada:** Forma tradicional de almacenar y organizar la información en arreglos o tablas con un orden definido.
- **No estructurada:** No tiene un modelo de organización definido, por ejemplo, las búsquedas realizadas a páginas web, la información de redes sociales, etc.

El *Big Data* está estrechamente relacionado con los conceptos de volumen (gran cantidad de información), variedad (diferentes tipos de datos) y velocidad(capacidad de respuesta inmediata), cualidades que no se podrían realizar en herramientas tradicionales (Provost & Fawcett, 2013).

Sin embargo, la verdadera funcionalidad de mantener este tipo de tecnología es explorar, analizar y generar valor de la información almacenada. Es aquí, en donde se torna relevante el concepto de ciencia de datos.

La ciencia de datos, se enfoca en la consolidación de técnicas (estadística, matemática, conocimiento del negocio, programación) para analizar y levantar modelos que permitan predecir situaciones, preferencias o incluso detectar patrones, generando ventajas competitivas (Provost & Fawcett, 2013).

En conclusión, Business Intelligence y Business Analytics trabajan con información estructurada y realizan análisis descriptivos y predictivos respectivamente, a este último análisis se le aplica conocimientos estadísticos. La ciencia de datos emplea técnicas adicionales que permiten entender comportamientos y patrones en base a información no estructurada. Adicionalmente, ciencia de datos y *Big Data* están relacionados, ya que el proceso de ciencia de datos, con ayuda de sus herramientas y técnicas, facilitan el análisis de gran cantidad de información (*Big Data*).

2.1.5. Arquitectura

Una vez que se conoce el concepto de *Big Data*, es necesario analizar la arquitectura que permita soportar una tecnología de este nivel. Se ha escogido analizar la arquitectura mencionada por la NIST(2018), instituto que creó un framework de Interoperabilidad para Big Data, con una serie de volúmenes para entender y gestionar a esta tecnología. Existe un capítulo específico dedicado a comprender la arquitectura de Big Data, denominada NBDRA (NIST Big Data Reference Architecture) la cual se ha tomado como referencia para el presente trabajo:

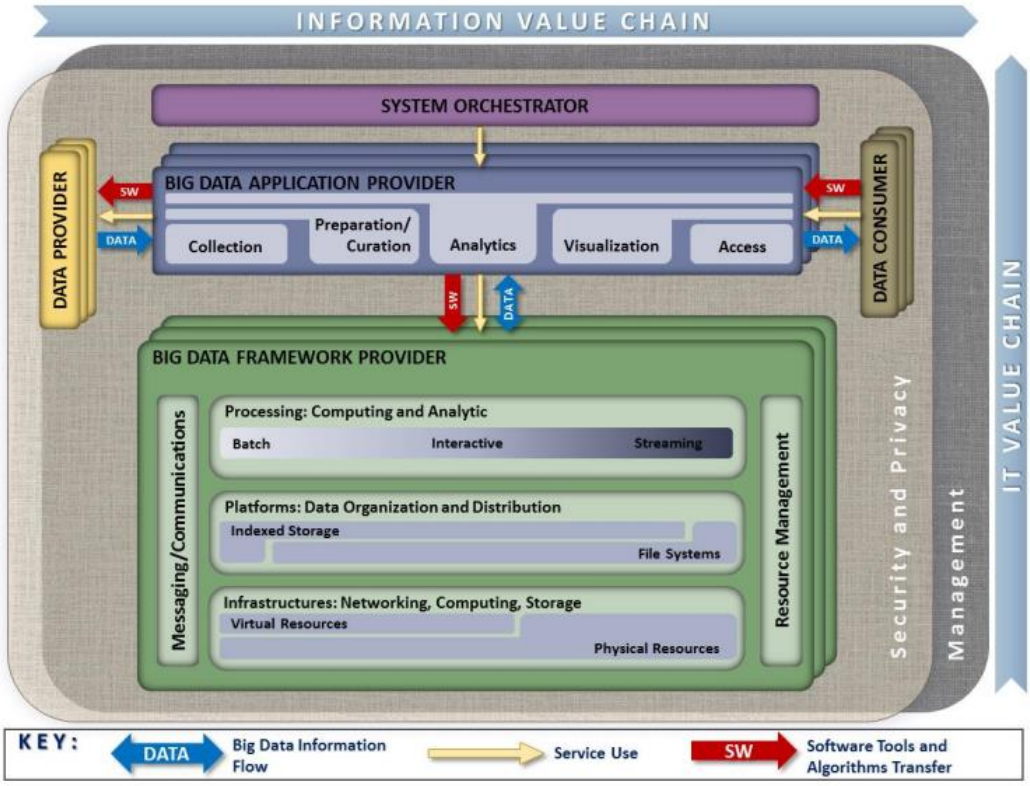

Figura 3. Arquitectura Big Data. **Fuente:** NIST(2018)

La arquitectura presentada por la NIST, refleja dos ejes que se mantienen relacionados, el eje horizontal que es la Información y el eje vertical que son las Tecnologías de Información. El eje vertical hace referencia a la recolección, análisis y procesamiento de la información, mientras que el eje horizontal hace referencia a las herramientas, infraestructura y servicios necesarios para soportar una implementación de Big Data.

Por otro lado, existen dos capas que se encuentran sosteniendo la arquitectura, que son la capa de gestión y la capa de seguridad y privacidad. La interpretación es que son temas que abarcan a toda la arquitectura y que no se puede descartar controles que garanticen una adecuada gestión, seguridad y privacidad de la información en todos los componentes que conforman a Big Data.

Adicionalmente, existen 5 componentes que completan el modelo de arquitectura propuesto por la NIST:

- **Sistema de orquestación:** Se refiere a los requerimientos que debe cumplir la implementación de Big Data, como son la arquitectura, los recursos y las necesidades del negocio. También están incluidas tareas de seguimiento y evaluaciones de auditorías, para verificar el cumplimiento de los requisitos.
- **Proveedor de datos:** Es un rol encargado de mantener disponibles los datos para sí mismo como para otros roles. Los datos son puestos a disposición a través de diferentes interfaces y las fuentes de datos pueden ser diversas (audios, imágenes, sensores consultas web, etc.).
- **Proveedor de aplicaciones Big Data:** Es el encargado de ejecutar operaciones para cumplir con los requisitos definidos en el Sistema de Orquestación. Es la etapa en la que se desarrolla la funcionalidad y lógica de negocio solicitada en los requerimientos iniciales. Existen subprocesos dentro de este componente:
    - **Recolección de datos**: Está relacionada al componente proveedor de datos y hace referencia a un servicio como un servidor de archivo implementado por un sistema de orquestación para extraer información.
    - **Preparación:** Actividad en donde se llevan a cabo los procesos de extracción, transformación y carga, adicionalmente, se realizan tareas relacionadas a la limpieza y validación de datos.
    - **Análisis:** Es el levantamiento de técnicas, algoritmos y métodos para generar conocimiento de acuerdo a los requerimientos solicitados inicialmente.

  - **Visualización:** Etapa en la que se alistan los resultados para la presentación al consumidor de datos.
  - **Acceso:** Son las actividades relacionadas a entregar y comunicar el requerimiento al Consumidor de datos.

- **Proveedor de framework de Big Data:** Componente en donde se especifican los recursos o servicios necesarios para el proveedor de aplicaciones de Big Data en el levantamiento de un proyecto. El proveedor de framework se divide en 3 subcomponentes: Plataforma de infraestructura, plataforma de datos y plataforma de procesamiento:
  - **Infraestructura:** Se refiere a todos los servicios y medios técnicos tales como: redes, computación (procesadores, memoria física), almacenamiento, ambiente (seguridad de infraestructura, servicios en la nube).
  - **Plataformas de datos:** Se refiere a la organización lógica de la información y los métodos de acceso. Estos métodos pueden variar desde APIs de acceso hasta lenguaje SQL.
  - **Plataforma de procesamiento: Se** refiere al software utilizado para implementar las soluciones de Big Data. El procesamiento a este nivel puede darse por batch o por streaming, dependerá del requerimiento. Adicionalmente hace referencia a los algoritmos y modelos matemáticos, entre los más reconocidos se encuentra Map Reduce y BSP.
- **Consumidor de datos:** Es el actor que recibe el resultado final del procesamiento.

2.1.6. Evolución de ciencia de datos

El concepto de *Data Science* nace después de una evaluación a propuestas de revisión de las áreas técnicas existentes alrededor a la estadística, con el objetivo de mejorar el análisis de datos realizado en aquellos tiempos con *Data Mining* y su aplicación en algunos procesos de negocio. Adicionalmente, existía un avance acelerado con respecto a recolección y análisis de datos. Posterior a estas evaluaciones, el concepto de *Data Science* se fue consolidando como la unión de diferentes áreas de conocimientos (estadística y matemáticas, habilidades de programación, dominio del contexto) que sostienen el concepto actual del análisis de datos (Friedman, 2001).

2.1.7. La ciencia de datos y las empresas

Galimany (2014) menciona que a través de los años, las empresas han cambiado sus objetivos estratégicos, dejando de lado el vender por vender y centrándose en mejorar la experiencia del usuario al ofertar nuevos productos, es decir, se ha generado el factor "añadir valor" para crear fidelización por parte del cliente a través de análisis predictivo de información.

Adicionalmente, en años anteriores la información que se obtenía se utilizaba únicamente para ordenar la data por módulos y posteriormente cuadrar estados financieros, sin embargo, para Galimany (2014) actualmente existe la alternativa de obtener información y conocimiento valioso a través de la aplicación de técnicas y herramientas, que permita pronosticar situaciones y en función de las mismas tomar decisiones.

Por lo tanto, según Galimany (2014), la ciencia de datos permite a las empresas aprovechar el gran volumen de información que mantiene, para analizarla, procesarla y transformarla en un resultado predictivo que genere valor.

2.1.8. Uso de ciencia de datos en entidades financieras

Las entidades financieras, al igual que otras empresas se enfrentan a un acelerado cambio tecnológico, el cual repercute en forma directa a sus actividades. Entre los factores principales que han cambiado, se puede mencionar:

- La utilización de dispositivos móviles para acceder a la información y con el paso del tiempo el uso de "Internet de las cosas", en el que los dispositivos que se utilicen en el diario vivir ya cuenten con tecnología móvil incorporada (Rayo, 2016).
- Las diferentes alternativas para almacenar información, desde métodos tradicionales con mejoras en su capacidad (repositorios, bases de datos), como métodos modernos como el almacenamiento en la nube.
- La capacidad de ejecutar instrucciones por segundo, permitiendo la ejecución simultánea de operaciones.
- Capacidad de uso de modelos que permiten la toma de decisiones de manera automática, lo que genera algunos beneficios como eficiencia y objetividad. (Management Solutions, 2015)

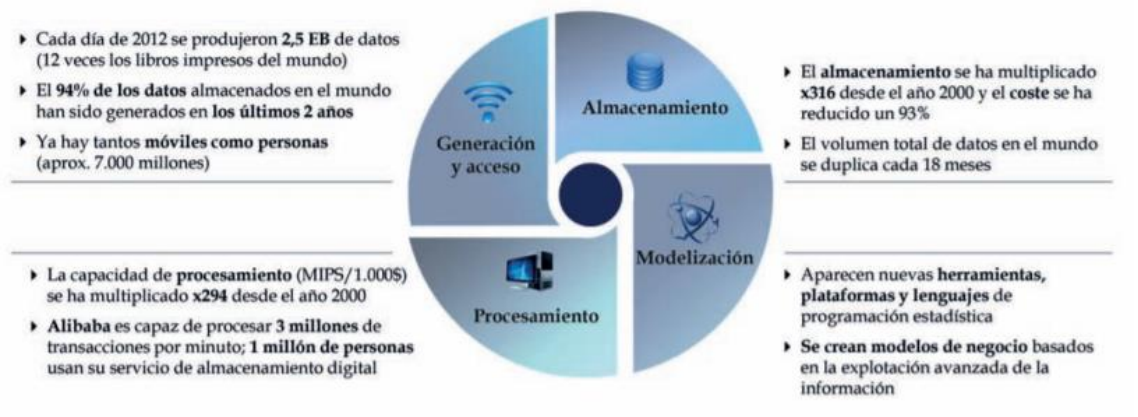

Figura 4. Revolución tecnológica. **Fuente:** Management Solutions (2015)

Las entidades financieras son una de las industrias más beneficiadas con este aceleramiento de tecnología, ya que un adecuado manejo de la información, puede permitirles desarrollar propuestas de valor a sus clientes (entendimiento de necesidad, personalización de productos, uso de canales, etc.) (Management Solutions, 2015). Sin embargo, existen riesgos inherentes por la naturaleza de la entidad, que deben ser tomados en cuenta para cumplir con las normas de la Superintendencia de Bancos y a la vez mantener la fidelización de los clientes.

El aspecto normativo sin duda presenta grandes retos para las instituciones bancarias, si bien actualmente no existe una norma que mencione temas relacionados al análisis de la información bajo procesos de ciencia de datos, las instituciones financieras están obligadas a cumplir con la normativa de riesgo operativo, resolución No. SB-2018-771 emitida por la Superintendencia de Bancos. Adicionalmente, es importante mantener planes contingentes de respuesta inmediata que permitan adaptarse a nuevas resoluciones emitidas por la Superintendencia.

Otro aspecto a ser tomado en cuenta es que los clientes cada vez se vuelven más exigentes, esto debido a las diversas ofertas que se le presentan, así como la facilidad de automatizar algunas transacciones a través de canales electrónicos y dejando atrás los trámites presenciales,

por lo que toma mayor importancia que las instituciones bancarias maduren su estrategia de negocio y se enfoquen en cómo fidelizar al cliente a través de servicios y productos de vanguardia.

2.1.9. Ventajas de usar ciencia de datos

Según un informe realizado por Ernst & Young (2017), las ventajas relacionadas al uso de ciencia de datos son:

- Toma de decisiones
- Fidelización de clientes
- Mejoras en las estrategias de marketing
- Incremento de la eficiencia y disminución de los costes
- Desarrollo de una relación más estrecha entre proveedores y socios

2.1.10. Implementación de un proceso de ciencia de datos

El artículo "Metodología fundamental para la ciencia de datos" publicado por Rollins (2015) de IBM, presenta las etapas para la obtención de conocimiento mediante el uso de datos:

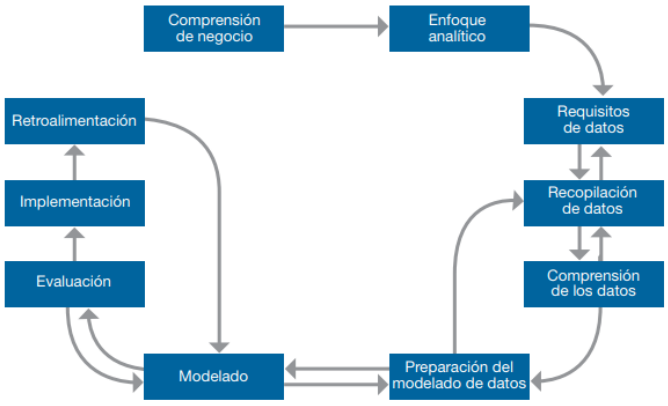

Figura 5. Metodología para la ciencia de datos. **Fuente:** Rollins (2015)

- **Etapa 1 - Comprensión del negocio:** Hace referencia que existe un usuario solicitante, quien debe mantenerse en todo el ciclo del desarrollo, ya que tiene conocimientos especializados y puede garantizar que el resultado del proyecto vaya acorde a lo inicialmente solicitado.
- **Etapa 2 - Enfoque analítico:** Es la etapa en el que el Científico de Datos analiza el problema y lo expresa en el contexto de técnicas estadísticas y aprendizaje automático.
- **Etapa 3 - Requisitos de datos:** Una vez escogido el enfoque analítico, se requiere determinar los formatos, contenido y representaciones de datos que requerirá el modelo a implementarse.
- **Etapa 4 - Recopilación de datos:** En esta etapa se determinan las fuentes de información que se tomarán para resolver el requerimiento. Se considera importante conocer la arquitectura que soporta el proceso, ya que de eso dependerá la cantidad de información que pueda ser extraída y procesada.
- **Etapa 5 - Comprensión de los datos:** Se refiere al uso de técnicas de visualización y estadística descriptiva para entender a los datos y evaluar su calidad.
- **Etapa 6 - Preparación de los datos:** Hace referencia a la ejecución de todas actividades realizadas previo a la construcción del requerimiento, incluye: combinación de información de diferentes fuentes, limpieza de datos.
- **Etapa 7 - Modelado:** Es el desarrollo de los modelos predictivos o descriptivos que solucione el requerimiento solicitado. En el caso de los modelos predictivos los Científicos de Datos utilizan un conjunto de entrenamiento para construir el modelo.

Se suelen probar varios algoritmos para encontrar el modelo que mejor se adopte a la solución.

- **Etapa 8 - Evaluación:** La evaluación del modelo consiste en medir la calidad y eficacia del modelo a través de tablas y gráficos. En el caso de modelos predictivos también se emplean conjuntos de prueba independiente del conjunto de entrenamiento el cual se utiliza para ajustar y evaluar al modelo.
- **Etapa 9 – Implementación:** Una vez que el modelo ha sido probado y certificado por los usuarios solicitantes, e procede a pasarlo a producción.
- **Etapa 10 – Retroalimentación:** Posterior a la implementación del requerimiento, se puede obtener retroalimentación del rendimiento del modelo y cómo impactó en el esquema bajo el cual fue implementado.

2.2. Auditoría

2.2.1. Definición de Auditoría

El proceso de Auditoría consiste en evaluar la eficacia de los procesos de gestión de riesgos y control en una organización, con el objetivo de agregar valor y en base a recomendaciones, mejorar de la ejecución de los procesos (Piattini & Del Peso, 2001).

2.2.2. Definición de Auditoría de Sistemas

Está enfocada a revisar y verificar la integridad de la información y datos almacenados en las bases de datos de los sistemas de información y los procesados por éstos. (Morrobert, 2011).

2.2.3. Fases de Auditoría

Según la Contraloría General del Estado (2014), existen tres fases de Auditoría:

- **Fase de Planificación:** Fase en la que se define la estrategia de auditoría a emplear en el proceso escogido. Es la etapa más crítica ya que se define el alcance de la revisión y los programas de Auditoría a ejecutarse durante la siguiente fase.
- **Fase de Ejecución:** Es la ejecución del programa de Auditoría levantado en la anterior fase, aplicando la estrategia definida. Se establecen las reuniones necesarias para realizar las revisiones y solicitar las pruebas y evidencias que sustenten lo mencionado.
- **Fase del informe final de auditoría:** Es la conclusión del trabajo, en donde se da a conocer los resultados de la evaluación. Se emiten comentarios y recomendaciones.

Cabe mencionar que el presente trabajo de tesis está enfocado en el diseño de una metodología en la que se base el programa de auditoría de ciencia de datos, es decir, las actividades a realizarse en la fase de planificación. Las fases de ejecución e informe final no estarían dentro del alcance a revisar.

2.2.4. Desarrollo de Programas de Auditoría

ISACA, el Instituto de Auditores Internos (IIA) y otras organizaciones, han levantado algunos programas de auditoría o fuentes de aseguramiento que sirven como base para realizar un adecuado proceso de Auditoría. Cooke (2017) menciona que, si bien los programas de auditoría son una base y guía fundamental para ejecutar una revisión, su enfoque es general y queda al criterio y experiencia del auditor el detalle que le asigne al programa conforme al proceso que se esté revisando.

Figura 1—Programas de auditoria/ aseguramiento existentes

Fuente	Descripción
ISACA	Programas de Auditoria/Aseguramiento[4]
IIA	Guías de Auditoría de Tecnología Global (GTAGs)[5]
AuditNet	Programas de Auditoría[6]

Figura 6. Programas de Auditoría. **Fuente:** Cooke (2017)

Un programa de auditoría es un conjunto de pasos que son ejecutados para cumplir con el alcance de auditoría y emitir un informe de resultados.

Según Cooke (2017) en su artículo "*Auditoría de Sistemas de Información: Herramientas y Técnicas para la Creación de Programas de Auditoría*" publicado por ISACA en el 2017, se determina que existen 5 pasos para levantar un programa de auditoría:

1. **Determinar la materia de la auditoría:** Identificar el área a ser auditada.
2. **Determinar el objetivo de la auditoría:** Identificar el propósito de la revisión.
3. **Establezca el alcance de la auditoría:** Identificar sistemas o funciones a ser incluidas en la revisión, dependiendo el tiempo que se le vaya a destinar a la revisión de auditoría.
4. **Realice la planificación previa a la auditoría:** Realizar una evaluación del riesgo, identificar personal para realizar entrevistas.
5. **Determine los pasos para la recolección de datos:** Se identifica la estrategia de auditoría, requisitos normativos, scripts de validación y criterios de evaluación de las pruebas.

2.3. Metodologías

El presente trabajo se enfoca en diseñar una metodología de evaluación para ciencia de datos, por tal motivo se investigó a CISA como un proceso auditor que mencione los lineamientos a seguir para realizar una revisión de auditoría.

2.3.1. Certificación CISA

CISA propone un proceso de auditoría en el que se aplican una serie de estándares, directrices, herramientas y técnicas que el responsable de auditoría debe realizar para obtener finalmente un informe en el que se reflejen los resultados de la revisión realizada. (ISACA, 2015). Se encuentra compuesto por los siguientes dominios:

- Dominio 1: El proceso de auditoría de sistemas de información
- Dominio 2: Gobierno y gestión de TI
- Dominio 3: Adquisición, desarrollo e implementación de sistemas de información
- Dominio 4: Gestión de servicios, mantenimiento y operaciones de sistemas de información
- Dominio 5: Protección de los activos de información

2.3.2. NIST

Instituto Nacional de Estándares y Tecnología de Estados Unidos, encargado de crear guías para distintos tipos de tecnologías. Su principal objetivo es promover la innovación y la competitividad industrial mediante la mejora de ciencia de la medición, los estándares y la tecnología en formas que mejoren la seguridad económica y mejorar la calidad de vida.

NIST desarrolló una publicación con 7 volúmenes orientados a Big Data, los mismos forman un marco de interoperabilidad que busca identificar componentes clave de la arquitectura y la relación entre sus interfaces (NIST, 2000):

- **Definiciones de Big Data**: Conceptos fundamentales para comprender el nuevo paradigma y los procesos analíticos del proceso de ciencia de datos.
- **Taxonomía Big Data:** Organización de los componentes de la arquitectura para sentar las bases en torno al concepto de Big Data.
- **Casos de uso y requisitos Big Data:** Recopilar casos de uso y extraer requisitos para formar la arquitectura de referencia de Big Data NIST
- **Seguridad y privacidad Big Data:** Identificar problemas de seguridad y privacidad que son específicos de Big Data. Los dominios de aplicaciones Big Data incluyen atención médica, descubrimiento de fármacos, seguros, finanzas, venta minorista y muchos otros de los sectores privado y público. Los dominios de tecnología de seguridad incluyen identidad, autorización, auditoría, red y seguridad.
- **Encuesta de arquitectura Big Data:** Sirve para facilitar la comprensión de las complejidades operacionales en Big Data y para servir como herramienta para desarrollar arquitecturas específicas del sistema utilizando un marco de referencia común.
- **Arquitectura de Big Data:** Caracteriza a Big Data desde la perspectiva de la arquitectura, presenta el modelo conceptual NIST de Big Data Reference Architecture (NBDRA) y analiza sus componentes.
- **Estándares de hoja de ruta Big Data:** Resúmenes del trabajo presentado en los otros seis volúmenes y una investigación de estándares relacionados con Big Data.

Si bien estas publicaciones están orientadas a Big Data, pueden ser una fuente importante para entender la arquitectura y componentes de un proceso de ciencia de datos y de esta forma generar una guía de evaluación consistente.

2.3.3. COBIT

Proporciona un conjunto de buenas prácticas y controles para la gerencia, proporcionando una visión integral y sistemática del gobierno y la gestión de las TI basada en objetivos de control lo que permite asegurar que las TI y el negocio están alineados. Su última publicación de COBIT5, integra en su redacción los modelos de procesos RisKIT y ValIT (Otro de los productos de la familia COBIT5). (ISACA, 2012).

COBIT 5 permite que las tecnologías de la información se gobiernen y administren de una manera holística a nivel de toda la Organización, incluyendo el alcance completo de todas las áreas de responsabilidad funcionales y de negocios, considerando los intereses relacionados con la TI de las partes interesadas internas y externas (ISACA, 2012).

Dentro de la metodología se han definido habilitadores para responder a las metas definidas por la organización. Son componentes que hacen que los procesos y políticas puedan interpretarse mejor dentro de las organizaciones, y que en consecuencia se van a reflejar en su rendimiento y alcance de objetivos (Braga, 2012):

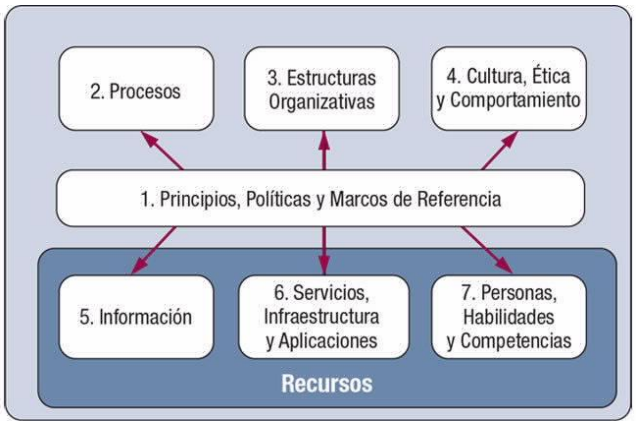

Figura 7. Habilitadores de COBIT5. **Fuente:** Braga (2012)

Adicionalmente, existen principios que se han desarrollado para gobernar y gestionar efectivamente la información y su tecnología y pueden beneficiar a cualquier empresa, sin importar su tamaño, ubicación o industria.

Figura 8. Principios de COBIT5. **Fuente:** Braga (2012)

Cobit 5, se basa en cinco dominios, con 37 procesos distribuidos por cada dominio (Braga, 2012):

Evaluar, Dirigir y Monitorear

- EDM01 Asegurar que se fija el Marco de Gobierno y su Mantenimiento
- EDM02 Asegurar la Entrega de Valor
- EDM03 Asegurar la Optimización de los Riesgos
- EDM04 Asegurar la Optimización de los Recursos
- EDM05 Asegurar la Transparencia a las partes interesadas

Alinear, Planear y Organizar

- APO01 Administrar el Marco de la Administración de TI
- APO02 Administrar la Estrategia
- APO03 Administrar la Arquitectura Corporativa
- APO04 Administrar la Innovación
- APO05 Administrar el Portafolio
- APO06 Administrar el Presupuesto y los Costos
- APO07 Administrar el Recurso Humano
- APO08 Administrar las Relaciones
- APO09 Administrar los Contratos de Servicios
- APO10 Administrar los Proveedores
- APO11 Administrar la Calidad
- APO12 Administrar los Riesgos
- APO13 Administrar la Seguridad

Monitorear, Evaluar y Valorar

- MEA01 Monitorear, Evaluar y Valorar el Desempeño y Cumplimiento
- MEA02 Monitorear, Evaluar y Valorar el Sistema de Control Interno
- MEA03 Monitorear, Evaluar y Valorar el Cumplimiento con Requisitos Externos

Construir, Adquirir e Implementar

- BAI01 Administrar Programas y Proyectos
- BAI02 Administrar la Definición de Requerimientos
- BAI03 Administrar la Identificación y Construcción de Soluciones
- BAI04 Administrar la Disponibilidad y Capacidad
- BAI05 Administrar la Habilitación del Cambio
- BAI06 Administrar Cambios
- BAI07 Administrar la Aceptación de Cambios y Transiciones
- BAI08 Administrar el Conocimiento
- BAI09 Administrar los Activos
- BAI10 Administrar la Configuración

Entregar, Servir y Dar Soporte

- DSS01 Administrar las Operaciones
- DSS02 Administrar las Solicitudes de Servicios y los Incidentes
- DSS03 Administrar Problemas
- DSS04 Administrar la Continuidad
- DSS05 Administrar los Servicios de Seguridad

- DSS06 Administrar los Controles en los Procesos de Negocio

2.3.4. Normas para instituciones financieras

Las instituciones financieras se encuentran regidas por las normativas expedidas por la Superintendencia de Bancos. Existe la normativa de riesgo operativo, resolución No. SB-2018-771, la cual analiza la probabilidad de que se ocasionen pérdidas financieras debido a la falla de ejecución de 4 factores (Superintendencia de Bancos, 2018):

- Procesos
- Personas
- Tecnologías de la Información
- Eventos externos.

En la metodología propuesta para auditar al proceso de ciencia de datos, se tomará en cuenta la presente normativa, con el objetivo de alinearla a las buenas prácticas y evaluar eventos de fallas en procesos, en personas relacionadas al proceso, la tecnología empleada y los eventos externos que podrían afectar a la institución.

2.3.5. Estado del arte

El avance acelerado de la tecnología ha persuadido a las empresas a mantenerse constantemente actualizadas y a la vanguardia en el uso de las herramientas y técnicas modernas. Una muestra de esto, según aporte de Marr (2018) en el sitio web de Forbes, es el crecimiento inmensurable de información y la velocidad con la que se genera, por ejemplo, se estima que el volumen total de datos en el mundo creados por día es de 2,5 quintillones de bytes,

adicionalmente, se conoce que más del 90% de los datos que existen han sido creados en los últimos dos años.

Este fenómeno, se debe a algunos factores, por ejemplo, la aparición de dispositivos inteligentes que permiten el acceso a la información, a realizar consultas, a ejecutar transacciones, etc., así como la facilidad de obtención de estos dispositivos que genera mayor cantidad de internautas por día (Domo, 2018).

El mantener a la innovación como pilar empresarial, se ha convertido en una necesidad para las empresas, ya que se genera ventajas competitivas frente a otras instituciones.

El ciclo de venta del bien o servicio que brindan las empresas no se quedaría únicamente en la comercialización, sino en añadir valor y fidelizar a los clientes, ofrecer productos de acuerdo a sus expectativas, garantizando transacciones seguras y mejorando la experiencia del cliente. Esto lo corroboran diversos estudios realizados por investigadores de estas nuevas tecnologías disruptivas. López(2012) y Punyol(2014), mencionan que los procesos deben estar enfocados en mejorar continuamente, empleando herramientas que permitan, en base al conocimiento de la información, generar propuestas de valor y facilitar la toma de decisiones encaminados a cumplir los objetivos estratégicos de la entidad.

Realizando una recopilación de los conceptos previamente mencionados, se podría considerar a la ciencia de datos como un proceso de ayuda idónea para gestionar y aprovechar toda la información que se está generando diariamente y que no necesariamente se almacena en formatos estructurados convencionales. Este proceso consiste en la extracción de la información de diferentes fuentes, la realización de una minería de datos (limpieza de la información), el

procesamiento de la información a través de métodos matemáticos y/o de aprendizaje de máquina y convertirlo en conocimiento que se verán reflejados en reportes empresariales que sirvan para generar decisiones estratégicas de valor (Provost & Fawcett, 2013).

Cabe mencionar que la ciencia de datos no está enfocada únicamente a segmentación de clientes, sino también a optimización de procesos, rendimientos de máquinas y dispositivos, mejoras en seguridad y salud, cómo, por ejemplo, el trabajo de investigación de Anchiraico (2017), en donde realizó el "Diseño de una arquitectura Big Data para la predicción de crisis en el trastorno bipolar", en donde se pretende, en función de diferentes síntomas (variables), pronosticar y detectar este trastorno con tiempo. Es así como se comprobó que la ciencia de datos puede estar presente en diferentes campos de acción.

El análisis de los beneficios que se presentan al incursionar en esta nueva tecnología, ha causado que las empresas opten por implementarlas. Según Galimany (2014) existen varios trabajos de investigación relacionados al desarrollo del proceso de ciencia de datos, en los mismos se han reconocido aspectos importantes a tomarse en cuenta para su implementación, como la arquitectura, el volumen de información, las herramientas en las que se puede implementar y principalmente el valor que se genera al explotarla. Este último punto puede llegar a ser el aspecto más importante a evaluar, puesto que se debe aterrizar a la naturaleza de la actividad de la empresa para la implementación del proceso.

Serrano (2017) afirma en un artículo de la revista Vistazo publicado en septiembre 2017, que Big Data nace en el mundo en el año 2000, sin embargo, en Ecuador se empiezan a ver proyectos relacionados al análisis de información a partir del año 2010. De acuerdo a los registros de la Asociación Ecuatoriana de Software (AESOFT) existen alrededor de 20

empresas a nivel nacional dedicadas a ofrecer servicios de Inteligencia de Negocio, no obstante, según Serrano (2017), en Ecuador no se ha adoptado todavía un número considerable de estas nuevas tecnologías, lo que genera un reto para quienes la implementan.

Emplear este tipo de tecnologías sin duda genera muchos beneficios, sin embargo, su implementación en las empresas genera grandes desafíos. En este contexto, se puede tomar en consideración la opinión del grupo de trabajo Working Party(2014), en la que se mencionan los principales retos a los que debería hacer frente una organización que quisiese desarrollar una estrategia de implantación de ciencia de datos y, entre ellos, enumera la pérdida de control sobre la información personal, la necesidad de regular la prestación de servicios tecnológicos asociados a estos fenómenos por parte de terceros, las limitaciones a la posibilidad de que el usuario del servicio no resulte identificado o identificable, la posible monitorización intrusiva, así como la calidad del consentimiento del afectado y las medidas de seguridad encaminadas a evitar la alteración, pérdida, tratamiento o acceso no autorizado de la información personal.

Las instituciones financieras no están lejos de esta realidad, un informe de BBVA(2017), menciona que, en la actualidad, sólo el 15% de control de riesgo bancario se realiza a través de la analítica y se estima que el porcentaje ascenderá a 40% para el año 2025. Bajo esta premisa, se considera que las instituciones financieras deberían cambiar su visión hacia la innovación, con el objetivo de ajustarse a los nuevos retos y ganar ventaja competitiva:

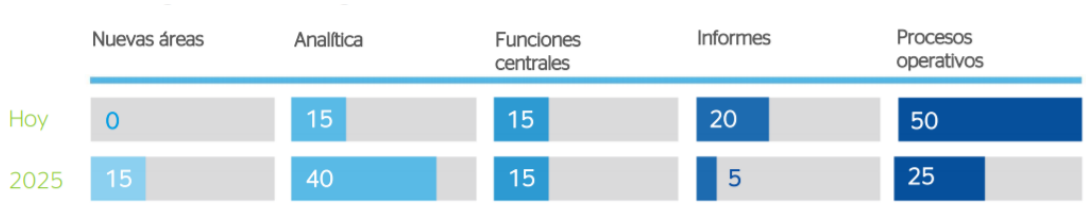

Figura 9. Perfiles en la gestión de riesgo bancario. **Fuente:** BBVA (2017)

Por otro lado, este tipo de entidades son afectadas por un sinnúmero de temas relacionados a seguridad, por lo que se implementan controles para facilitar que los riesgos se mitiguen, para este objetivo se crean políticas, procedimientos con la descripción de estos controles los cuáles necesitan ser evaluados, con el objetivo de verificar que se encuentren correctamente implementados. Ante esta necesidad existen buenas prácticas y metodologías que guían en la creación de controles adecuados.

Existe escasa información hacia un proceso que compruebe que se cumplen las expectativas de los controles implementados dentro del proceso de ciencia de datos o que las políticas de este son actualizadas de acuerdo con las necesidades de las partes interesadas de la organización. De acuerdo a las investigaciones realizadas se concluye en que en la mayoría de metodologías se proponen estándares para ser adaptados a los sistemas de información que existan dentro de una organización a nivel general, no existe nada específico. Esta afirmación se corrobora en "Trabajos futuros" del proyecto de investigación realizado por García (2017), cuyo tema es "Metodología para la auditoría de Sistemas" y menciona que se debería establecer estándares y normas específicas para sistemas Big Data, en materia de implementación y evaluación.

Con respecto a las metodologías que se emplearon, se identificaron algunas buenas prácticas en el trabajo de investigación realizado por García(2017), y se determinó que existen varias buenas prácticas cuyo objetivo es garantizar la seguridad de la data. García(2017) afirma que si bien los principios de estas metodologías se utilizan de manera general sobre cualquier contexto de una empresa, no existe un modelo de revisión exclusivo para la ciencia de datos. Por lo tanto,

se decidió investigar metodologías como COBIT, GTAG y NIST que permitan el análisis de la realidad de la empresa y su respectiva adaptación.

Adicionalmente, otro motivo por los que se decide trabajar con COBIT, es por la disposición general de la Superintendencia de Bancos (2013), en su artículo 20 de las Normas Generales para las instituciones del Sistema Financiero, que menciona que se deben tomar en consideración las directrices de auditoría previstas por ISACA.

2.4. Marco conceptual

- **Ciencia de datos:** La ciencia de datos ese trata de un ámbito que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, combinando disciplinas como la estadística, la minería de datos, el aprendizaje automático y la analítica predictiva. (Liu, 2018).
- **CISA:** La Certificación de Auditor de Sistemas de Información (Certified Information Systems Auditor), es la principal Certificación de ISACA, desde 1978. (ISACA, 2015). La acreditación CISA es una certificación reconocida universalmente para profesionales en auditoría, control y seguridad de SI.
- **COBIT:** COBIT fue creado para ayudar a las organizaciones a obtener el valor óptimo de TI manteniendo un balance entre la realización de beneficios, la utilización de recursos y los niveles de riesgo asumidos. COBIT 5 posibilita que TI sea gobernada y gestionada en forma holística para toda la organización, tomando en consideración el negocio y áreas funcionales de punta a punta, así como los interesados internos y externos. COBIT 5 se puede aplicar a organizaciones de todos los tamaños, tanto en el sector privado, público o entidades sin fines de lucro (ISACA, 2012).

- **NIST:** El Instituto Nacional de Normas y Tecnología (NIST por sus siglas en inglés, National Institute of Standards and Technology) es una agencia de la Administración de Tecnología del Departamento de Comercio de los Estados Unidos. La misión de este instituto es promover la innovación y la competencia industrial mediante avances en metrología, normas y tecnología de forma que mejoren la estabilidad económica y la calidad de vida. (NIST, 2000).

- **GTAG:** Las Guías de Auditoría de Tecnología Global publicadas por el IIA son parte de las Guías para la Práctica, que a su vez son un componente fuertemente recomendado para su aplicación, dentro del Marco para la Práctica Profesional de la Auditoría Interna (IPPF por sus siglas en inglés). Identifican diversos niveles de conocimiento de TI necesarios en toda la organización para proporcionar un "enfoque sistemático y disciplinado que permita evaluar y mejorar la eficacia de los procesos de gestión de riesgos, control y gobierno". Las publicaciones GTAG destacan que los conocimientos sobre cómo se utilizan las TI, sus riesgos asociados y la capacidad de utilizar las TI como recursos son esenciales para que el auditor interno sea eficaz en todos los niveles.

2.4.1. Adopción de una perspectiva teórica

Entre las buenas prácticas investigadas se han identificado las siguientes que se utilizarán como base para el diseño de la metodología de evaluación, las cuales se considera, aportarán a la ejecución del objetivo de este proyecto:

- **CISA:** Es una certificación reconocida universalmente para profesionales en auditoría, control y seguridad de TI. Esta certificación se divide en 5 dominios, sin embargo, para la presente metodología estará enfocado al primer dominio: "El proceso de auditoría

de sistemas de información", el cual establecerá directrices para llevar a cabo una auditoría durante todo el proceso de evaluación.

- **GTAG:** Según The Institute of Internal Auditors (2017) en su artículo "Understanding and Auditing Big Data" publicado en sus guías GTAG, las principales áreas de riesgo que deben ser analizadas en un sistema Big Data son:
    - Gobierno del programa de Big Data
    - Disponibilidad de la tecnología y performance
    - Seguridad y privacidad
    - Calidad de los datos, manejo y reporte

  Los mismos que surgen de factores tanto internos como externos a la organización y requieren ser identificados y controlados de acuerdo a la realidad de la institución y del país.

- **COBIT:** Es un conjunto de buenas prácticas que proporciona una visión integral y sistemática del gobierno y la gestión de las TI basada en objetivos de control lo que permite asegurar que las TI y el negocio están alineados. En el presente trabajo se identificarán los procesos que apliquen a la institución financiera privada para llevar a cabo la función de auditoría.
- **NIST:** De las 7 publicaciones relacionadas al Big Data, la metodología se apalancará en el módulo referente a la Arquitectura de Big Data.

CAPÍTULO III. ANÁLISIS SITUACIONAL

3.1. Uso de ciencia de datos.

Según Galimany (2014) La ciencia de datos se torna un proceso necesario para mantenerse a la vanguardia del mundo actual, ya que no sólo permite procesar la información, sino también sacar provecho mediante análisis predictivo y optimizar la toma de decisiones.

Ante esta nueva tendencia las instituciones han optado por obtener los recursos para levantar un adecuado proceso de ciencia de datos, lo que involucra mantener un entendimiento de los componentes necesarios. A pesar de que, en los países europeos, esta no es una tecnología nueva, los países latinoamericanos están incursionando en implementarla y no cabe duda que la experiencia de los otros países fortalecerá el proceso que se está iniciando en los países latinos, entre alguno de los temas importantes se pueden resaltar: el manejo de la capacidad de almacenamiento, el procesamiento y el número de decisiones tomadas automáticamente, en base a modelos proporcionados por el proceso de ciencia de datos, etc.

De acuerdo a entrevistas mantenidas con la gerencia de Ciencia de datos de la institución del caso de estudio, se conoce que la ciencia de datos está permitiendo aprovechar la minería de datos y el modelado predictivo para personalizar ofertas, reducir riesgos, crear nuevos productos disruptivos, expandir mercados, minimizar gastos operativos, automatizar procesos tradicionalmente manuales, etc. Sin embargo, en el sector financiero, esta revolución viene a sumarse a un entorno complejo, que combina elementos coyunturales y de gobierno, con una fuerte

presión regulatoria. Adicionalmente impacta en diferentes ámbitos de una entidad, pasando por las políticas y procedimientos internos, hasta las herramientas y los sistemas de información por el cual fluye la data.

Frente a esta realidad se identifica que existen una cantidad de componentes que requieren ser evaluados en este proceso, con el objetivo de garantizar que los esfuerzos involucrados en su implementación y puesta en marcha, están cosechando frutos y se está obteniendo información de valor que apalanque la toma de decisiones y fortalezca la ejecución de los objetivos estratégicos de la institución.

3.1. Metodologías existentes

3.1.1. Metodologías para ciencia de datos

Se han realizado investigaciones acerca de buenas prácticas que permitan evaluar a la ciencia de datos, sin embargo, no se obtuvieron resultados en la búsqueda de una metodología específica y completa para este proceso, lo que genera el producto de la presente investigación.

Adicionalmente, dentro de las normativas vigentes que norman a las instituciones financieras no se han levantado temas referentes a los procesos de ciencia de datos, pero si hacen referencia a la seguridad que se debe mantener en torno a la información que se almacena, transmite y procesa de los clientes. Al ser una institución financiera, se debe cumplir con las disposiciones de los entes de control, por lo que la metodología planteada para auditar al proceso de ciencia de datos en el Banco, tomará en cuenta a la Normativa 771.

También se indagó en las normativas de la Superintendencia de Bancos(2013) por sugerencias de buenas prácticas internacionales en las cuales se pueda apoyar la metodología y

se encontró que en la las Normas Generales para las instituciones del Sistema Financiero se menciona que las auditorías informáticas deben considerar las directrices sugeridas por ISACA, motivo por el cual se escogió a COBIT, al ser un marco de trabajo que se enfoca en comprender el gobierno y la gestión de las tecnologías de la información aplicadas a todo nivel en la organización.

COBIT no es un marco de trabajo que abarque temas técnicos específicos de arquitectura o infraestructura, ya que los controles sobre estos temas son abordados de forma general, razón por la cual fue necesario investigar otra buena práctica que abarque estos temas, encontrándose aportes relacionados a Big Data como son la NIST y GTAG.

El Instituto Nacional de estándares y Tecnología (NIST), desarrolló un framework con 7 módulos para comprender y gestionar el entorno Big Data. De estos módulos, se tomó el número 5, el cual da a conocer cómo debería implementarse una arquitectura que soporta a este tipo de tecnologías, aspecto que complementa el análisis que se realizará bajo COBIT.

Por último, se identificó la existencia de una publicación de las Guías de Auditoría de Tecnología Global (GTAG), denominada “Understanding and Auditing Big Data”, en dónde se abarcan temas relacionados a la gestión de datos, así como los riesgos y desafíos que se pueden presentar al utilizar esta tecnología, razón por la cual se tomarán en cuenta los argumentos sugeridos en esta publicación para el desarrollo de la metodología.

3.1.2. Metodologías para Auditoría

Tomando en cuenta a las normativas de la Superintendencia de Bancos (2013) y su disposición de basarse en ISACA, se ha seleccionado al primer dominio de la certificación CISA: "El proceso de auditoría de sistemas de información", en donde se encuentran los lineamientos y directrices para efectuar una evaluación y levantar un programa de auditoría.

La explicación realizada en el apartado anterior, se resume en el siguiente gráfico, en dónde se puntualizan las buenas prácticas y normativa en las que se basará la metodología planteada:

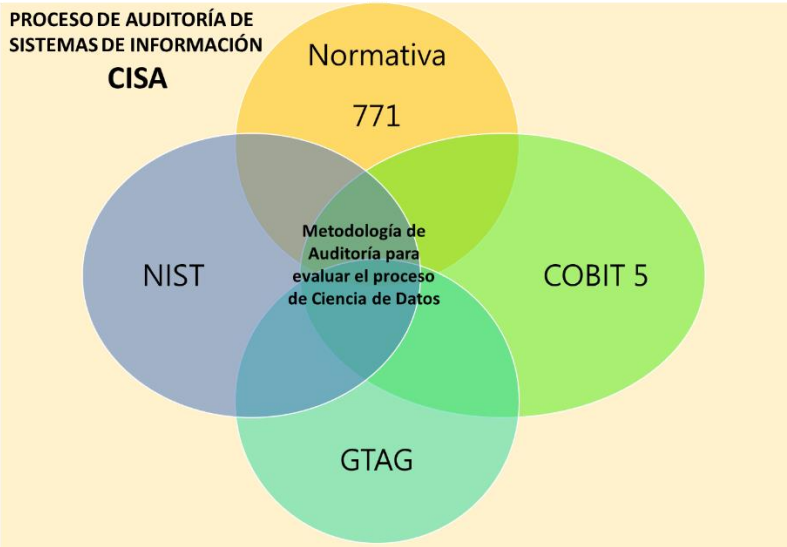

Figura 10. Metodología de Auditoría para evaluar el proceso de ciencia de datos. **Fuente:** Elaborado por autor

3.1.3. Mapeo de metodologías

A continuación, se detalla el procedimiento realizado para obtener el mapeo de los procesos basados en COBIT, GTAG, NIST y la normativa, partiendo de los dominios internos de revisión de la institución, para el proceso de ciencia de datos:

- **Identificación de marcos de referencia y buenas prácticas**

  En base al análisis de diferentes buenas prácticas definir las que apliquen para el desarrollo de la metodología planteada. Se ha definido utilizar: COBIT, GTAG, NIST y la Normativa 771 de Riesgo Operativo.

- **Identificación de dominios internos estándar utilizados actualmente para revisiones de Auditoría.**

  En la Tabla 1, se indica los dominios de evaluación con los que cuenta la institución dentro de su metodología interna:

Tabla 1
Dominios para revisiones de Auditoría Interna

Dominio	Descripción
Gobierno de datos	Se refiere a la gestión de la información que se maneje en el proceso y su alineación con los objetivos de la institución.
Selección y capacitación de personal	Controles que se lleven a cabo para garantizar un adecuado proceso de contratación y capacitación de personal, de acuerdo a las necesidades del área.
Seguridad de infraestructura	Análisis de la infraestructura y sus componentes. Adicionalmente se verifica el análisis de capacidad de acuerdo a las necesidades del área
Seguridad lógica	Referente a garantizar el acceso a los componentes de hardware y software a personal autorizado con el concepto de menor privilegio.
Gestión de requerimientos de desarrollo	Hace referencia a un adecuado proceso de gestión de desarrollo de los requerimientos solicitados por el cliente.
Gestión de operaciones	Dominio que garantiza la disponibilidad del proceso de ciencia de datos a través de controles relacionados a: monitoreo de ejecución de procesos batch, gestión de incidentes, gestión de respaldos y gestión de la continuidad y gestión de contratos de servicios externos.
Controles de aplicación	Controles que garanticen que la lógica funciona de acuerdo a lo requerido, garantizando la integridad y veracidad de la información que se genera del proceso.

Fuente: Elaborado por autor

- **Mapeo de buenas prácticas de GTAG con los dominios internos estándar de la institución**

  En la Tabla 2 se establece un mapeo entre los procesos internos de la institución con los que se relaciona la documentación de GTAG. Cabe mencionar que se seleccionaron todos los temas sugeridos por la GTAG:

  - Gobierno del programa de Big Data
  - Disponibilidad de la tecnología y performance
  - Seguridad y privacidad
  - Calidad de los datos, manejo y reporte

Tabla 2
Mapeo dominios internos y GTAG

Proceso interno de la institución	**Proceso GTAG**
Gobierno de datos	Gobierno del programa de Big Data
Selección y capacitación de personal	-
Seguridad de infraestructura	Disponibilidad de la tecnología y performance
Gestión de operaciones	
Seguridad lógica	Seguridad y privacidad
Gestión de requerimientos de desarrollo	-
Controles de aplicación	Calidad de información

Fuente: Elaborado por autor

- **Selección de procesos de COBIT5**

  En base a la investigación y al mapeo levantado en el punto anterior, en la Tabla 3 se seleccionan los procesos COBIT5 que guarden relación con la metodología propuesta y su respectiva justificación:

Tabla 3
Selección de procesos COBIT

Dominio	Proceso	Justificación
Alinear, planificar y organizar	APO01 - Gestionar el Marco de Gestión de TI	Se ha seleccionado este proceso debido a la importancia de garantizar la alineación de los objetivos estratégicos de la institución con los objetivos del proceso de ciencia de datos. Asimismo, este proceso de COBIT se enfoca en mantener políticas y procedimientos que garanticen la adecuada gestión de información.
Alinear, planificar y organizar	APO04 - Gestionar la Innovación	Investiga tecnologías emergentes, adoptar este conocimiento a la institución y realizar seguimientos para conocer su evolución. Se define la capacidad de la institución para adoptar un proceso de ciencia de datos.
Alinear, planificar y organizar	APO07 - Gestionar los Recursos Humanos	El proceso de ciencia de datos hace especial énfasis en mantener personal capacitado en diferentes áreas de conocimiento, para generar valor dentro del proceso.
Alinear, planificar y organizar	APO10 Gestionar los Proveedores	Se revisará que exista una adecuada gestión de contratos con los proveedores con los que se mantenga servicios externos.
Construir, Adquirir e implementar	BAI04 - Gestionar la capacidad y la disponibilidad	Es necesario evaluar la capacidad y rendimiento a nivel de recursos para soportar el proceso de ciencia de datos, con el objetivo de mantener los servicios disponibles y prever requerimientos futuros.
Construir, Adquirir e implementar	BAI01 - Gestión de Programas y Proyectos BAI06 - Gestionar los Cambios BAI07 - Gestionar la Aceptación del Cambio y la Transición	Para llevar a cabo el proceso de ciencia de datos es necesario mantener un adecuado proceso de Gestión de requerimientos de desarrollo, para asegurar que los mismos se cumplan conforme a lo solicitado.
Construir, Adquirir e implementar	BAI08 - Gestionar el conocimiento	Este proceso está directamente ligado a la ciencia de datos, ya que menciona la importancia de gestionar el conocimiento obtenido a través del análisis de la data. Con este proceso se garantiza que la información es fiable y

Dominio	Proceso	Justificación
		con ella se puede tomar decisiones que favorezcan los objetivos estratégicos de la institución.
Entregar, dar servicio y soporte	DSS01 - Gestionar Operaciones	Relacionado al monitoreo de ejecución de actividades operativas del proceso tales como: ejecución de procesos batch y en línea, monitoreo de infraestructura e instalaciones. Garantizar un plan de acción en línea frente a posibles alertas. Adicionalmente, se hace mención a la gestión de contratos ante posibles servicios externalizados.
Entregar, dar servicio y soporte	DSS02 - Gestionar las Peticiones y los Incidentes del Servicio	Proceso enfocado en gestionar los incidentes que pueden presentarse. Se habla de monitoreo y controles para garantizar la disponibilidad de la información.
Entregar, dar servicio y soporte	DSS04 - Gestionar la Continuidad	Control que se enfoca en la gestión sobre eventos y los planes de acción levantados ante los mismos. Adicionalmente, se analiza el grado de preparación de la institución frente a un evento contingente, los escenarios identificados y los planes de continuidad y contingencia levantados.
Entregar, dar Servicio y Soporte	DSS05 - Gestionar los Servicios de Seguridad	Se selecciona este proceso debido a la importancia de la gestión de seguridad de la información que se manipula, analiza y almacena dentro del proceso de ciencia de datos. Se verifica específicamente el proceso definido para otorgar accesos y privilegios a la información.
Entregar, dar Servicio y Soporte	DSS06 - Gestionar Controles de Proceso de Negocio	Control enfocado a verificar que la información que se procesa dentro de ciencia de datos se ejecuta en función al requerimiento solicitado. Verifica la integridad y calidad de la información resultante.

Fuente: Elaborado por autor

- **Mapeo de buenas prácticas de GTAG, procesos internos de la institución y COBIT5**

  En la Tabla 4 se establece un mapeo entre los procesos internos de la institución con los procesos seleccionados de las GTAG y COBIT5.

Tabla 4
Mapeo procesos internos, GTAG y COBIT.

Proceso interno de la institución	**Proceso GTAG**	**Proceso COBIT**
Gobierno de datos	Gobierno del programa de Big Data	APO01 - Gestionar el Marco de Gestión de TI APO04 - Gestionar la Innovación
Selección y capacitación de personal	-	APO07 - Gestionar los Recursos Humanos
Seguridad de infraestructura	Disponibilidad de la tecnología y performance	BAI04 - Gestionar la capacidad y la disponibilidad DSS01 - Gestionar Operaciones
Seguridad lógica	Seguridad y privacidad	DSS05 - Gestionar los Servicios de Seguridad
Gestión de requerimiento de desarrollo	-	BAI01 Gestión de Programas y Proyectos BAI06 - Gestionar los Cambios BAI07 - Gestionar la Aceptación del Cambio y la Transición
Gestión de operaciones	Disponibilidad de la tecnología y performance	APO10 Gestionar los Proveedores DSS01 - Gestionar Operaciones DSS02 - Gestionar las Peticiones y los Incidentes del Servicio DSS04 - Gestionar la Continuidad
Controles de aplicación	Calidad de información	DSS06 - Gestionar Controles de Proceso de Negocio BAI08 - Gestionar el conocimiento

Fuente: Elaborado por autor

- **Selección de procesos a evaluar de la Normativa de la Superintendencia No. SB-2018-771**

  En base a la investigación y al mapeo levantado en el punto anterior, en la Tabla 5 se seleccionan los procesos relacionados con la metodología propuesta.

Tabla 5
Selección de procesos Normativa de la Superintendencia No. SB-2018-771

Sección	Artículo	Tema	Justificación
III	Artículo 10 Literal b: Personas	Personas Permanencia de personal	Aspecto relacionado a garantizar el compromiso de la institución por perfeccionar las habilidades del personal.
III	Artículo 10 Literal c: Tecnología de la Información Subliteral ii	iii. Políticas y procedimientos – Alineación estratégica	Relacionado a mantener políticas y procedimientos alineados a la institución.
III	Artículo 10 Literal c: Tecnología de la Información Subliteral iii	iii. Gestión de operaciones	Tienen relación con la gestión de incidentes, respaldos y operaciones del servicio.
III	Artículo 10 Literal c: Tecnología de la Información Subliteral iv	iv. Gestión de desarrollo	Tienen relación con la gestión de desarrollo, tema que servirá para analizar la implementación de requerimientos de desarrollo
III	Artículo 10 Literal c: Tecnología de la Información Subliteral v	v. Infraestructura	Tienen relación con la gestión de la infraestructura que soporta al proceso de ciencia de datos

Sección	Artículo	Tema	Justificación
III	Artículo 10 Literal d: Eventos externos Subliteral ii	ii. Plan de Continuidad	Se debe contar con un plan de acción ante posibles escenarios que se puedan presentar y afecten a la disponibilidad del proceso de ciencia de datos.
V	Artículo 13 Literal a, b y c	Administración de proyectos	Relacionado a la gestión sobre los proyectos que se levanten dentro del proceso de ciencia de datos.
VI	Artículo14 Literal b y f	Servicios provistos por terceros	Relacionado a contratos que se levanten por posibles servicios externalizados.
VII	Artículo 16 Literal j: Protección de información Subliteral i, ii, iii, iv, vii, ix	Seguridad de la Información	Se deben tomar en cuenta los controles necesarios para garantizar el adecuado acceso a la información analizada en el proceso de ciencia de datos. Adicionalmente, en estos apartados se menciona la importancia de ejecutar análisis de riesgos sobre la infraestructura.

Fuente: Elaborado por autor

- **Mapeo de buenas prácticas de GTAG, procesos internos de la institución, COBIT5 y Normativa de la Superintendencia No. SB-2018-771**

  En la Tabla 6 se establece un mapeo entre los procesos internos de la institución con los procesos seleccionados de las GTAG, COBIT5 y la Normativa de la Superintendencia No. SB-2018-771.

Tabla 6
Mapeo procesos internos, GTAG, COBIT y Normativa

Proceso interno de la institución	Proceso GTAG	Proceso COBIT	Normativa
Gobierno de datos	Gobierno del programa de Big Data	APO01 - Gestionar el Marco de Gestión de TI APO04 - Gestionar la Innovación	Artículo 10 - Literal c: Tecnología de la Información – Subliteral ii - Políticas y procedimientos – Alineación estratégica
Selección y capacitación del personal	-	APO01 - Gestionar el Marco de Gestión de TI APO07 - Gestionar los Recursos Humanos	Artículo 10 - Literal b: Personas
Seguridad de infraestructura	Disponibilidad de la tecnología y performance	BAI04 - Gestionar la capacidad y la disponibilidad DSS01 - Gestionar Operaciones	Artículo 10 - Literal c: Tecnología de la Información - Subliteral v - Infraestructura
Seguridad lógica	Seguridad y privacidad	DSS05 - Gestionar los Servicios de Seguridad	Artículo 16: Literal j: Protección de información - Seguridad de la Información
Gestión de requerimiento de desarrollo	-	BAI01 Gestión de Programas y Proyectos BAI06 - Gestionar los Cambios BAI07 - Gestionar la Aceptación del Cambio y la Transición	Artículo 10 - Literal c: Tecnología de la Información - Subliteral iv Artículo 13 - Literal a, b y c: Administración de proyectos
Gestión de operaciones	Disponibilidad de la tecnología y performance	APO10 - Gestionar los Proveedores DSS01 - Gestionar Operaciones	Artículo 14: Servicios provistos por terceros. Artículo 10- Literal c: Tecnología de la

Proceso interno de la institución	Proceso GTAG	Proceso COBIT	Normativa
		DSS02 - Gestionar las Peticiones y los Incidentes del Servicio DSS04 - Gestionar la Continuidad	Información - Subliteral iii Artículo 10 - Literal d: Eventos externos - Plan de Continuidad
Controles de aplicación	Calidad de información	DSS06 - Gestionar Controles de Proceso de Negocio BAI08 - Gestionar el conocimiento	-

Fuente: Elaborado por autor

- **Selección de componentes de la arquitectura de Big Data del Framework NIST**

En base a la investigación y al mapeo levantado en el punto anterior, en la Tabla 7 se seleccionan los componentes de arquitectura relacionados con la metodología:

Tabla 7
Selección de componentes del Framework de Arquitectura de BIG DATA propuesto por la NIST

Componente de arquitectura	Justificación
Sistema de orquestación	Se selecciona este componente debido a que se requiere mantener un control sobre los requerimientos levantados inicialmente y su cumplimiento. Se considera importante verificar de donde provienen los requerimientos y las autorizaciones respectivas.
Proveedor de datos	Aspecto correspondiente a las fuentes de donde se extraerá la información. Es necesario evaluar el método de extracción de información y verificar su integridad cuando llegue la información al destino.
Proveedor de aplicaciones Big Data	Componente correspondiente al desarrollo de la funcionalidad y lógica de negocio solicitada en los requerimientos iniciales. Se considera importante evaluar los controles levantados para garantizar que los desarrollos respondan a los requerimientos iniciales.

Proveedor de framework de Big Data	Aborda los recursos tecnológicos necesarios para procesar información en este tipo de tecnologías. Aspecto de revisión indispensable, ya que se analizan los componentes tecnológicos de soporte al proceso.
Consumidor de datos	Componente final que obtiene los resultados levantados durante el procesamiento. Se requiere revisar este componente debido a la necesidad de mantener un aseguramiento de la veracidad de la información.

Fuente: Elaborado por autor

- **Mapeo de buenas prácticas de GTAG, procesos internos de la institución, COBIT5, Normativa de la Superintendencia No. SB-2018-771 y NIST**

  En la Tabla 8, se establece un mapeo entre los procesos internos de la institución con los procesos seleccionados de las GTAG, COBIT5, la Normativa de la Superintendencia No. SB-2018-771 y NIST.

Tabla 8
Mapeo procesos internos, GTAG, COBIT, Normativa y NIST

Proceso interno de la institución	**Proceso GTAG**	**Proceso COBIT**	**Normativa**	**NIST**
Gobierno de datos	Gobierno del programa d Big Data	APO01 - Gestionar el Marco de Gestión de TI APO04 - Gestionar la Innovación	Artículo 10 - Literal c: Tecnología de la Información – Subliteral ii - Políticas y procedimientos – Alineación estratégica	Sistema de orquestación
Selección y capacitación del personal	-	APO01 - Gestionar el Marco de Gestión de TI APO07 - Gestionar los Recursos Humanos	Artículo 10 - Literal b: Personas	-

Seguridad de infraestructura	Disponibilidad de la tecnología y performance	BAI04 - Gestionar la capacidad y la disponibilidad DSS01 - Gestionar Operaciones	Artículo 10 - Literal c: Tecnología de la Información - Subliteral v – Infraestructura	Proveedor de framework de Big Data
Seguridad lógica	Seguridad y privacidad	DSS05 - Gestionar los Servicios de Seguridad	Artículo 16: Literal j: Protección de información - Seguridad de la Información	Proveedor de aplicaciones Big Data
Gestión de requerimiento de desarrollo	-	BAI01 Gestión de Programas y Proyectos BAI06 - Gestionar los Cambios BAI07 - Gestionar la Aceptación del Cambio y la Transición	Artículo 10 - Literal c: Tecnología de la Información - Subliteral iv Artículo 13 - Literal a, b y c: Administración de proyectos	-
Gestión de operaciones	Disponibilidad de la tecnología y performance	APO10 - Gestionar los Proveedores DSS01 - Gestionar Operaciones DSS02 - Gestionar las Peticiones y los Incidentes del Servicio DSS04 - Gestionar la Continuidad	Artículo 14: Servicios provistos por terceros. Artículo 10- Literal c: Tecnología de la Información - Subliteral iii Artículo 10 - Literal d: Eventos externos - Plan de Continuidad	-

Controles de aplicación	Calidad de información	DSS06 - Gestionar Controles de Proceso de Negocio BAI08 - Gestionar el conocimiento	-	Consumidor de datos Proveedor de datos

Fuente: Elaborado por autor

CAPÍTULO IV. PROPUESTA

En base al análisis del estado situacional de la institución, se ha levantado una metodología basada en mejores prácticas aceptadas y reconocidas a nivel internacional, con la cual se podrá ejecutar una evaluación del proceso de ciencia de datos, para identificar principales riesgos y emitir un informe de resultados con recomendaciones de valor.

El objeto del presente trabajo está enfocado en determinar una metodología estándar para identificar los componentes esenciales a ser evaluados en una auditoría del proceso de ciencia de datos, con el objetivo de garantizar la existencia de controles sobre la seguridad e integridad de la información.

Cabe mencionar que la metodología se concentrará en la fase de planificación debido a que las fases de ejecución e informe final no se encuentran dentro del alcance del presente trabajo, además por temas de confidencialidad de la institución, no pueden ser expuestas.

La metodología abarca el definir claramente el objetivo, el alcance y el programa de auditoría a emplearse en la revisión.

4.1. Objetivo de la revisión

El objetivo de la revisión es determinar el qué, cómo y para qué se deberá realizar una evaluación a este proceso. Para el caso de esta metodología se ha definido el siguiente objetivo: Ejecutar una revisión de auditoría al proceso de ciencia de datos, a través de la aplicación de una

metodología específica, para garantizar la existencia de controles de seguridad e integridad de la información, a través de buenas prácticas reconocidas internacionalmente.

4.2. Alcance de la revisión

El alcance está asociado a identificar los procesos que serán objeto de revisión y prueba, por lo que es necesario conocer la razón de ser del área de ciencia de datos, identificar el flujo por el que atraviesa la información en el proceso y finalmente identificar los controles establecidos en cada fase.

Adicionalmente, el alcance se refiere al corte, es decir en qué periodo de tiempo se enfocará la revisión. Este aspecto se deberá definir en función de la necesidad de la institución. Como referencia de corte, podría limitarse la revisión de por lo menos un año lectivo.

4.3. Creación de plan de auditoría

Para levantar un adecuado plan o programa de auditoría, se ha tomado como referencia los temas señalados por ISACA(2018) en su artículo "*Herramientas y Técnicas para la Creación de Programas de Auditoría"*:

1. Determinar la materia de auditoría
2. Definir el objetivo de la auditoría
3. Establecer el alcance de la auditoría
4. Realizar una planificación previa de la auditoría
5. Determinar los procedimientos de auditoría y los pasos para la recopilación de datos

Cabe mencionar que los puntos 2 y 3 ya se abordaron en los numerales 4.1 y 4.2 del presente capítulo.

4.3.1. Determinar la materia de la auditoría

Se determinan las áreas inmersas parcial o totalmente dentro del proceso de ciencia de datos, a continuación, se presenta una tabla con las posibles áreas a evaluar y los respectivos temas a tratar:

Tabla 9
Áreas involucradas en la revisión - Temas a tratar

Área	Temas a tratar
Ciencia de datos	• Plan operativo • Estructura del área • Gestión de requerimiento de desarrollo • Gestión de calidad • Planes de continuidad y contingencia
Tecnología	• Gestión de infraestructura • Gestión de software • Gestión de respaldos • Gestión de contratos
Seguridad de la Información	• Política de seguridad de la información • Procedimientos de accesos a la información • Análisis de riesgos • Análisis de vulnerabilidades
Riesgos	Ejecución de pruebas en escenarios de continuidad y contingencia
Talento Humano	• Procesos de selección de personal • Programas de capacitación y especialización de personal

Fuente: Elaborado por autor

Se mantienen reuniones iniciales con las gerencias de cada una de las áreas mencionadas anteriormente, para explicar el objetivo y el alcance de la revisión de auditoría e indagar en los temas propuestos. Adicionalmente, se efectúa el entendimiento de las actividades realizadas

para soportar el proceso de ciencia de datos. Se hace uso de algunas técnicas de investigación como la observación, las entrevistas y cuestionarios con preguntas genéricas acerca del proceso.

4.3.1.1. Área de ciencia de datos.

Es necesario identificar la razón de ser del área, cuál es su planificación operativa, cuáles son sus actividades principales y como todo esto se ha alineado al cumplimiento de los objetivos estratégicos de la institución. El dominio de gobierno de datos toma relevancia en este punto, ya que se verificará que exista concordancia entre lo que se realiza en el área y su alineación a los objetivos de la institución financiera. Adicionalmente, se garantizará un entendimiento y compromiso total por parte de la Alta Gerencia, con respecto a los resultados finales de la ciencia de datos, sus beneficios y generación de valor, para que se apoye con un adecuado financiamiento a los recursos que soporten el proceso de ciencia de datos. Asimismo, se conocerá los riesgos en los que esta inmiscuido el proceso y levantar, en la medida de lo posible, un plan de respuesta ante la posible materialización de un riesgo.

El área de ciencia de datos también debe mantener políticas y procedimientos, con el objetivo de dar a conocer qué y cómo se realizan las actividades en el área y de esta manera controlar los procesos asociados a la generación de los entregables y resultados.

Otro tema importante a considerarse, es la selección del personal para el área de Ciencia de datos, ya que es necesario contar con especialistas competentes para ejecutar las actividades en este proceso. Es imprescindible que, en el perfil profesional del personal de Ciencia de datos, se destaquen habilidades matemáticas, estadísticas, de programación y de conocimiento del negocio, de tal manera que sus funciones en el área, generen valor. Adicionalmente, la capacitación constante y permanente del personal es un punto de suma

importancia, ya que se asegura la actualización de conocimientos en una tecnología que cambia constantemente. Estos temas se considerarán dentro del apartado de Selección y capacitación de personal.

Por último, es necesario conocer el cumplimiento de los objetivos planteados por el área, razón por la cual se debe considerar levantar métricas que permitan medir el desempeño de las actividades realizadas. Mantener métricas no sólo le servirá a la Gerencia del área para conocer el nivel de cumplimiento de objetivos, sino también, se comunicará de una manera más dinámica y transparente los resultados a la Alta Gerencia. En el apartado 4.3.3. Determinar los pasos para la recolección de datos, se sugieren métricas que aplicarían a cada uno de los dominios levantados en la metodología.

4.3.1.2. Flujo de información

Existen tres fases principales a través de las cuáles fluye la información en el proceso de Ciencia de Datos: Fase de recopilación de información, fase de procesamiento de información y fase de generación de información.

A fin de entender las fases por las que atraviesa la información se ha levantado el siguiente gráfico y su respectiva explicación:

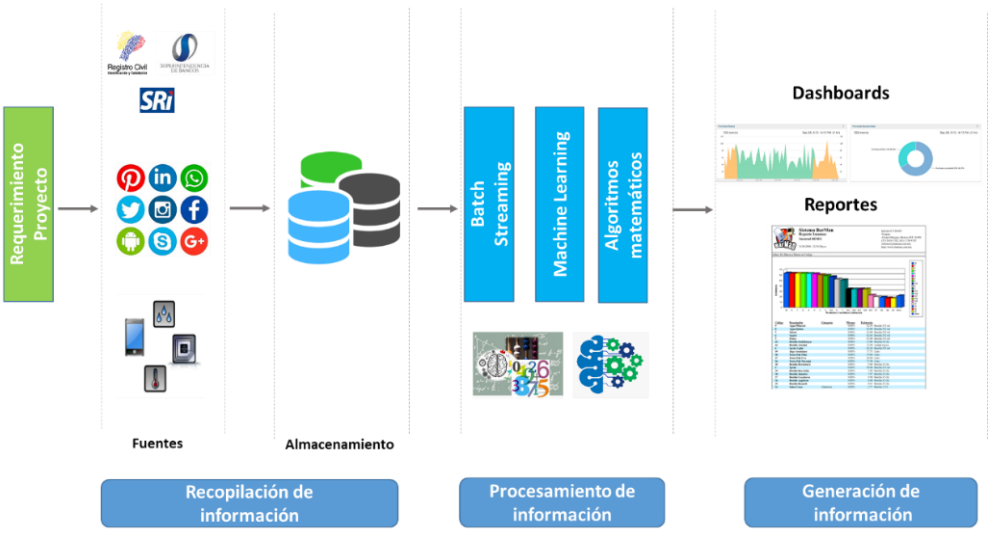

Figura 11. Flujo de información de ciencia de datos. Fuente: Elaborado por autor

- **Requerimiento de información:** El flujo nace de una necesidad de la institución, que previamente debió ser analizada entre el personal de ciencia de datos y el personal que solicita el requerimiento para verificar la factibilidad del resultado deseado, las fuentes que faciliten su implementación y los recursos necesarios.
- **Recopilación de información:** Para la extracción de información es necesario conocer el tipo de información que se va a procesar, en ciencia de datos se puede trabajar con información estructurada (bases internas de la institución o bases de información pública de entidades externas) y no estructurada (consultas a navegadores, información de redes sociales, sensores, dispositivos inteligentes, etc.).

  En este punto existen procesos de extracción de información que pueden ser ejecutados en línea o a través de lotes programados, por lo que este proceso es un punto importante a ser considerado en la metodología, ya que se debe

garantizar la integridad de data al trasladar la información desde el origen al destino (repositorios del proceso de ciencia de datos). Cabe mencionar la importancia de la ejecución de un proceso de Data Mining (Limpieza de datos), necesario para garantizar un nivel de calidad aceptable de la información con la que se pueda trabajar.

En esta fase es importante analizar la infraestructura de almacenamiento de la información, las tecnologías utilizadas y la arquitectura levantada en ciencia de datos. También, se deben tomar en cuenta controles sobre la gestión de contratos, con el fin de asegurar garantías legales y disponibilidad del servicio. En esta sección se evaluará los controles que se deben mantener, en el apartado de Seguridad de infraestructura y Seguridad lógica de la metodología.

Por otro lado, también se debe salvaguardar la información ante un posible evento, razón por la cual es necesario mantener controles de respaldos y planes de continuidad ante la materialización de un posible incidente. En la metodología se analizará este punto en el apartado de Gestión de Operaciones.

- **Procesamiento de información:** Una vez que la información está en las bases de datos, se define inicialmente el método de procesamiento (batch o streaming), el cual dependerá del proyecto que se esté ejecutando y en cada caso se aplicará diferentes controles. Posteriormente, la información se procesa y transforma en conocimiento, es decir, se genera información de valor para la institución. Para llegar a este punto se deberá evaluar los controles establecidos para levantar un adecuado modelado matemático que se adapte a resolver la necesidad planteada. Adicionalmente, se deberá identificar el algoritmo de machine learning escogido

para entrenar el modelo (supervisado o no supervisado) y al igual que en el anterior caso, se deberá evaluar los controles levantados para garantizar que el algoritmo elegido se ajuste a la necesidad y además sea el método más ágil para la resolución del problema.

Por último, es importante mantener controles sobre la Gestión de requerimientos de desarrollo, con el objetivo de garantizar un adecuado proceso de solicitud, entendimiento, desarrollo, pruebas y certificación de los proyectos realizados con la información y además que cubra las necesidades del requerimiento solicitado.

- **Generación de información:** Una vez que la información se encuentra probada y certificada se deberá mantener controles en el paso a producción de los cambios, a fin de asegurar que se mantenga la última versión de cambios certificados en los resultados de reportes, tableros de control y demás repositorios finales.
- **Procesos transversales:** Cabe mencionar que hay procesos transversales que deben cubrirse en todas las fases y deben tomarse en cuenta para el aseguramiento del proceso de ciencia de datos:
    - Se debe mantener controles que aseguren el cumplimiento de la normativa de riesgo operativo, resolución No. SB-2018-771 (las secciones que apliquen), a fin de evitar observaciones por parte del ente de control.

- Todos los procesos deben contar con políticas y procedimientos con el objetivo de mantener documentación actualizada de lo que se realiza y cómo se realiza, para dar cumplimiento a lo estipulado.

4.3.2. Realizar la planificación previa a la auditoría

Se realiza una evaluación del riesgo, con el objetivo de identificar la probabilidad y el impacto de ocurrencia de un evento. A continuación, se presenta un análisis de los riesgos, de acuerdo a cada buena práctica:

- **GTAG:**

  En la Tabla 10 se establecen los riesgos identificados de acuerdo a las guías GTAG:

Tabla 10
Riesgos identificados con GTAG

Tema de metodología	Proceso GTAG	Riesgos identificados
Gobierno de datos	Gobierno del programa de Big Data	Falta apropiada de financiación
		Falta de cumplimiento de metas estratégicas
		Falta de gobierno de información
		Soluciones tecnológicas ineficaces
Seguridad infraestructura	Disponibilidad	Indisponibilidad de la infraestructura
		Indisponibilidad de la infraestructura
		Seguridad ineficaz
Seguridad lógica		Seguridad ineficaz

Tema de metodología	Proceso GTAG	Riesgos identificados
	Seguridad y privacidad	Calidad de datos incorrecta
		Información no confiable e inexacta
Selección y capacitación	-	Personal no capacitado, acciones incorrectas
		Personal no capacitado, acciones incorrectas

Fuente: Elaborado por autor

- **COBIT5**

En la Tabla 11 se establecen los riesgos identificados de acuerdo a las buenas prácticas de COBIT:

Tabla 11
Riesgos identificados con COBIT5

Tema Metodología	Proceso COBIT	Riesgos identificados
Gobierno de datos	Gestionar el Marco de Gestión de TI	Riesgos a nivel de TI desconocidos por la alta gerencia, presidencia, directorio
	Gestionar la estrategia	Decisiones sobre inversiones y priorización no basadas en estrategia corporativa
		Estructura organizacional no soporta las operaciones de negocio
		Baja aceptación y confianza de las gerencias del negocio con respecto a los esfuerzos del área para cumplir los objetivos
		Procedimientos de tecnología no alineados a seguridad, servicio y reporte
		Fuga de información
Selección y capacitación del personal	Gestionar los recursos humanos	Dependencia de personal
		Personal de TI desmotivado, crecimiento interno limitado.
Seguridad de infraestructura	Gestionar la capacidad y la disponibilidad	Limitación en la capacidad de procesamiento

Tema Metodología	Proceso COBIT	Riesgos identificados
Gestión de operaciones	Gestionar la capacidad y la disponibilidad Gestionar Operaciones	Saturación de infraestructura
		Servidores con fallos inadvertidos - pérdida de disponibilidad
	Gestionar la Continuidad	Crecimiento de costos para recuperar las operaciones en caso de emergencias
	Gestión de operaciones	- Paso de información incompleta - Falta de integridad de la data
	Gestión de contratos	Paralización del servicio por falta de definición de SLA's
	Gestionar los respaldos	Pérdida de información no recuperada desde los respaldos generados
Seguridad lógica	Gestionar Servicios de Seguridad	Accesos inadecuados, escalamiento de privilegios
		Usurpación de identidades
Gestión de requerimientos de desarrollo	Gestión de Programas y Proyectos	Debilidad en la claridad de requerimientos
		Inserción de código no contralado
	Gestionar los Cambios	Afectaciones no advertidas a otros programas
	Gestionar la Aceptación del Cambio y la Transición	Pruebas que no cubren todos los escenarios
		Programas con errores funcionales
		Borrado accidental o intencional de parte del código, librerías o script
		Incumplimiento normativo

Fuente: Elaborado por autor

- **NIST**

En la Tabla 12 se establecen los riesgos identificados de acuerdo a la metodología NIST:

Tabla 12
Riesgos identificados con NIST

Tema Metodología	Proceso NIST	Riesgos identificados
Gobierno de datos	Sistema de Orquestación	- Proyectos ineficientes, que no responden al requerimiento solicitado - Esfuerzos y recursos desaprovechados

Seguridad lógica	Proveedor de datos Proveedor de aplicaciones Big Data Consumidor de datos	Falta de entrega de información integra y veraz
Seguridad de infraestructura	Proveedor de framework de Big Data	Indisponibilidad del servicio

Fuente: Elaborado por autor

Cabe mencionar, que en esta etapa de planificación y una vez realizado un entendimiento inicial del proceso y los posibles riesgos asociados se deberá definir un cronograma para las fases de trabajo de campo y emisión de informes. También se deberá asignar los recursos para ejecutar la evaluación.

4.3.3. Determinar los pasos para la recolección de datos

En esta etapa se identifica el programa de auditoría a aplicarse en la revisión y se determinan las evidencias que se solicitaran en el trabajo de campo. A continuación, se detalla el programa de auditoría, segmentado por los siete dominios de revisión definidos por el área de Auditoría Interna:

4.3.3.1. Dominio de Gobierno de Datos

Generalidades:

Se han establecido los siguientes aspectos que abordan datos generales del dominio de Gobierno de Datos:

Tabla 13
Dominio de Gobierno de Datos - Generalidades

Gobierno de datos	
Descripción del dominio	Los objetivos estratégicos de la institución deben proyectarse hacia la adopción de nuevas tecnologías, por lo que es importante que la Alta Gerencia realice un análisis de estado situacional y

Gobierno de datos	
	determine cuál es su expectativa al implementar un proceso de ciencia de datos y aterrizarlo en un objetivo de la institución. Una vez que se levanta este objetivo, la institución debe analizar cómo llevar a cabo un gobierno de información orientado a nuevas tecnologías y la inversión que conlleva. Asimismo, se deberá realizar un análisis que permita reconocer los riesgos que impidan llevar a cabo la consecución del objetivo. Se debe mantener un área de Ciencia de datos, que se encargue de llevar a cabo las actividades de este proceso, para lo cual se deben establecer iniciativas alineadas al objetivo planteado y mantener un seguimiento de cumplimiento de los resultados esperados. Por otro lado, las actividades clave de control que se deben mantener en este proceso incluyen la identificación de propietarios de datos, consumidores, datos maestros y fuentes de información autorizados. El dueño de datos tendrá la responsabilidad de asegurar la calidad y seguridad de los datos.
Objetivos del dominio	El objetivo principal del proceso de Gobierno de Datos es mantener controles en torno a: • Comunicación con la Alta Gerencia • Alineación estratégica • Definición de propietarios de datos y sus responsabilidades • Mantener la agilidad del proceso; garantizando la integridad, veracidad y accesibilidad de la información.
Políticas	Se debe considerar levantar las siguientes políticas: • Políticas que establezcan la frecuencia, temas a tratar y participantes para la presentación de resultados del proceso de ciencia de datos • Políticas de definición de propietarios de datos y sus respectivas responsabilidades
Roles y responsabilidades	Los principales roles en este dominio son: **Alta Gerencia** • Aprobar financiamiento • Dar seguimiento a las actividades del proceso de ciencia de datos y verificar su alineación estratégica • Establecer dueños de datos y aprobar sus responsabilidades • Verificar las métricas que demuestren el logro de objetivos **Gerencia de Ciencia de datos:** • Generar estrategias para alinearse a los objetivos levantados por la institución.

Gobierno de datos	
	• Levantar una planificación operativa alineada a las estrategias del área. • Presentación de resultados de la gestión de ciencia de datos a la Alta Gerencia • Proponer indicadores de cumplimiento de resultados **Gerencia de Ciencia de datos, Gerencia de Seguridad de la Información, Gerencia de Riesgos:** • Levantar un análisis de los activos de información que soportan al proceso de ciencia de datos • Realizar un análisis de vulnerabilidades que puedan afectar a los activos de información • Levantar un análisis de riesgo que afecten a la consecución de objetivos estratégicos del área y de la institución.
Métricas	Se recomienda levantar las siguientes métricas: • Porcentaje de objetivos de la institución basados en las metas estratégicas del área de Ciencia de datos. • Nivel de satisfacción de la Alta Gerencia con el alcance de los servicios y proyectos paneados. • Nivel de satisfacción de los usuarios finales con la capacidad de respuesta del proceso de ciencia de datos a nuevos requerimientos

Fuente: Elaborado por autor

Controles a evaluar

A continuación, se detallan los controles que se consideran necesarios evaluar en el Dominio de Gobierno de Datos:

Tabla 14
Dominio de Gobierno de Datos - Controles a evaluar

Tema	Detalle	Evidencias	Normativa/ Buena práctica
Alineación estratégica	Se debe garantizar que exista alineación con los objetivos de la institución	Actas de Comités	**COBIT:** - APO01: Gestionar el Marco de Gestión de TI - APO04: Gestionar la Innovación
	Se debe levantar un Comité de Ciencia de datos que asegure el seguimiento del cumplimiento de actividades.		

Tema	Detalle	Evidencias	Normativa/ Buena práctica
	Se debe definir dueños de datos y responsabilidades sobre la calidad de los datos.	Política de asignación de dueños de datos	**GTAG:** Gobierno del programa de Big Data **Normativa**: Artículo 10 - Literal c: Tecnología de la Información – Subliteral ii - Políticas y procedimientos – Alineación estratégica
	Asegurar un análisis del presupuesto para implementar un proceso de ciencia de datos	Presupuesto financiero	
Análisis de riesgo	Identificar activos críticos que soporten al proceso de ciencia de datos	Reporte de activos críticos del proceso	**NIST:** Sistema de Orquestación
	Realizar análisis de vulnerabilidades sobre activos críticos	Informe de análisis de vulnerabilidades	**COBIT:** APO01: Gestionar el Marco de Gestión de TI
	Levantar un análisis de riesgos del proceso	Análisis de riesgo del proceso	
Seguimiento de resultados	Llevar cabo reuniones que garanticen el seguimiento de las actividades en el proceso de ciencia de datos	Actas de comité	**COBIT:** APO01: Gestionar el Marco de Gestión de TI
Indicadores	Asegurar el levantamiento de métricas que garanticen el rendimiento y cumplimento de las actividades en el proceso de ciencia de datos	Presentaciones de indicadores	**COBIT:** - APO01: Gestionar el Marco de Gestión de TI **GTAG:** - Gobierno del programa de Big Data

Fuente: Elaborado por autor

Ciclo del dominio

Se presenta un resumen de las entradas y salidas que se mantienen en el dominio de Gobierno de Datos:

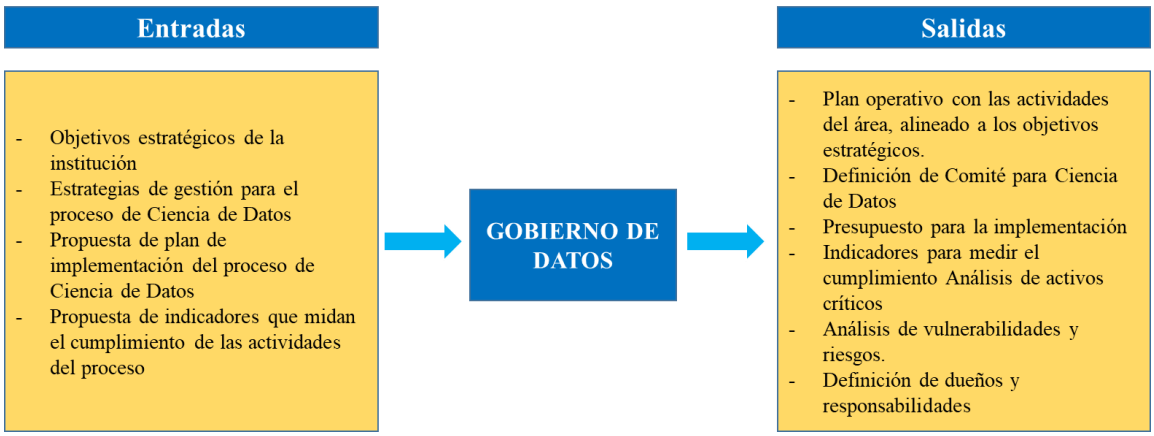

Figura 12 Dominio de Gobierno de Datos - Ciclo del dominio. **Fuente:** Elaborado por autor

4.3.3.2. Dominio de Selección y capacitación de personal

Generalidades:

Se han establecido los siguientes aspectos que abordan datos generales del dominio de Selección y capacitación de personal:

Tabla 15
Dominio de Selección y capacitación de personal - Generalidades

Selección y capacitación del personal	
Descripción del dominio	El dominio de selección y capacitación de personal hace referencia a los controles que se deben mantener para asegurar la contratación de personal competente para la asignación de funciones en el proceso de ciencia de datos. Asimismo, es importante mencionar a los planes de desarrollo y formación del personal, para asegurar que se mantiene conocimientos sólidos y actualizados que permitan generar valor a los proyectos que se realicen.
Objetivos del dominio	El objetivo principal del dominio de Selección y capacitación personal es mantener controles en torno a: • Contratación de personal eficiente para cargos del área de Ciencia de datos • Capacitación permanente al personal garantizando la disponibilidad de la información.
Políticas	Se debe considerar levantar o complementar las siguientes políticas: • Políticas de contratación de personal

	• Políticas de capacitación de personal
Roles y responsabilidades	Los principales roles en este dominio son: **Alta Gerencia:** • Aprobar descriptivos funcionales de funcionarios del área de Ciencia de datos. • Autorizar presupuesto para contratación y capacitación de personal de Ciencia de datos. **Gerencia de Ciencia de datos:** • Analizar y levantar un documento donde se describan las habilidades y conocimientos necesarios para contratar personal para ocupar un cargo en el área de Ciencia de datos. • Definir y aprobar descriptivo funcional • Indagar por capacitaciones que complementen y fortalezcan los conocimientos del personal de Ciencia de datos. **Gerencia de Recursos Humanos:** • Asesorar a Gerencia de Ciencia de datos para levantar el descriptivo funcional del personal. • Acompañar en el proceso de contratación de personal para garantizar las habilidades y conocimientos requeridos. Adicionalmente, ayudarán a verificar el nivel de estudio, la experiencia y los cursos tomados por los candidatos. • Mantener alineados los procesos de contratación de personal del área de Ciencia de datos acorde a las políticas y procedimientos de la institución.
Métricas	Se recomienda levantar las siguientes métricas: • Porcentaje de rotación del personal en el área de Ciencia de datos • Duración media de las vacantes

Fuente: Elaborado por autor

Controles a evaluar

A continuación, se detallan los controles que se consideran necesarios evaluar en el Dominio de Selección y capacitación del personal:

Tabla 16
Dominio de Selección y capacitación del personal - Controles a evaluar

Tema	Detalle	Evidencias	Normativa/ Buena práctica
Roles y responsabilidades	Mantener descriptivos funcionales en donde se detallen las actividades que ejecuta el personal del área de Ciencia de datos. Asimismo, el documento debe incluir habilidades y competencias requeridas, así como el tiempo y tipo de experiencia necesario.	Descriptivo funcional	**COBIT:** APO01 - Gestionar el Marco de Gestión de TI
Contratación de personal	Se deben definir las habilidades y conocimientos que requiere el personal que conformará el área de Ciencia de datos, Cabe mencionar que las disciplinas de programación, matemáticas, estadística y de conocimiento del negocio son fundamentales para estas posiciones.	Descriptivo funcional	**COBIT:** APO07 - Gestionar los Recursos Humanos **Normativa**: Artículo 10 - Literal b: Personas
	Confirmar la información del candidato en cuanto a nivel de estudio, capacitaciones tomadas, experiencia adquirida, etc.	Informes de contratación de personal	
	Alinear el proceso de contratación del personal del área de Ciencia de datos a los procedimientos de la institución	Políticas y procedimientos de contratación de personal	
	Evaluar las necesidades del personal de Ciencia de datos de forma regular, debido a que es un proceso cambiante.		
Capacitación de personal	Analizar las posibles capacitaciones requeridas por el personal del área de Ciencia de datos	Plan de capacitaciones	**COBIT:** APO07 - Gestionar los Recursos Humanos

Tema	Detalle	Evidencias	Normativa/ Buena práctica
	Evaluar las necesidades de capacitación para personal del área de Ciencia de datos de forma regular, debido a que es un proceso cambiante en el que se requiere estar actualizado.		**Normativa**: Artículo 10 - Literal b: Personas
	Mantener prácticas de entrenamiento cruzado	Reportes de ejecución de prácticas de entrenamiento cruzado	

Fuente: Elaborado por autor

Ciclo del dominio

Se presenta un resumen de las entradas y salidas que se mantienen en el dominio de Selección y capacitación del personal.

Figura 13. Dominio de Selección y capacitación del personal - Ciclo del dominio. Fuente: Elaborado por autor

4.3.3.3. Dominio de Seguridad de infraestructura

Generalidades:

Se han establecido los siguientes aspectos que abordan datos generales del dominio de

Seguridad de Infraestructura

Tabla 17
Dominio de Seguridad de infraestructura - Generalidades

Seguridad de infraestructura	
Descripción del dominio	El dominio de Seguridad de infraestructura describe los controles y mecanismos de seguridad implementados para proteger el hardware y los medios de almacenamiento de información. De tal manera que el proceso de ciencia de datos mantenga un esquema seguro y disponible sobre el cual sea soportado. El análisis de capacidad de almacenamiento y procesamiento son temas relevantes al hablar de este dominio, ya que se debe garantizar que el proceso responda de forma ágil y su infraestructura se adapte a los requerimientos de la institución. Otro factor relevante en este dominio es el levantamiento de logs o pistas de auditoría, en donde se deben registrar las actividades realizadas de los usuarios, fallas o eventos de seguridad sobre los componentes de infraestructura.
Objetivos del dominio	El objetivo principal del dominio de Seguridad de infraestructura es mantener controles en torno a: • Análisis adecuado de la infraestructura necesaria para soportar el proceso de ciencia de datos. • Levantamiento de una red de alta disponibilidad • Diseño de redes que mitiguen congestión con esquema de rendimiento eficiente y consistente • Diseño de arquitectura escalable • Monitoreo la infraestructura ante cualquier tipo de evento garantizando la disponibilidad y seguridad de los componentes del proceso de ciencia de datos.
Políticas	Se debe considerar levantar o complementar las siguientes políticas: • Políticas de gestión de infraestructura • Políticas de diseño de redes • Políticas de seguridad informática
Roles y responsabilidades	Los principales roles en este dominio son: **Alta Gerencia**

Seguridad de infraestructura	
	Aprobar presupuesto para implementación de infraestructura **Gerencia de Ciencia de datos:** • Levantar informe de solicitud en el que se detalle los requerimientos funcionales y técnicos necesarios para llevar a cabo la estrategia planteada para las actividades del área de Ciencia de datos • Solicitar la ejecución de análisis de vulnerabilidades a la infraestructura que soporta el proceso. **Gerencia de Tecnología:** • Asesorar a Gerencia de Ciencia de Datos en el levantamiento de requerimientos técnicos para la implementación de la infraestructura de este proceso. • Diseñar la arquitectura que soportará este proceso. El diseño debe tomar en cuenta las necesidades del proceso de ciencia de datos: trabajar bajo un esquema distribuido, infraestructura resistente a fallas y aplicación de controles para minimizar la congestión del tráfico. **Gerencia de Seguridad Informática:** • Gestionar la ejecución de análisis de vulnerabilidades a los componentes que conforman la infraestructura de ciencia de datos. • Analizar accesos a la infraestructura de ciencia de datos **Especialista de Infraestructura:** Monitorear la infraestructura con el objetivo de mantenerla disponible.
Métricas	Se recomienda levantar las siguientes métricas: • Número de interrupciones del proceso de ciencia de datos debido a incidentes de disponibilidad de infraestructura • Número de actualizaciones de temas referentes a infraestructura que soporta al proceso de ciencia de datos que no fueron planificadas.

Fuente: Elaborado por autor

Controles a evaluar

A continuación, se detallan los controles que se consideran necesarios evaluar en el Dominio de Seguridad de infraestructura

Tabla 18
Dominio de Seguridad de infraestructura - Controles a evaluar

Tema	Detalle	Evidencias	Normativa/ Buena práctica
Implementación y diseño de infraestructura	Mantener reuniones con el personal de Tecnología con el fin de evaluar los requerimientos funcionales de la estrategia de ciencia de datos y evaluar los requisitos técnicos necesarios para implementar la infraestructura, en ocasiones puede ser más rentable adquirir servicios de almacenamiento en la nube.		**COBIT:** BAI04 - Gestionar la capacidad y la disponibilidad
	Diseñar la arquitectura distribuida que soporte a ciencia de datos, donde prevalezca la disponibilidad, la tolerancia a fallas, la resiliencia en redes (ser capaz de identificar errores y elegir conectarse a otro nodo, de algunos nodos existentes). Este punto es importante debido a que incluso se mitiga la congestión de la red, al tener más de una alternativa por la cual pasar. Adicionalmente, la arquitectura debe estar diseñada de tal forma que considere un escalamiento de la infraestructura.	Arquitectura del proceso de Ciencia de Datos	**Normativa:** Artículo 10 - Literal c: Tecnología de la Información - Subliteral v – Infraestructura **NIST:** Proveedor de framework de Big Data **GTAG:** Disponibilidad de la tecnología y performance
	Realizar un análisis de capacidad de la infraestructura que soporta al proceso de ciencia de datos con una frecuencia por lo menos anual para	Análisis de capacidad de infraestructura	

Tema	Detalle	Evidencias	Normativa/ Buena práctica
	corroborar o mejorar la capacidad mantenida		
	Validar que la infraestructura de ciencia de datos se mantenga bajo los esquemas de seguridad de red de la institución, cuyos componentes se encuentren protegidos por los diferentes sistemas de seguridad (IDS, IPS, Firewall)	Arquitectura de red	
Monitoreo de Infraestructura	Levantar pistas de auditoría y logs sobre los dispositivos de infraestructura de tal manera que se pueda conocer los accesos o fallas de algún componente	Muestra de logs de componentes de infraestructura	**Normativa:** Artículo 10 - Literal c: Tecnología de la Información - Subliteral v – Infraestructura **COBIT:** DSS01 - Gestionar Operaciones
	Centralizar el monitoreo y la configuración, de tal manera que se facilite la gestión de alertas sobre la infraestructura.	Informes de resolución de casos en función de muestra de monitoreo de logs	
	Mantener procedimientos de monitoreo y atención de alertas en la infraestructura de ciencia de datos.	Procedimientos de monitoreo y atención de alertas	

Fuente: Elaborado por autor

Ciclo del dominio

Se presenta un resumen de las entradas y salidas que se mantienen en el dominio de Seguridad de Infraestructura:

Figura 14. Dominio de Seguridad de infraestructura - Ciclo del dominio. Fuente: Elaborado por autor

4.3.3.4. Dominio de Seguridad lógica

Generalidades:

Se han establecido los siguientes aspectos que abordan datos generales del dominio de Seguridad lógica:

Tabla 19
Dominio de Seguridad de Seguridad Lógica - Generalidades

Seguridad lógica	
Descripción del dominio	El dominio de Seguridad lógica está enfocado a la aplicación de medidas que garanticen el adecuado acceso a los sistemas de información del proceso de ciencia de datos con el concepto del mínimo privilegio tanto de usuarios internos de la institución, como proveedores. Además de asegurar estos componentes ante posibles fallos o ataques de seguridad.
Objetivos del dominio	El objetivo principal del dominio de Seguridad lógica es mantener controles en torno a: • Garantizar controles de acceso a personal autorizado, tanto interno como externo. • Monitoreo de los sistemas de información ante posibles ataques o fraudes externos • Proteger los componentes tecnológicos que soportan el proceso de ciencia de datos contra software malicioso

Seguridad lógica	
	• Mantener esquemas de seguridad aplicados a los dispositivos que soportan el proceso desde equipos de usuario final hasta servidores. • Asegurar el enmascaramiento de la información sensible existente dentro de los repositorios de ciencia de datos. • Garantizar, en la medida de lo posible, la encriptación de archivos y directorios que puedan ser accedidos por usuarios finales. • Levantamiento de procedimientos debidamente documentados y aprobados garantizando la seguridad y privacidad de la información.
Políticas	Se debe considerar levantar o complementar las siguientes políticas: • Políticas de administración de accesos a los sistemas de información • Políticas de seguridad informática
Roles y responsabilidades	Los principales roles en este dominio son: **Gerencia de Ciencia de datos:** • Levantar con Seguridad de la Información los usuarios y perfiles requeridos para el acceso a los sistemas de información de ciencia de datos. • Acompañar en el proceso de encriptación de la información de ciencia de datos, de tal manera que se asegure que datos no deberían ser enmascarados. • Acompañar en el proceso de encriptación de los archivos o directorios considerados sensible **Gerencia de Tecnología:** • Levantar con Gerencia de Seguridad de la Información los usuarios y perfiles requeridos para el acceso a la infraestructura de ciencia de datos, justificando su uso. • Acompañar en el proceso de encriptación de la información que se procesa de ciencia de datos, de tal manera que se asegure que datos pueden ser enmascarados. • Definir la encriptación de los archivos o directorios considerados sensible **Gerencia de Seguridad Informática:** • Gestionar la ejecución de análisis de vulnerabilidades a los componentes que conforman la infraestructura de ciencia de datos. • Llevar a cabo procedimientos de hardening en los sistemas de información.

Seguridad lógica	
	• Gestionar la encriptación de la información que se procesa en ciencia de datos. • Gestionar la inclusión de los equipos dentro del esquema de control de antivirus de la institución. • Gestionar la encriptación de los archivos o directorios considerados sensible **Gerencia de Seguridad de la Información** • Analizar accesos a los sistemas de información del proceso de ciencia de datos • Levantar matrices de usuarios y perfiles con acceso a los sistemas de información del proceso de ciencia de datos. • Analizar y asignar permisos a proveedores bajo lo reglamentado en la política de administración de accesos y bajo el concepto de menos privilegio. Justificar dicha asignación de acceso. **Especialista de Infraestructura:** • Incluir a los componentes de sistemas de información que apliquen dentro del esquema de protección antivirus de la institución. • Llevar a cabo procesos de encriptación de los archivos o directorios considerados sensible
Métricas	Se recomienda levantar las siguientes métricas: • Número de vulnerabilidades descubiertas • Número de usuarios no autorizados detectados en los sistemas de información • Número de equipos que se encuentran con el antivirus desactualizado

Fuente: Elaborado por autor

Controles a evaluar

A continuación, se detallan los controles que se consideran necesarios evaluar en el Dominio de

Seguridad lógica:

Tabla 20
Dominio de Seguridad Lógica - Controles a evaluar

Tema	Detalle	Evidencias	Normativa/ Buena práctica
Análisis de vulnerabilidades	Gestionar la ejecución de un análisis de vulnerabilidades	Informe de resultados de análisis de vulnerabilidades	**Normativa:** Artículo 16 - Literal j: Protección de información - Subliteral ix
Hardenización de equipos	Llevar a cabo procedimientos de aseguramiento de equipos, con el objetivo de limitar privilegios a usuarios finales	Detalle de las seguridades aplicadas a los equipos Equipos en los que se aplicó el aseguramiento de equipos	**COBIT:** DSS05 - Gestionar los Servicios de Seguridad **Normativa**: Artículo 16: Literal j: Protección de información – Seguridad de la Información **GTAG:** Seguridad y privacidad **NIST:** Proveedor de aplicaciones Big Data
Protección antivirus	Incluir a los componentes de sistemas de información dentro del esquema de protección antivirus de la institución	Verificación de actualización de antivirus de una muestra de equipos	
Enmascaramiento de información	Enmascarar la información considerada sensible, siempre y cuando sea posible y no dificulte el procesamiento de la información.	Procedimiento de designación de información sensible Listado de información que se enmascaró	
Encriptación de archivos	Llevar a cabo un proceso de encriptación de archivos y directorios, basándose en un análisis realizado por el dueño de datos y Tecnología	Llevar a cabo una prueba en sitio para verificar la encriptación	
Asignación de acceso a sistemas de información	Determinar usuarios y perfiles autorizados a tener acceso al software que soporta al proceso de ciencia de datos (acceso a bases de datos, herramientas de reporte, procesamiento y	Listado de usuarios y perfiles con acceso a software de información	

Tema	Detalle	Evidencias	Normativa/ Buena práctica
	sincronización de información).		
	Determinar los usuarios autorizados a acceder a los componentes de infraestructura por temas de soporte, mantenimiento y monitoreo.	Listado de usuarios con permisos de acceso a la infraestructura	
	Realizar un análisis de concordancia de los accesos otorgados a los sistemas de información de ciencia de datos	-	
	Asignar accesos a los sistemas de información de ciencia de datos, basado en las políticas de administración de accesos.	Listado de usuarios con permisos de acceso a la infraestructura	

Fuente: Elaborado por autor

Ciclo del dominio

Se presenta un resumen de las entradas y salidas que se mantienen en el dominio de Seguridad Lógica:

Figura 15. Dominio de Seguridad lógica - Ciclo del dominio. Fuente: Elaborado por autor

4.3.3.5. Dominio de Gestión de requerimientos de desarrollo

Generalidades

Se han establecido los siguientes aspectos que abordan datos generales del dominio de Gestión de requerimientos de desarrollo:

Tabla 21
Dominio de Gestión de requerimientos de desarrollo - Generalidades

Gestión de requerimientos de desarrollo	
Descripción del dominio	Este dominio se enfoca en mantener una metodología para la gestión de requerimientos, que cubra el ciclo de una implementación: análisis, desarrollo, pruebas y paso a producción, asegurando el cumplimiento de las funcionalidades acordadas con el cliente.
Objetivos del dominio	El objetivo principal del dominio de Gestión de requerimientos de desarrollo es mantener controles en torno a: • Gestionar adecuadamente las solicitudes, manteniendo un único canal de recepción de requerimientos y priorizándolos en base a lo definido en la política. • Identificar las fuentes de información de dónde se extraerá la información para la implementación del requerimiento (identificar si requiere de proceso de Data Mining) y definir el alcance • Mantener reuniones frecuentes con el usuario solicitante para comprender el requerimiento, definir alcance y acordar tiempos de entrega. • Identificar el tipo de procesamiento (batch o streaming) que se apega a la solicitud del cliente. • Verificar que se mantenga una metodología para el desarrollo de este tipo de proyectos • Identificar los procesos de entrenamiento que apliquen al modelo desarrollado • Validar que se cuenten con métodos de prueba de los modelos que permitan medir su calidad y eficacia • Identificar el uso de técnicas de seguridad en procesos de desarrollo en base a directrices de codificación segura.
Políticas	Se debe considerar levantar o complementar las siguientes políticas: • Políticas de proyectos • Políticas de desarrollo de requerimientos

Gestión de requerimientos de desarrollo	
Roles y responsabilidades	Los principales roles en este dominio son: **Usuarios finales:** • Solicitar requerimientos de implementación • Participar en las pruebas de usuario • Autorizar el paso a producción, después de certificar los resultados. **Gerencia de Ciencia de datos:** • Establecer políticas y procedimientos de implementación de requerimientos de ciencia de datos, en donde se defina la forma de priorizar los requerimientos y la certificación o paso a producción de una solicitud. • Levantar una metodología para el desarrollo de este tipo de soluciones, donde se especifique: levantamiento de casos de uso, tipo de procesamiento, plan de pruebas y requerimientos para pasar a producción • Verificar priorización asignada a los requerimientos solicitados **Especialista de Desarrollo de ciencia de datos:** • Levantar casos de uso en función de reuniones mantenidas con el usuario solicitantes, para establecer alcance y tiempo de implementación • Identificar las fuentes de donde se requiere extraer información, para afinar el alcance y el tiempo de implementación • Cumplir con lo detallado en las políticas, procedimientos y metodología de desarrollos establecidos para este tipo de implementaciones. • Levantar y ejecutar plan de pruebas que aseguren la calidad y funcionalidad del modelo desarrollado
Métricas	Se recomienda levantar las siguientes métricas: • Porcentaje de beneficios en base al funcionamiento de los proyectos de ciencia de datos • Nivel de satisfacción de los usuarios solicitantes al cierre del proyecto • Porcentaje de proyectos sin éxito debido al diseño incorrecto de los casos de uso

Fuente: Elaborado por autor

Controles a evaluar

A continuación, se detallan los controles que se consideran necesarios evaluar en el Dominio de Gestión de requerimientos de desarrollo:

Tabla 22
Dominio de Gestión de requerimientos de desarrollo - Controles a evaluar

Tema	Detalle	Evidencias	Normativa/ Buena práctica
Metodología, políticas y procedimientos de desarrollo	Mantener políticas, procedimientos y metodología de implementación de soluciones de ciencia de datos, la cual sea una guía para los especialistas para desarrollar un modelo. Deberá constar: levantamiento de casos de uso, criterios de priorización, tipo de procesamiento, procesos de entrenamiento, plan de pruebas y requerimientos para pasar a producción	Metodología de desarrollo	Artículo 13 - Literal a: Administración de proyectos
Entendimiento del requerimiento	Mantener reuniones con el usuario solicitante para realizar el entendimiento del requerimiento, a través de casos de uso: comprender el objetivo, definir el tipo de procesamiento y en base a esto determinar el alcance y tiempo de entrega.	Casos de uso generado	**Normativa**: Artículo 10 - Literal c: Tecnología de la Información - Subliteral iv Artículo 13 - Literal a, b y c: Administración de proyecto
Desarrollo y pruebas	Desarrollar en base a directrices de codificación segura Identificar el procedimiento de entrenamiento que aplique al modelo desarrollado.	Metodología de desarrollo	**COBIT:** BAI01 Gestión de Programas y Proyectos BAI06 - Gestionar los Cambios

Tema	Detalle	Evidencias	Normativa/ Buena práctica
	Llevar a cabo pruebas que certifiquen la calidad y el funcionamiento del modelo desarrollado		BAI07 - Gestionar la Aceptación del Cambio y la Transición
Paso a producción	Realizar reuniones con los usuarios solicitantes para que certifiquen el resultado deseado	Plan de certificación	
	Pasar a producción el modelo posterior a la confirmación del usuario solicitante	Confirmación de usuario solicitante	

Fuente: Elaborado por autor

Ciclo del dominio

A continuación, se presenta un resumen de las entradas y salidas que se mantienen en el dominio de Gestión de requerimientos de desarrollo:

Figura 16. *Dominio de Gestión de requerimientos de desarrollo - Ciclo del dominio. Fuente: Elaborado por autor*

4.3.3.6. Dominio de Gestión de operaciones

Generalidades

Se han establecido los siguientes aspectos que abordan datos generales del dominio de Gestión de operaciones:

Tabla 23
Dominio de Gestión de operaciones - Generalidades

Selección y capacitación del personal	
Descripción del dominio	El dominio de Operaciones está orientado a garantizar que los servicios provistos por ciencia de datos mantengan un alto grado de disponibilidad, haciendo referencia a una adecuada gestión sobre respaldos, planes de continuidad y contingencia tecnológica. También garantiza que se mantenga una adecuada gestión de incidentes, en donde se establezcan claramente los acuerdos de tiempo de respuesta ante un caso de evento. Asimismo, se verifica que se mantengan SLA's definidos para servicios que se mantengan externalizados. Por otro lado, se verifica que los procesos batch o en línea se ejecuten completa y correctamente.
Objetivos del dominio	El objetivo principal del dominio de Gestión de Operaciones es mantener controles en torno a: • Garantizar la ejecución de respaldos y su replicación a un sitio externo • Mantener planes de continuidad y contingencia en donde se especifiquen posibles escenarios de riesgo • Mantener procedimientos de gestión de incidentes en donde se establezcan los niveles de escalamiento y tiempos de resolución de un caso • Asegurar que se hayan estipulado SLA's en los contratos que se mantienen con proveedores que ofrecen servicios al proceso de ciencia de datos. • Monitorear los procesos calendarizados y en línea para garantizar que se hayan ejecutado correcta y completamente. En caso de presentarse errores garantizar el seguimiento y solución correspondiente. garantizando la disponibilidad del servicio.

Selección y capacitación del personal	
Políticas	Se debe considerar levantar o complementar las siguientes políticas: • Políticas de Gestión de Operaciones • Políticas de Planes de Continuidad y Contingencia • Políticas de Gestión de incidentes • Políticas de Contrataciones y Compras
Roles y responsabilidades	Los principales roles en este dominio son: **Gerencia de Ciencia de datos:** • Levantar contrato con proveedores que otorguen servicios externalizados en donde se detallen los acuerdos de niveles de servicio y su correspondiente penalización en caso de incumplimiento. • Levantar solicitud de generación de respaldos conforme a la necesidad de las operaciones de ciencia de datos. **Especialista de Operaciones:** • Ejecutar la generación de respaldos y asegurar que se envíe una copia a un lugar externo **Especialista de Infraestructura:** • Levantar planes de contingencia del servicio, detallando el plan de respuesta y tiempo de recuperación. • Levantar planes de continuidad en donde se analicen escenarios de riesgo en los que se pueda perder el servicio. Detallar planes de acción y mencionar tiempos de recuperación del servicio y de la información. **Gerencia de Atención de Incidentes:** Levantar procedimientos en donde se detalle el monitoreo de incidentes realizado, los niveles de escalamiento existentes, los responsables asignados y los tiempos de respuesta ante un caso determinado. **Especialista de Contrataciones y Compras** • Asesorar en el levantamiento de contratos y SLA's correspondiente, a la Gerencia de ciencia de datos. • Verificar que los contratos se apeguen a lo estipulado en las políticas internas de la institución.
Métricas	Se recomienda levantar las siguientes métricas: • Porcentaje de casos resueltos por el proveedor en un tiempo razonable • Porcentaje de incidentes resueltos en los tiempos establecidos en los procedimientos

Selección y capacitación del personal	
	• Porcentaje de copias y restauración de respaldos satisfactorias. • Porcentaje de casos que dieron cumplimiento a los tiempos de respuesta establecidos en los planes de continuidad y contingencia. • Porcentaje de casos con error de ejecución de procesos batch o en línea.

Fuente: Elaborado por autor

Controles a evaluar

A continuación, se detallan los controles que se consideran necesarios evaluar en el Dominio de Gestión de operaciones:

Tabla 24
Dominio de Gestión de operaciones - Controles a evaluar

Tema	**Detalle**	**Evidencias**	**Normativa/ Buena práctica**
Gestión de respaldos	Generación de respaldos de acuerdo a las necesidades del proceso de ciencia de datos	Logs de ejecución de respaldos	**COBIT:** DSS04 - Gestionar la Continuidad **GTAG:** Disponibilidad de la tecnología y performance
	Envío de copias de respaldos a un site alterno	Pruebas de restauración	
	Pruebas de restauración de información	Logs de ejecución de restauración	
Gestión de continuidad y contingencia	Mantener planes de continuidad y contingencia técnicos del proceso de ciencia de datos	Planes de continuidad y contingencia	**COBIT:** DSS04 - Gestionar la Continuidad **Normativa**: Artículo 10 - Literal d: Eventos externos - Plan de Continuidad
	Ejecutar pruebas por lo menos una vez al año de lo mencionado en los planes de continuidad y contingencia técnica	Informes de resultados de pruebas	
Gestión de incidentes	Mantener procedimientos de gestión de incidentes detallando: actividades de monitoreo de incidentes, tipos de incidentes,	Procedimiento de gestión de incidentes	**COBIT:**

Tema	Detalle	Evidencias	Normativa/ Buena práctica
	tiempo de respuesta a incidentes, niveles de escalamiento		DSS02 - Gestionar las Peticiones y los Incidentes del Servicio
	Registrar los casos de atención de incidentes en una bitácora, con el objetivo de generar una biblioteca de soluciones para futuros casos.	Bitácora de incidentes	**Normativa**: Artículo 10- Literal c: Tecnología de la Información - Subliteral iii
Gestión de contratos	Asegurar el levantamiento de acuerdos de niveles de servicio en los contratos que se mantiene con terceros	Contratos que se mantenga con proveedores	**COBIT:** APO10 - Gestionar los Proveedores **Normativa:** Artículo 14: Servicios provistos por terceros.
	Garantizar la inclusión de cláusulas en los contratos que se mantiene con los proveedores, en donde se encuentren: definición de productos y servicios a ser entregados por el proveedor, garantías técnicas, multas y penalizaciones, detalle de personal que brindará el servicio, confidencialidad, derechos de propiedad intelectual, etc.		
Gestión de ejecución de procesos	Mantener procedimientos para monitoreo de la ejecución de procesos batch y en línea	Procedimiento de monitoreo de ejecución de procesos Inventario de procesos que se ejecutan en ciencia de datos Bitácora con el registro de errores y solución	**COBIT:** DSS01 - Gestionar Operaciones **Normativa:** Artículo 10- Literal c: Tecnología de la Información - Subliteral iii
	Verificar que la ejecución de procesos sea en batch o en línea se ejecuten correctamente		
	Garantizar seguimiento de casos de error hasta que se le haya otorgado una solución al caso.		
	Mantener una bitácora con el registro de errores y solución		

Tema	Detalle	Evidencias	Normativa/ Buena práctica
		Muestra de logs de ejecución de procesos para realizar prueba	

Fuente: Elaborado por autor

Ciclo del dominio

A continuación, se presenta un resumen de las entradas y salidas que se mantienen en el dominio de Gestión de operaciones:

Figura 17. Dominio de Gestión de operaciones - Ciclo del dominio. Fuente: Elaborado por autor

4.3.3.7. Dominio de Controles de aplicación

Generalidades:

Se han establecido los siguientes aspectos que abordan datos generales del dominio de Controles de aplicación:

Tabla 25
Dominio de Controles de aplicación - Generalidades

Controles de aplicación	
Descripción del dominio	El dominio de controles de aplicación es aquel que verifica que la lógica funciona de acuerdo a lo requerido, garantizando la integridad de la información que se genera del proceso. Adicionalmente, se verifica que el modelo diseñado para solventar los requerimientos mantenga un nivel de precisión alto.
Objetivos del dominio	El objetivo principal del dominio de Controles de aplicación es mantener controles en torno a: • Integridad de la información • Confiabilidad y precisión del modelo implementado en el proceso de ciencia de datos garantizando la disponibilidad de la información.
Políticas	Se debe considerar levantar o complementar las siguientes políticas: • Políticas de desarrollo de requerimientos • Política de Gestión de Operaciones
Roles y responsabilidades	Los principales roles en este dominio son: **Especialista de Operación de Ciencia de Datos:** Asegurar que la información que se extrae del origen es la misma que llega al repositorio, aplicando controles que validen la integridad de la data. **Especialista de Desarrollo de Ciencia de datos:** Garantizar que el modelo matemático empleado para implementar el requerimiento mantiene un alto grado de precisión.
Métricas	Se recomienda levantar las siguientes métricas: • Porcentaje de beneficios en base al funcionamiento de los modelos • Nivel de satisfacción de los usuarios solicitantes de un modelo

Fuente: Elaborado por autor

Controles a evaluar

A continuación, se detallan los controles que se consideran necesarios evaluar en el Dominio de Controles de aplicación:

Tabla 26
Dominio de Controles de aplicación - Controles a evaluar

Tema	Detalle	Evidencias	Normativa/ Buena práctica
Integridad de la data	Mantener controles en todas las fases por las que atraviesa la información, garantizando la integridad de la data. Los nodos que deben garantizar la integridad son: - Desde el origen a los repositorios de ciencia de datos - Desde los repositorios a las unidades de procesamiento - Desde las unidades de procesamiento a los reportes expuestos.	Cuadre de integridad de la información fuente a los resultados finales	**GTAG:** Calidad de información **NIST:** Consumidor de datos
Precisión del modelo	Aplicar controles de medidas de precisión del modelo basados en métodos sobre conjuntos de datos de entrenamiento y prueba.	Métodos de precisión de modelos utilizados	**COBIT:** BAI08 - Gestionar el conocimiento **GTAG**: Calidad de información

Fuente: Elaborado por autor

Ciclo del dominio

Se presenta un resumen de las entradas y salidas que se mantienen en el dominio de Controles de aplicación:

Figura 18. Dominio de Controles de aplicación - Ciclo del dominio. Fuente: Elaborado por autor

CAPÍTULO V. CONCLUSIONES Y TRABAJOS FUTUROS

5.1. Conclusiones

El presente trabajo de investigación, propone una metodología para llevar a cabo una auditoría al proceso de ciencia de datos, enfocándose en garantizar la seguridad de la información que es procesada dentro de este entorno.

Para la elaboración del diseño de la metodología de auditoría se usaron las directrices del proceso auditor CISA, las cuales permitieron identificar que el alcance, el objetivo y el programa de auditoría, son tres componentes necesarios para construir la estructura de la metodología.

Se realizó un mapeo de buenas prácticas como COBIT, NIST y GTAG, las cuales permitieron analizar las fases del proceso de ciencia de datos desde varias aristas e identificar los dominios que fueron tomados en cuenta para establecer los controles de la metodología: gobierno de datos, seguridad lógica, seguridad física, gestión de operaciones, gestión de requerimientos de desarrollo y gestión de selección de personal. Adicionalmente, se tomó en cuenta a la normativa de riesgo operativo, resolución No. SB-2018-771, ya que la evaluación está orientada a una entidad financiera que está obligada a acatar lo estipulado por la Superintendencia de Bancos. Esta normativa sirvió para reforzar los temas de la metodología relacionados a continuidad, contingencia, seguridad lógica y física.

Se identificaron las etapas por las que se extrae, procesa y genera información en el proceso de ciencia de datos, reconociendo así los componentes que debían ser considerados para la evaluación de riesgos y levantamiento de controles. Se verificó que los controles más

representativos y exclusivos para ciencia de datos están relacionados a la capacidad de la infraestructura, el monitoreo de ejecución correcta de procesos, la confiabilidad y precisión de los modelos y los controles en torno a la integridad de la data.

El concepto de *Big Data* fue considerado dentro de la metodología, ya que es un componente que a corto o largo plazo se debe adoptar, tanto por el volumen de información que abarca como por el tipo de datos que se procesa y la exigencia de respuesta inmediata requerida. Razón por la cual se consultaron buenas prácticas relacionadas a la arquitectura (NIST, 2018) y al gobierno y calidad de información emitidas en las guías GTAG (The Institute of Internal Auditors, 2017).

De acuerdo al análisis de las disciplinas que conforman al proceso de ciencia de datos, se identifica la necesidad de mantener personal especializado y capacitado en los campos de estadística, matemática, programación y conocimiento del negocio.

5.2. Trabajos futuros

El objetivo de este trabajo se centró en el levantamiento de la metodología para auditar el proceso de ciencia de datos, dejando para trabajos futuros los resultados de la ejecución de la metodología y la evaluación de riesgos correspondiente, para verificar si el nivel de riesgo ha subido o ha bajado y tomar las medidas que ameriten, de ser el caso.

Se puede complementar la metodología, con controles dirigidos a la disponibilidad y seguridad de la información, en torno a la infraestructura que soporta el proceso, analizando alternativas modernas y funcionales, tales como el almacenamiento de información en la nube.

Existen otras alternativas de buenas prácticas que establecen controles por dominio, tal como la ISO27002, bajo la cual también se podría levantar una nueva metodología y comparar los resultados obtenidos con el presente trabajo.

La metodología debe ser analizada con una periodicidad de por lo menos un año, con el objetivo de mantenerla actualizada y se adapte a los cambios de una tecnología que evoluciona a pasos agigantados.

BIBLIOGRAFÍA

Chen, H., Chiang, R. H., & Storey, V. C. (s.f.). Business intelligence and analytics: From big data to big impact. *MIS quarterly, 36*(4).

Anchiraico, J. C. (Febrero de 2017). Diseño de una arquitectura Big Data para la predicción de crisis en el trastorno bipolar. *(Master's thesis)*.

BBVA. (2017). Big Data en el ecosistema "Fintech". 2-13.

Braga, G. (2012). COBIT 5 aplicado al sistema de registro contable informático argentino.

Contraloría General del Estado. (2014). Planificación de la Auditoría. *CGE, V*, 82-87.

Conway, D. (2013). *The Data Science Venn diagram*. Obtenido de http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Cooke, I. (2017). Auditoría básica de SI: Programas de auditoría.

Domo. (2018). Data Never Sleeps 6.0. Obtenido de Data Never Sleeps 6.0.

Ernst & Young. (Marzo de 2017). Data & Advanced Analytics Survey. 3-4.

Friedman, J. H. (2001). The Role of Statistics in the Data Revolution? *International Statistical Review, 69*(1), 5-10.

Galimany Suriol, A. (Julio de 2014). La creación de valor en las empresas a través de Big Data. *(Master's thesis)*.

García Ceca, J. (Octubre de 2017). Metodología para auditar Big Data. *(Master thesis)*.

ISACA. (2012). COBIT 5 Procesos catalizadores. 35-215.

ISACA. (2015). *Manual de preparación al examen CISA 2015.* Rolling meadows: ISACA. Obtenido de CISA.

Jiménez, P. (12 de Diciembre de 2018). Proceso de ciencia de datos en la institución. (V. Ramírez, Entrevistador) Quito.

Liu, A. (2018). *Data Science and Data Scientist.* New York: IBM Analytics. Obtenido de IBM Analytics.

López García, D. (2012). Análisis de las posibilidades de uso de Big Data en las organizaciones. *(Master's thesis)*, 3-6.

Management Solutions. (2015). Data Science y la transformación del sector financiero.

Marr, B. (21 de Mayo de 2018). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. (Forbes, Ed.) Obtenido de Forbes.

Morrobert, E. (26 de Mayo de 2011). Auditoría de Sistemas. (ISACA, Ed.) *ISACA*.

NIST. (2000). Definitions of Terms and Modes Used at NIST for Value-Assignment of Reference Materials for Chemical Measurements.

NIST. (Junio de 2018). NIST Big Data Interoperability Framework: Volume 6, Reference Architecture. *6*(2), 10-18.

O'Neil, C., & Schutt, R. (2013). *Doing Data Science: Straight Talk from the Frontline.* California: O'REILLY.

Pérez Marqués, M. (2015). *Big Data Técnicas, herramientas y aplicaciones* (Vol. 2). México: Alfaomega.

Piattini, M., & Del Peso, E. (2001). Auditoria informática. Un enfoque práctico. *6*, 1-2.

Provost, F., & Fawcett, T. (2013). Data Science and its relationship to Big Data and Data Driven decision making. *1*(1), 2-4.

Puyol, J. (2014). Una aproximación a Big Data. *Revista de Derecho UNED*(14).

Rayo, A. (6 de julio de 2016). El mundo Internet of Things (IoT) y su aportación al Big Data. *Big Data Fundations*.

Rollins, J. (Junio de 2015). Metodología fundamental para la ciencia de datos. Obtenido de IBM.

Serrano, C. (1 de Septiembre de 2017). El Boom del Big Data. *Vistazo*. Obtenido de El Boom del Big Data.

Superintendencia de Bancos. (2013). Normas generales para las instituciones del sistema financiero. 1287.

Superintendencia de Bancos. (13 de Agosto de 2018). Resolución No. SB-2018-814.

The Institute of Internal Auditors. (2017). *Understanding and Auditing Big Data.* Lake Mary: IPPF.

Van der Aalst, W. (2016). *Process Mining - Data Science in action.* Berlin: Springer.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *34*(2), 77-84.

Working Party. (16 de septiembre de 2014). Opinion 8/2014 on the on Recent Developments on the Internet of Things.